

# 制約つきグラフ探索を実現する異種データベース統合技術

鈴木 源吾<sup>1,a)</sup> 鬼塚 真<sup>1,b)</sup> 榎本 俊文<sup>1,c)</sup> 小林 伸幸<sup>1,d)</sup>

**概要：**交通検索・地図検索・SNS 検索等でグラフ情報を活用するサービスが、近年多く開発され実用化されている。今後は、情報の表現形式や探索能力に異種性のあるグラフデータベースと既存のデータベースや Web サービスを組み合わせた統合検索が重要になると考えられる。本論文では、異種データベース統合技術をグラフデータ操作に拡張し、グラフ探索とノード・エッジのプロパティに対する制約条件とを組み合わせる、制約つきグラフ探索を可能とする手法を提案する。グラフデータベースを含む複数のデータベーススキーマを統合し、対応関係のメタデータを構築する。検索要求時にそのメタデータを探索し、情報源に対する問合せの組合せを決定する動的な側面と、情報源の能力に応じてグラフ探索を最適化できる点に特長がある。グラフデータベースへの探索の一部を、情報源側に適切にプッシュダウンすることで高速化することが可能である。時間が限定されたサービスが実施されているノードを経由する最短経路問題の一種である、時間制約つき寄り道経路探索に本手法を適用し有効性を示した。交通情報とサービス情報の分散状況・グラフデータベースの探索能力の差異・探索に用いる解法に基づいたグラフ探索操作のプッシュダウン方法を明らかにし、分散グラフデータベース環境における実験で評価し、プッシュダウンが有効になる基本導出法の有効範囲が広いことを示した。

## 1. はじめに

交通検索・地図検索・SNS 検索等でグラフ情報を活用するサービスが、近年多く開発され実用化されている。交通・地図検索としては駅探のような鉄道路線検索やバス路線の検索サービス、Google Map による道路網のナビゲーションサービス等が Web 上で容易に利用可能となっている。また、最近公開された Facebook のグラフ検索では、人の属性や関係性（例：サイクリングが好きで近くに住んでいる友達）を表す問合せによって SNS グラフを検索することができる。これらのサービスは、グラフ情報を扱う専用のデータベースであるグラフデータベース [1][2] によって実現されるようになりつつある。グラフデータベースでは、情報はノードとそれをつなぐエッジ、それぞれに付与されるプロパティでモデル化され、最短経路探索やグラフパターンマッチを高速に実行することができる。

このようなグラフ情報を探索するサービスは、既存の情報サービスと組合せることによって更に価値が高まると考えられる。例えば、交通検索と、レストラン情報サービス（例：食べログ）と、チケット予約サービス（例：JR のチ

ケット予約）を組合せ、旅行のルートと日程を決定し、鉄道の予約の空き情報を調べる等といった探索である。このようなグラフ探索は、経由するノードのプロパティに関する条件を制約とするグラフ探索となることが多く、本論文では、それを制約つきグラフ探索と呼ぶこととする。制約つきグラフ探索の一例として、時間制約つき寄り道探索がある [3][4]。それは、利用者の時間制約（例：門限）とサービス側の時間制約（例：開店時間）を満たす条件で、指定したサービス（例：店の開店）を持つノードを経由する最短経路を求める探索である。

このような制約つきグラフ探索では、分散し独立に設計運用されている、道路・鉄道等の交通路に関するグラフデータベースと店・開店時間等のサービスに関するデータベースを組み合わせて利用する必要がある。しかし、それらは相互利用を前提としていないため、実現には以下の 3 つの課題がある。

**課題 1：情報源の異種性の解消** 独立して設計された情報源には、情報の構造・表現形式・命名に違い（異種性）がある。その違いを解消しつつデータ統合検索を実現する必要がある。

**課題 2：サービスの仮想化と適切な問合せ候補生成** 制約つきグラフ探索は、時間制約つき寄り道探索のようなサービスであり、それを如何に仮想的に情報源とみなし、どのような場合に問合せ候補とすべきかという点

<sup>1</sup> NTT ソフトウェアイノベーションセンタ  
NTT, Minato-ku, Tokyo 108-0075, Japan  
a) suzuki.gengo@lab.ntt.co.jp  
b) onizuka.makoto@lab.ntt.co.jp  
c) enomoto.toshifumi@lab.ntt.co.jp  
d) kobayasi.nobuyuki@lab.ntt.co.jp

が課題となる。

**課題3：データ統合検索の最適化** グラフデータベースの機能は、RDBにおけるSQLのような標準言語が存在しないため、情報源毎にばらつきがある。また、Web経由で能力が公開されるケースが多いため、そのポリシーによって能力がすべて公開されるとも限らない。そのような制約つきグラフ探索の能力のばらつきに対応し、探索処理を分解してできるだけ情報源へ処理を移譲するプッシュダウンを実施する最適化が必要がある。そこで本論文では、独立に構築された複数のデータベースを仮想的に統合する異種分散データベース統合技術を拡張し、制約つきグラフ探索操作を含め、複数情報源を連携し統合利用を実現する手法を提案する。応用で想定しているサービス情報において、サービス自体の追加・削除、機能拡張が多く行われるため、筆者らが提案した、サービス変更に柔軟な動的な特性を持つメタデータを検索要求時に探索して異種性解消するデータ統合検索技術 [5][6] を利用した。制約つきグラフ探索操作を階層的にモデル化することにより、情報源の処理できる能力を考慮し、情報源側に適切なプッシュダウンを行い高速化できることが特長である。文献 [7][8] で提案した概念グラフを用いたスキーマ統合手法をグラフデータベースへ拡張しメタデータを構築する。

