

機関番号：35311

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500107

研究課題名（和文） 次世代パーソナル情報管理システムにおけるデータモデル・検索に関する研究

研究課題名（英文）

研究代表者 プラダン スジツト (Pradhan Sujeet R)

倉敷芸術科学大学・産業科学技術学部・教授

研究者番号：90320001

## (1) 研究成果の概要（和文）：

リレーショナルデータベースをグラフ構造を用いて表現し、次に、その構造に対してキーワード検索を行い、グラフの一部を適切な解として見なす研究がすでにある。これらの研究では、検索結果の単位が、検索時に指定される全てのキーワードが出現する極小サブグラフ、即ち従来の“minimality”に基づくものである。しかし、パーソナル情報管理システムにおける異種性を持つ大量の情報を表現するグラフにおいては、検索キーワードが出現するノード数が必ずしも1には限らず、利用者にとってより適切な解は、極小サブグラフを更に拡大する必要があると考えられる。従って、従来の“minimality”概念に基づいたアプローチによる検索結果の単位の決定法が不十分であり、“beyond minimality”概念による検索結果の単位を生成するために必要な代数的な演算の提案を行った。

次に、従来の情報検索にシステムにおいては、適合度の高い順に検索結果の要素を並べることが一般的である。つまり、検索結果は、単なる要素の集合型ではなく、ランク付けリスト型である。従って、従来のデータベースにおいては、代数的な演算を適用し、検索結果を効率よく求めることが可能であった。しかし、パーソナル情報管理システムにおいては、それと同様な手法を直接導入することが混乱であることが明らかになった。

更には、パーソナル情報管理システムにおける異種性を持つ大量の情報中でも、電子教科書やそれをもとに作られた教材（パワーポイントなどの資料）は、特定な性質を持っている。それらの情報は、同期化メディアストリーム的一种であり、それぞれに補足情報が含まれているといえる。これらの同期化メディアストリームにおける部分的な情報を効率よく効果的に検索するためには、まずそれらの情報のモデル化が必要であると考え、同期化メディアストリームの相対関係が表現可能な2次元モデルを提案した。

## 研究成果の概要（英文）：

In recent years, there has been growing interest in keyword search over traditional relational databases. Usually a relational database is represented as a graph structure; an appropriate subgraph is computed as an answer to a keyword query. In these studies, the units of query results is based on the concept of ‘minimality’, that is, each answer subgraph is the smallest subgraph containing each query keyword. However, the concept of ‘minimality’ cannot be applied to keyword search over a graph representing heterogeneous information in a personal information management system. It is because an appropriate answer is not necessarily a subgraph comprising the nodes having at least one occurrence of each query. As a result, further expansion of the concept of ‘minimality’ needs to be need to be considered. In this research, we proposed the concept of "beyond minimality" for computing answer units in keyword search over a graph representing heterogeneous information. We also explored algebraic operations to generate such answer units.

In a traditional information retrieval system, it is common to rank search

result in a descending order of relevance. In other words, search result, rather than just a collection of elements, is a ranked list type. Thus, as query result in a traditional database system is a simple collection of tuple elements, query optimization is achieved through various strategies --- one of them being the manipulation of algebraic operations. However, it is obvious that the similar kind of optimization cannot be guaranteed in keyword search over a personal information management system.

Electronic textbooks and PowerPoint slides which refer to those textbooks have specific characteristics in a personal information management system. Such data can be categorized as a special kind of synchronized media streams can contain additional temporal information. In order to efficiently and effectively find information in a large collection of synchronized media streams, we have proposed a data model that represents information in a two-dimensional space. This space can be manipulated by our query processor.

#### 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
20 年度	1,400,000	420,000	1,820,000
21 年度	1,100,000	330,000	1,430,000
22 年度	900,000	270,000	1,170,000
23 年度	0	0	0
24 年度	0	0	0
総 計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：パーソナル情報データベース、検索結果単位、代数的な検索モデル、問い合わせ処理最適化

#### 1. 研究開始当初の背景

人類によって創出される情報量は、企業、研究機関や政府機関などにおける大規模な計算機のみならず、個人が利用するデスクトップやノートパソコンまで、爆発的に増大していることが明らかになりつつある。これらの膨大な情報は、従来のテキスト形式や表形式で表現できる情報に限らず、動画像、音楽、メール、写真、ブログやスケジュールなど様々な構造や形式、いわゆる異種型データ (heterogeneous data) から構成されていることが一般的であり、それを管理するパーソナル情報管理システム (Personal Information Management Systems) 技術の必要性が急増するとと思われる。

従来のデータベースシステムでは、大量の構造化されたデータ、また XML 形式のような半構造化データを構築し、効率良く検索する機能が充実しているが、パーソナル情報管理システムにおける多様な構造を持つデータや動画像、音楽、写

真のようなマルチメディアデータを組織化し、統合的な枠組みにより管理する機能やそれをシームレス (seamless) に検索する機能が不足している。一方、デスクトップにおけるパーソナル情報の検索を、Web 情報の検索手段と同様な手段を用いて支援するデスクトップサーチツールが開発されつつある。これを代表するものとしては、

① Google デスクトップ検索 (<http://desktop.google.com/>),

②マイクロソフト社のデスクトップサーチ

(<http://www.microsoft.com/windows/desktopsearch>) や

③アップル社の Apple Spotlight (<http://www.apple.com/macosx/features/spotlight/>)