本論文では、上記手法の技術の詳細を述べ、時間制約つき寄り道経路探索へ適用し有効性を検証する。2章で異種分散データベース技術とグラフ探索に関する関連研究を述べる。3章と4章が本手法の主となる技術的な提案であり、それぞれスキーマ統合手法とデータ統合検索手法である。5章で時間制約つき寄り道探索に本手法を適用し、その評価結果を示す。6章で本手法の適用範囲と効果を考察し、7章でまとめを述べる。

## 2. 関連研究

データの形式や表現に異種性があり分散した情報源を、自律性を維持しつつ（既存のデータベースやサービスに極力手を入れない）、仮想的に統合検索を行う技術は、連邦データベース（federated database）技術 [9]・メディエータ技術と呼ばれている [10][11][12][13]。連邦データベースは、関係データベース等のレガシーな情報源を対象としていたが、メディエータ技術は、XML・Web 情報・マルチメディア情報等の木構造・半構造のデータを対象とすることが特長である。様々な異種性を持つデータベースのスキーマを統合する技術は、スキーマ統合技術 [14][15] と呼ばれており、ER モデルを対象とする研究 [14] から、近年では、論理記述や知識処理を用いた研究 [16][17][18] に至っている。

本論文は、近年応用が進展しているグラフデータベースを対象情報源とする点で既存研究と異なっている。グラフ情報に特有な探索操作に対して、スキーマ統合したり問合

せ最適化を行ったりする研究はこれまで行われていなかった。メディエータにおいては、情報源の能力のばらつきを考慮する技術の研究がなされている [19]。本論文は、既存のメディエータ技術では実現していたなかったグラフデータベースを対象とした情報源の能力のばらつきを考慮した最適化を行えることが特長である。

制約つきグラフ探索は、旅行計画問合せの一種であり、最短経路探索などを求めるグラフアルゴリズム、一定の制約のもとでタスクを最適に配置するスケジューリングアルゴリズム等に関連が深い。寄り道探索 [4] と Optical Sequential Route 探索といった既存研究がある。両手法とも寄り道先である POI（Point of Interest）を経由する探索であるが、寄り道先が1つ・複数という差がある。これらの手法は、単一情報源のメモリ・ファイル・データベースにすべての情報と能力が集中していることを前提としている。本研究は、複数の情報源を組み合わせ、これらのグラフ探索を実現する点で既存研究と異なっている。また、旅行計画問合せ的なグラフ探索としては、過去の旅行道程のグラフデータベースに対して、自分の行きたい旅行道程パターンのグラフそのものを検索キーとして入力し、検索結果を得るグラフ内容検索や軌道検索（trajectory search）の研究 [20] もある。本手法は、グラフデータベース機能を仮想的な表としてモデリングしているため、これらの研究に適用することは不可能ではないが、仮想的なデータ項目を探索し、探索機能におけるデータ項目の依存関係を利用する本研究は、グラフ内容検索では有効に利用することはできない。

## 3. グラフデータベースのスキーマ統合

本章では、データ統合検索で利用するメタデータを構築するためのスキーマ統合について述べる。

### 3.1 概念グラフを用いたスキーマ統合

本論文では、筆者らが提案している概念グラフ [21] を用いたスキーマ統合手法を適用する [7][8]。この手法は、同じ概念を異なる型（データモデル構成要素）で表すことにより生じる異種性である構造異種を減少させることに特長がある。概念グラフは、実世界と概念と関係という2つのデータモデル構成要素のみによってモデル化し、ER モデル（UML も同等）や関係モデルが持つ項目の集まりに相当するデータモデル構成要素を持たない。各情報源の ER スキーマを概念グラフに変換すると、その過程で構造異種が減少し発生しにくくなり、スキーマ統合の比較も容易にすることができる。

### 3.2 グラフデータベースからの概念グラフ変換

グラフデータベースからスキーマを作成し、概念グラフに変換することにより、前記の概念グラフを用いたスキーマ

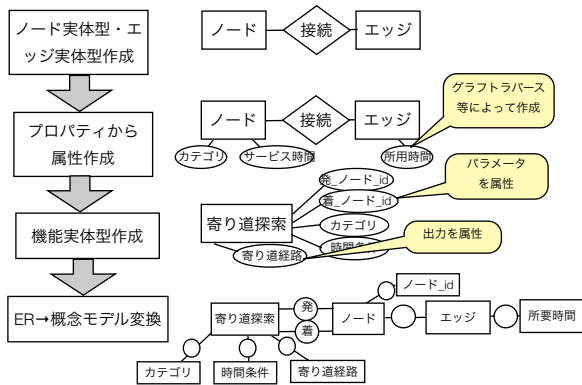


図 1 グラフ DB スキーマから概念グラフへの変換

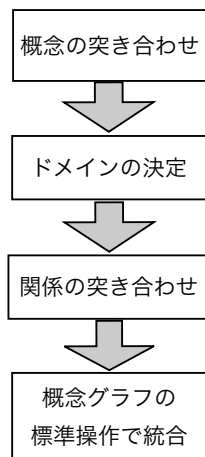


図 2 概念と関係の突き合わせと統合手順

マ統合手法を利用することができる。変換の基本方針は、「ノード」「エッジ」「機能」に相当する実体型を持つ ER モデルを作成し、既存の ER モデルからの概念グラフ変換手法を適用することである。変換の手順を図 1 に示す。

まず、ノードとエッジに相当する実体型を作成し、ノードとエッジが持つプロパティを属性として追加する。これはグラフをトラバースする等の手段によって作成する（ノード・エッジにより持つプロパティが異なる可能性があるため）。また、駅ノードと店ノードのように種類の異なるノードは区別する。

次に機能実体型を作成する。これは、例えば、ダイクストラ探索や寄り道探索のようなグラフデータベース機能を仮想的な表としてとらえた実体型である。機能のパラメータと機能の出力を属性として生成する。

以上で ER モデルが作成されるので、既存手法によりそれを概念モデルに変換する。

### 3.3 概念と関係の突き合わせと統合

変換された複数の概念グラフの概念と関係を突き合わせ統合する。その手順を図 2 に示す。

まず概念を類似度によってランキングし、突き合わせを

行い同一性を決定する。その結果、概念と情報源のデータ項目の関係が決定するため、それを用語辞書と呼ぶメタデータに蓄積する。

次に概念とそれに対応するデータ項目のドメインの決定を行う。ドメインとは抽象的な表現形式を表し、同じ意味を表す表現形式の集まりであるドメイングループによりグループ化されている（詳細は文献 [5]）。例えば、日付というドメイングループに西暦日付、和暦日付というドメインが含まれる。概念をドメイングループに対応させ、対応する情報源のデータ項目のドメインを決定する。ドメイン間の変換関数がなければ作成する。この関係はドメイン辞書に登録する。

最後に同一と判定された概念間にある関係を名称類似度でランキングして比較し、同一性を判断する。

最後に概念グラフの標準操作により統合を行い統合概念グラフを作成する。同一視された概念と関係は重複して生成しないようにする。

## 4. 動的なデータ統合検索方式とグラフデータベース拡張

本章では、スキーマ統合により構築されたメタデータを利用した動的なデータ統合検索方式について述べ、それをグラフデータベースに拡張する。

### 4.1 方式の概要

方式の概要を図 3 に示す [5][6]。利用者は検索したい概念と概念に対する条件のみを指定し、その概念がどの情報源からどのように得られるかを指定しない。データ統合方式は、問合せ要求から、その要求する概念が存在する情報源を特定し、概念間の繋がりを補完し、問合せの処理順番を決定する「問合せ変換機能」と、利用者が選択した問合せ候補を実行し検索結果を得る「問合せ実行機能」の 2 つの機能から構成される。様々な情報源は仮想的な表としてモデル化され、問合せ処理も関係モデルにおける操作として記述される。関係モデルの採用は、XML 等の階層を表現するためには不十分であるが、データ統合検索を単純で整理されたモデルによって実現できるメリットがある。

問合せ変換機能は、指定された概念から、用語辞書を利用し対応する情報源のデータ項目を特定し、統合概念グラフを探索することにより概念間の関係を得て、ドメイン辞書から利用者側の表現形式と情報源の表現形式の関係から変換ルールを決定する。概念間の関係は、必要に応じてテーブル間やデータベース間を結合する問合せに変換される。

利用者は、適切な問合せ候補を選択し、データ統合検索を実行する。問合せ実行機能は、問合せ候補に含まれる情報を解釈し、複数の情報源から検索したデータを統合し、表現形式を変換して返却する。複数の候補を選択した場合、それぞれの検索結果の和集合が結果として返却される。

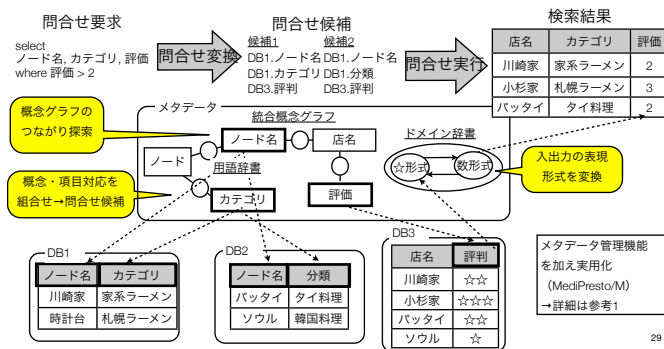


図3 動的なデータ統合検索方式

表1 既存手法との比較

| ビュー定義方式<br>観点 | SQL・XQuery | ルールベース        | 提案手法        |
|---------------|------------|---------------|-------------|
| 動的・柔軟性        | ×          | ○<br>(実行時組合せ) | ○           |
| 統合能力          | ○          | ○             | △           |
| 定義の再利用性       | ×          | ○<br>(ルール単位)  | ○           |
| 必要スキルレベル      | △          | ×             | ○<br>(知識処理) |

図3の例は、グラフデータベースを含む3つのデータベースから、統合概念グラフの探索、ドメインの変換、命名異種の解消等を行い問合せ候補を作成し実行している。

本方式の特長は、多くの既存方式で採用されているSQL等の言語による静的なビューを定義せずに、検索要求時に動的にメタデータを探索して、問合せ候補を生成することにある。この方式は、情報源の追加・スキーマ変更時の柔軟性が高い。メタデータは個々のビューに縛られておらず断片的であり、再利用性が高いからである。本方式と既存の方式との比較を表1に示す。動的で定義の再利用性のある点で、本方式はルールベースでビューを定義する手法に類似していると言える。しかし、ルールベースのビュー定義では知識処理言語を利用する必要があり、統合記述能力は豊富なものの汎用的すぎて定義が困難である問題がある。本方式は、異種性解消に特化してメタデータの知識を構築する方式であり、ルールベースの方式に比べ、統合を記述する能力はやや劣るものの、定義が容易であることがメリットである。

本方式の採用により、1章であげた3つの課題のうち課題1の異種性解消が解決される。

## 4.2 Web 情報源による制約つきグラフ探索の実現

1章の課題2のサービスの仮想化と適切な問合せ候補生成を解決するために、Web 情報源を利用したサービスの仮想化を利用する。よって、グラフデータベースへのアクセスはWeb 経由であるという前提を置く。現在、最もポピュラーと考えられる Neo4j には REST API が用意されてお

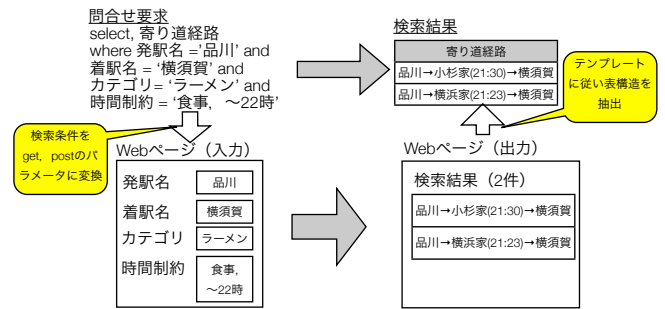


図4 Web 情報源機能を利用したグラフ探索の仮想化

り、多くのグラフ探索サービスがWeb 上から提供されているために、この前提は不自然ではない（セキュリティ上の課題はある）。

Web 情報源の利用例を図4に示す。問合せ要求に対し、検索条件を get, post のパラメータに変換し、Web ページの取得要求を行い、結果として得られる Web ページをテンプレートに従い表構造を抽出し、検索結果として返却する。

スキーマ統合の途中過程で得られたグラフデータベースの ER モデルの実体型（ノード、エッジ、機能）を、仮想的な表として、本手法の Web 情報源を定義する。そのとき、データ項目に対する制約を定義する。Web 情報源では、データ項目に入力のみ可能な項目、出力のみ可能な項目という制約がある。制約つきグラフ探索のパラメータである発ノードや着ノードは入力のみ指定可能であり、探索結果の経路は出力のみに可能な項目である。利用者の指定でこの制約を違反する場合は問合せ候補から除外する。ここまでは既存の Web 情報源技術で実現できる。この Web 情報源によって仮想化されたグラフデータベースと既存の情報源を組み合わせることにより、例えば、グラフデータベースに存在するカテゴリ情報より粗いカテゴリ情報を関係操作の結合により組合せ、他のデータベースにあるカテゴリ情報で寄り道探索を行うことが可能になる。また、複数のグラフ探索結果を関係操作の和として取得することも可能となる。

## 4.3 グラフ探索能力の階層化を利用した問合せ変換と最適化

前節の Web 情報源による手法で、課題2のサービスの仮想化と問合せ候補生成は基本的には実現されているが、情報源側が必要なグラフ探索能力を完全に持つ必要があるという点で、課題3に情報源の能力に応じた問合せ最適化が実現されていない。情報源側にその能力がなければ問合せ候補から除外されてしまう。そこで、グラフ探索能力を階層的に定義し、可能な限り情報源側に実行させる方式を提案する。

まず、単純な例として最短経路を求めるダイクストラ探索を用い説明する。ダイクストラ探索は、エッジに移動コストの数値プロパティを持つグラフに対し、始点・終点を



指定し合計移動コストが最小になる経路を求める探索である。グラフデータベース情報源として、ダイクストラ探索を持っていないケースを想定してみる。前述の Web 情報源による手法では、このグラフデータベース情報源にダイクストラ探索を実行することはできない。しかし、以下の条件が成立する場合を想定する。

- グラフデータベース情報源は、ノード・エッジ情報を取得する基本能力を外部に Web 経由で提供している。
- グラフデータベース情報源には、移動コストに相当するプロパティがエッジに存在している。
- データ統合検索機能側で、ダイクストラ探索を実行することができる。

この想定であれば、データ統合検索機能側でダイクストラ探索を実行し、グラフデータベース情報源にアクセスするときにノード・エッジ情報を取得する基本能力を利用すれば、ダイクストラ探索を機能として持たない情報源に対して、ダイクストラ探索を実行することが原理的に可能である。

この考え方を一般化し、ある制約つきグラフ探索があるとき、以下の仮定を置けば、情報源の能力の有無に関わらず、最終的に制約つきグラフ探索を実現できる。

- グラフデータベース情報源は、ノード・エッジ情報を取得する基本能力を外部に Web 経由で提供している。
- グラフデータベース情報源には、制約つきグラフ探索に必要なプロパティが存在している。
- データ統合検索機能側で、制約つきグラフ探索を実行することができる。

これは制約つきグラフ探索の能力の階層化の最も単純な例であり、最低レベルの基本能力と最高レベルの制約つきグラフ探索能力の 2 通りしかない場合である。一般の制約つきグラフ探索においては、探索と制約のチェックが分離でき、探索後に制約チェックを行い探索結果を求められる場合がある。そのような場合、グラフ探索能力の階層的な組合せとして、探索+制約としてメタデータに管理する。問合せ変換機能は、情報源がどのレベルの階層まで処理できるかをチェックし、グラフ探索実行に必要なプロパティの存在を調べ、存在すれば問合せ候補とし、さらに可能な限り高いレベルの階層の処理を情報源側にプッシュダウンし、適切な問合せ候補を生成する。能力の階層化の管理イメージを図 5 に示す。階層の高い能力は、下位の能力を利用して実現され、能力が高くなるほど、利用するプロパティが増えるという性質を持つ。同図で、k-top の最短経路探索をその近隣施設の値の条件でしばり込む制約つきグラフ探索の例を示している。

## 5. 提案手法を利用した異種分散データベース環境における時間制約つき寄り道探索

本章では、階層的な能力管理を用いた制約つきグラフ探

制約つきグラフ探索の能力階層



能力階層と必須プロパティの例

| level | 能力                   | 必須プロパティ               |
|-------|----------------------|-----------------------|
| 3     | k-top最短経路探索+近隣施設フィルタ | id, 移動コスト, 近隣施設 (ノード) |
| 2     | k-top最短経路探索          | id, 移動コスト (エッジ)       |
| 1     | 基本能力                 | id (ノード・エッジ)          |

図 5 グラフ探索能力の階層化の管理

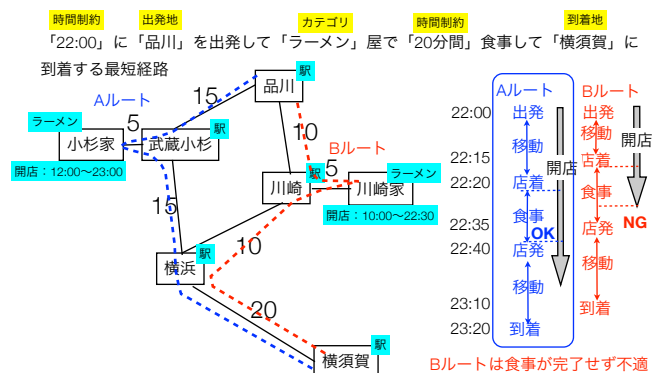


図 6 時間制約つき寄り道探索

索のプッシュダウン方式を利用し、時間制約つき寄り道探索を異種分散データベース環境で実現できることを示し、評価する。

### 5.1 時間制約つき寄り道探索の定義と解の導出法

時間制約つき寄り道探索は、筆者らが文献 [3] で提案したグラフ探索である。例えば、開店時間には 12 時～22 時のように制限がある店に到着できるようなルートを求めるグラフ探索である。夜遅めに出発するときには、閉店に間に合うように出発点に近い POI に寄り道する必要がある。「その店で 2 時間ゆっくりしたい」「銀行でお金をおろすのでぎりぎりに間に合えばよい」というような、ユーザにとっての時間制約の要求もあり、これらも満たすような探索となっている。詳細な定義は文献 [3] を参照してほしい。そのイメージを図 6 に示す。

その解の導出法には、大沢他 [4] で提案されている寄り道探索のための逐次拡大法をそのまま利用し、それに時間制約つき寄り道探索解を導出する方法を組み合わせた基本導出法と、さらに時間制約を積極的にグラフ探索に利用する動的導出法を提案している。動的導出法は、探索範囲と

解候補の個数を制限し性能を改善できる。

## 5.2 時間制約つき寄り道探索の能力の階層化

時間制約つき寄り道探索能力を階層化する。基本導出法は、時間制約のない寄り道探索と、制約のチェックに処理を分離することができるから、能力は以下の3レベルに階層化することができる。

- ・レベル1能力：ノード取得・エッジ取得の基本能力
- ・レベル2能力：時間制約のない寄り道探索
- ・レベル3能力：時間制約のある寄り道探索（基本導出法と動的導出法）

レベル2の必須プロパティは、エッジのコスト情報と、寄り道先の選択に利用するノードのカテゴリ情報である。レベル3の必須プロパティは、レベル2の必須プロパティに加えて、ノードの時間制約（開店時間等）情報となる。また、エッジのコスト情報が、距離ではなく時間長である制約もある。

寄り道探索は、両端からのダイクストラ探索を利用しているため、ダイクストラ探索を階層として設定できそうだが、単純な組合せとして寄り道探索を実現できないため、不可能である。

## 5.3 時間制約寄り道探索の問合せ最適化

本節では時間制約つき寄り道探索の処理が、情報源の能力や分散形態によって、どのように実行されるかを示す。ここで注意すべき点として、基本導出法は、時間制約なしの寄り道探索の解を求めてから、時間制約をチェックするという2段階処理であるために、単一データベースの場合は不利であったものの、分散データベース構成では寄り道探索を情報源側に行わせられる場合があるため、動的導出法に比べて有利になることが期待されることである。以下に最適化で考慮される3つの観点を示す。

**情報源に対する能力の仮定** 前節で述べたような3つのレベルが想定される。

**時間制約寄り道探索の処理方式の選択** データ統合側で基本導出法、または動的導出法を選択できる。

**データの分散パターン** データの分散には図5-7に示すような3パターンがある。パターン1は集中である。パターン2は分散1と呼ぶが、店へ至るまでの交通情報と、店の詳細情報が情報源が分かれているものである（駅と店の繋がりは交通情報側に持つ）。パターン3は分散2と呼ぶが、店を含まない交通情報と、店情報に情報源が分かれているものである（駅と店の繋がりは店情報側に持つ）。

この3つの観点に対し、本手法を適用するとき、まずデータの分散パターンに対しては、既存技術と同じ、可能な限り1つの情報源から情報を取得しようとする戦略に従うとする。情報源の能力と処理方式の選択については、前

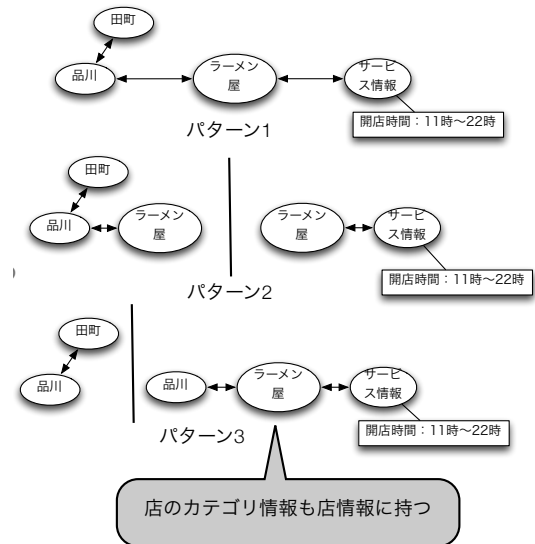


図7 データの分散パターン

表2 最適化観点の組み合わせと情報源側の処理

| 導出法      | 基本導出法 |     |       | 動的導出法 |    |       |
|----------|-------|-----|-------|-------|----|-------|
|          | L1    | L2  | L3    | L1    | L2 | L3    |
| 観点1→(能力) |       |     |       |       |    |       |
| 集中       | 基本    | 寄り道 | 時間寄り道 | 基本    | 基本 | 時間寄り道 |
| 分散1      | 基本    | 寄り道 | 寄り道   | 基本    | 基本 | 基本    |
| 分散2      | 基本    | 基本  | 基本    | 基本    | 基本 | 基本    |

・寄り道探索を情報源側で実行可能性→性能向上→分散向き

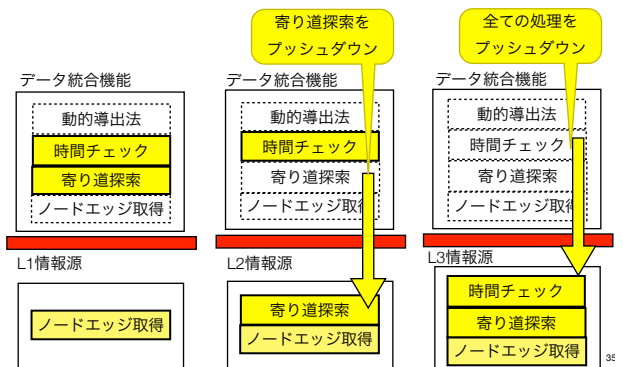


図8 情報源側への処理プッシュダウン：基本導出法

章で述べた階層管理を利用したプッシュダウン戦略に従う。

この3つの観点を組み合わせ、情報源側とデータ統合検索側の処理分担が決定される。表2は、その情報源側の処理を示している。残りの処理がデータ統合検索側で実行されることとなる。可能な限りレベルの高い処理を情報源側で実行する戦略に従うため、情報源能力がレベル2で、分散1で、基本導出法の場合は、寄り道探索処理はプッシュダウンすることが可能である。基本導出法と動的導出法のプッシュダウンの実現のイメージを図8、図9にそれぞれ示す。

- ・処理できない→プッシュダウンケース少ない→分散不向き

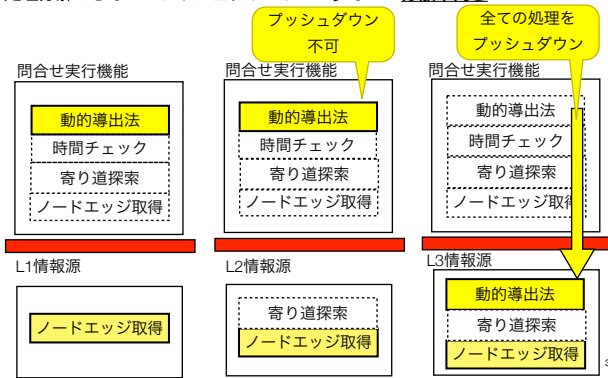


図9 情報源側への処理プッシュダウン：動的導出法

- ・実装機能
  - ▶ データ統合：問合せ実行機能（寄り道探索機能のみ）
  - ▶ 情報源：L1～L3能力
- ・クエリは1つ固定
  - ▶ 出力：片道約40分の寄り道
  - ▶ レスポンスタイム

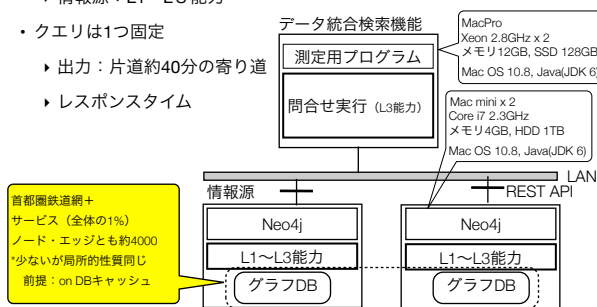


図10 実験環境の構成

## 5.4 実験と評価

プッシュダウン効果を評価する実験を行った。首都圏鉄道網をグラフデータベース化し、全ノード約1%のサービスノードを作成した（ノード・エッジとも約4000件）。データとしては小規模だが、グラフの局所的性質が同じであり、プッシュダウン効果の一次評価としては十分である。実験環境の構成を図10に示す。グラフデータベースにはNeo4j[2]を用いている。情報源は先に述べた分散パターンによって、3組（集中・分散1・分散2）のデータベース環境を構築した。グラフデータベースの基本能力は、Neo4jの機能をそのまま使い、寄り道探索、時間制約つき寄り道探索等はNeo4jのユーザ定義関数の組み込み機能を用いて実装し、REST API経由でアクセスしている。

行き・帰りともに40分以内という条件で、時間制約付き寄り道探索を、前節で述べた観点の組み合わせ毎に測定した。評価結果を表3に示す（単位：秒）。

表内の赤と橙で囲んだ部分がグラフ探索能力のプッシュダウンが適用されている範囲である。基本導出法においては、プッシュダウンを行える範囲が広く、動的導出法に比べて実用的な性能の領域が広いことがわかる。情報源能力がレベル1であったり、基本能力のみで探索を行う場合、ノードやエッジを取得するたびに、ネットワークアクセスが発生する。ノード・エッジの取得は、グラフ探索の過程

表3 プッシュダウン効果検証実験の結果

単位：秒

| 導出法      | 基本導出法 |       |       | 動的導出法 |       |       |
|----------|-------|-------|-------|-------|-------|-------|
| 観点1→(能力) | L1    | L2    | L3    | L1    | L2    | L3    |
| 集中       | 257.4 | 2.2   | 2.1   | 180.6 | 180.6 | 0.5   |
| 分散1      | 254.8 | 2.3   | 2.3   | 181.6 | 181.6 | 181.6 |
| 分散2      | 198.6 | 198.6 | 198.6 | 124.4 | 124.4 | 124.4 |

において非常に多いため、性能は劣悪である。プッシュダウンの適用がある場合、ない場合という条件が合致しているときに、基本導出法と動的導出法を比較すると、動的導出法の結果が優れている。これは文献[3]における単一データベースにおける検証と合致している。

この評価結果は、プッシュダウンを利用した本方式の最適化戦略が有効であることを示している。また、時間制約つき寄り道探索を異種分散データベース環境で実現する場合は、基本導出法の選択が有効であることも示唆している。

## 6. 考察

本手法の適用範囲について考察する。本手法は、グラフ探索処理をWebインターフェースとして仮想化されている場合に、一般的に利用できると言える。本論文では、寄り道探索のPOIは一箇所という仮定であったが、それが複数箇所になった場合でも適用は可能である。グラフ探索は、そのパラメータ指定はノードやそのプロパティであることが多いため、適用範囲は広い。また、処理のプッシュダウン制御によって、仮に情報源側で高度な能力を持っていない場合でも、基本能力が公開されていれば、基本能力を利用して、データ統合側で高度な能力を使うことによって、論理的にはすべてのグラフ処理を行うことが可能になる（グラフ探索は基本能力の組み合わせで実現できる）。ただし、実験結果に示したように実用的な性能を得ることは難しい。

ただし、本手法は情報源の仮想化は関係モデルによって行っている。寄り道探索の返却結果は、経路情報をJSON等の文字列で返却されることが想定されている。この文字列を分解し、必要な情報を取り出す処理は利用者（データ統合検索の開発者）側で実施する必要がある。経路情報の表現形式変換関数も利用者責任で作成する必要がある。また、グラフ類似検索・軌跡検索のようにグラフをパラメータとして渡す場合、それらを文字列等に変換して情報源側に渡せば適用は可能であるが、項目の対応を基礎としている本手法のメリットは必ずしも生かせない。

また、本手法は、情報源を跨るグラフ探索には対応していない。情報源を跨るデータ統合処理は、関係演算であ

る結合と和に限られている。例えば、鉄道網の最短経路探索が各鉄道会社毎に構築されているとき、それぞれの路線に閉じた最短経路を求め和として返却することは可能であるが、両社の路線を辿って検索する結果を求めることができない。このような分散的なグラフ探索は問題としては難しく、個別の探索毎に有効な手法も異なるため、一般的な異種分散データベースのフレームワーク内で整理するのは、今後の課題である。本手法は、例えば、情報源にない階層カテゴリを付加して寄り道探索を行うなど、単一情報源では実現できないデータ統合検索を容易に実現することができる有用な手法であって、最適化制御の考え方も導入したグラフデータベースを統合検索するための基本的ではあるが重要な一歩である。

## 7. おわりに

制約つきグラフ探索処理を階層的に管理することにより、情報源の能力に応じた、情報源側へのプッシュダウンが可能になる異種データベース統合手法を提案した。既存の関係モデルに基づく、動的な異種性解消を行うデータ統合検索における Web 情報源機能を利用して、制約つきグラフ探索処理向けに拡張した。統合に必要なメタデータを得るためのグラフデータベースのスキーマ統合方式も明らかにした。本方式を、時間制約つき寄り道探索へ適用し、情報源へのプッシュダウン効果が有効であり、特に基本導出法を用いる場合に性能改善できる場合が多いことを示した。

今後の課題としては、他のグラフ探索処理（例：POI 複数化）へ適用し、関係モデルの拡張方式の限界を見極めることと、探索以外のグラフ処理への適用がある。

## 参考文献

- [1] Cheng, J., Ke, Y., NG, W.: Efficient Query Processing on Graph Databases, ACM Transactions on Database Systems, Vol. 34, No. 1, pp. 1-48 (2009).
- [2] Neo4j WEB Page, <http://neo4j.org>
- [3] 鈴木源吾, 榎本俊文, 小林伸幸, 山室雅司, 鬼塚真: 時間制約をもつ寄り道経路探索システムの実現と評価, 情報処理学会論文誌 Vol.53 No.2 pp.857-867(2012).
- [4] 大沢裕, 藤野和久: 前処理を必要としない道路ネットワーク上での最短寄り道経路探索アルゴリズム, 電子情報通信学会論文誌 D, Vol.J93-D, No.3, pp. 203-210 (2010).
- [5] Honishi, T., Suzuki, G., Kobayashi, N., Konishi, K.: A Mediation System Based on Universal Relation Modeling, 20th International Conference on Conceptual Modeling Proceedings (ER2001), SE3, pp.1-4 (2001).
- [6] Suzuki, G., Ilzuka, Y., Kasuga, S.: Integration of Keyword Bases Source Search and Structure Bases Information Retrieval, 7th International CODATA Conference, pp.149-158 (2000).
- [7] 山室雅司, 鈴木源吾: データ標準化と概念グラフへの変換を利用したスキーマ統合支援法, 電子情報通信学会論文誌 D-1 Vol.J79-D-I No.11 pp.966-974 (1996).
- [8] Suzuki, G., Yamamuro, M.: Schema Integration Methodology Including Structural Conflict Resolution and Checking Conceptual Similarity - Conceptual Graphs Approach -, International Workshop on Database Reengineering and Interoperability, pp.229-242 (1995).
- [9] Sheth, A.P. and Larson, J.A.: Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Survey, Vol. 22, No.3, pp.183-236 (1990)
- [10] Hammer, J., Garcia-Molina, H., Ireland, K., Papakonstantinou, Y., Ullman, J., and Widom, J.: Information Translation, Mediation, and Mosaic-Based Browsing in the TSIMMIS System, Proceedings of the ACM SIGMOD International Conference on Management of Data, pp.483 (1995).
- [11] Levy, A.Y., Rajaraman, A. Ordille, J.J.: Querying Heterogeneous Information Sources Using Source Descriptions, Proceedings of the 22nd VLDB Conference (1996).
- [12] Halevy, A., Rajaraman, A. Ordille, J.J.: Data integration: the teenage years. In Proceedings of the 32nd international conference on Very large data bases, pp.9-16 (2006).
- [13] Zaman, M.: Information Integration for Heterogeneous Data Sources, IOSR Journal of Engineering Vol. 2(4) pp.640-643 (2012)
- [14] Batini, C., Lenzerini, M., Navathe, S.B.: A Comparative analysis of methodologies for database schema integration, ACM Computing Survey, Vol.18, No. 4, pp.323-364 (1986).
- [15] Kim, W. and Seo, J.: Classifying Schematic and Data Heterogeneity in Multidatabase Systems, Computer, Vol.24, No.12, pp.12-18 (1993).
- [16] Lenzerini, M.: Data Integration: A Theoretical Perspective, In Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '02) (2002).
- [17] Madhavan, J., Bernstein, P. A., and Rahm, E.: Generic Schema Matching with Cupid. Proc. VLDB, pp.49-58, (2001).
- [18] Bernstein, P. A., Madhavan, J., Rahm, E.: Generic Schema Matching, Ten Years Later. PVLDB 4(11): pp.695-701 (2011)
- [19] 林孝志, 小西一也, 堀口恭太郎, 綱川光明, 鈴木源吾, 芳西崇: 異種情報源統合のための XML 問い合わせ最適化と情報源問合せ能力管理, 情報処理学会論文誌 Vol.44 No.SIG12(TOD 19) pp.1-10 (2003).
- [20] Chen, Z., Shen, H.T., Zhou, X., and Zheng, Y.: Xing XieSearching Trajectories by Locations? An Efficiency Study, In ACM SIGMOD International Conference on Management of Data (SIGMOD 2010), Indianapolis, Indiana, USA.
- [21] Sowa, J.F.: Conceptual structures: information processing in mind and machine, Addison-Wesley (1984).