などのようなものがあり、IT 業界の大企業がパーソナル情報管理システム技術を重要な企業戦略の中に位置づけていることが明らかである。また、スイス ETH 大学で行われている iMeMex プロジェ

ク ト  
([http://www.dbis.ethz.ch/research/current\\_projects/iMeMex](http://www.dbis.ethz.ch/research/current_projects/iMeMex))では、個人が持つ大量で多様な情報をデータベース化し、それを従来のデータベース問い合わせ言語 SQL や XQuery と同様な言語を用いて検索する技術が開発中である。この検索手法では、問合せを行うための専門的知識や検索対象となる情報の構造を利用者が予め把握し、検索の際に情報構造を指定する必要があるため、依然として一般の利用者にとって利便性の高いものとは言えないという課題が残されている。パーソナル情報管理システムの IT 業界での重要性や現在の検索手法に残されている課題を鑑みると、パーソナル情報管理システムにおける検索手法の研究開発は、学術的にも優先度の高い重要な研究課題であることは明白である。

## 2. 研究の目的

本研究の目的は、従来のパーソナル情報管理システム研究における、

- ①デスクトップサーチツールのようなデータベース的な機能が欠如している、
- ②一般利用者にとっては使いにくいという状況を克服し、

デスクトップ上の異種データのような個人が持つ大量の異種情報を、統合的にデータベース化し、それを誰もが容易に細かい単位で検索できるパーソナル情報管理システムの開発である。

## 3. 研究の方法

研究期間の初年度は、パーソナル情報管理システムにおける異種性を持つ大量の情報を組織化、次に、関連する情報の統合化を行い、無用な情報重複や不整合が生じるのを防げる新たなデータモデルを提案した。これによって、①異種型データの統合表現、②異種型データ間の複雑な相互関連の表現、③検索結果の対象になりうる最小情報単位の表現が可能になった。

次に、このデータモデルによる構築される異種型情報の検索手法としては、誰もが容易に活用できるようにキーワード入力のみでの検索手法を用いた。しかし、この検索手法では、「情報全体を表わすグラフのどの部分が検索解になるのかが明確ではない」という点が、最も困難な問題点になると考えられる。この問題を解決するひとつの方法として、グラフの部分を動的に生成することが可能にさせる新たな問合せモデルを提案した。しかし、この検索手法においては効率の問題があり、その原因を明確にした。

パーソナル情報管理システムにおける異種性を持つ大量の情報中でも、同期化メディアストリームに注目した。これらの同期化メディアストリームにおける部分的な情報を効率よく効果的に検索するためには、まずそれらの情報のモデル化を提案し、今後、このデータモデルによる構築される同期化メディアストリームをキーワード入力のみでの検索手法を用いた場合、「検索結果になりうるメディア部分」を求めるために必要な検索モデルの理論を提案することを次の課題として残した。

## 4. 研究成果

リレーショナルデータベースをグラフ構造を用いて表現し、次に、その構造に対してキーワード検索を行い、グラフの一部を適切な解として見なす研究がすでにある。これらの研究では、検索結果の単位が、検索時に指定される全てのキーワードが出現する極小サブグラフ、即ち従来の“**minimality**”に基づくものである。しかし、パーソナル情報管理システムにおける異種性を持つ大量の情報を表現するグラフにおいては、検索キーワードが出現するノード数が必ずしも1には限らず、利用者にとってより適切な解は、極小サブグラフを更に拡大する必要がありうると考えられる。従って、従来の“**minimality**”概念に基づいたアプローチによる検索結果の単位の決定法が不十分であり、“**beyond minimality**”概念による検索結果の単位を生成するために必要な代数的な演算の提案を行った。

次に、従来の情報検索にシステムにおいては、適合度の高い順に検索結果の要素を並べることが一般的である。つまり、検索結果は、単なる要素の集合型ではなく、ランク付けリスト型である。従って、従来のデータベースにおいては、代数的な演算を適用し、検索結果を効率よく求めることが可能であった。しかし、パーソナル情報管理システムにおいては、それと同様な手法を直接導入することが混乱であることが明らかになった。

更には、パーソナル情報管理システムにおける異種性を持つ大量の情報中でも、電子教科書やそれをもとに作られた教材（パワーポイントなどの資料）は、特定な性質を持っている。それらの情報は、同期化メディアストリームの一つであり、それぞれに補足情報が含まれているといえる。これらの同期化メディアストリームにおける部分的な情報を効率よく効果的に検索するためには、まずそ

これらの情報のモデル化が必要であると  
考え、同期化メディアストリームの相対  
関係が表現可能な2次元モデルを提案  
した。

5. 主な発表論文等  
(研究代表者、研究分担者及び連携研究  
者には下線)

[雑誌論文] (計 2 件)

- ① Sujeet Pradhan, Nimit Pattanasri,  
Two-Dimensional Retrieval in Synchronized  
Media Streams, Proceedings of Database  
Systems for Advanced Applications,  
Springer, 236-240, 2009, 査読あり
- ② Sujeet Pradhan, Towards an integrated  
framework for querying collection of  
heterogeneous data, ACM Publication on  
Ubiquitous Information Management and  
Communication, CD-ROM, 2009, 査読あ  
り

[学会発表] (計 3 件)

- ① Sujeet Pradhan, Nimit Pattanasri,  
Two-Dimensional Retrieval in Synchronized  
Media Streams, Int'l Conf. on Database  
Systems for Advanced Applications,  
Springer, 236-240, 2009, Brisbane, Australia
- ② Sujeet Pradhan, Towards an integrated  
framework for querying collection of  
heterogeneous data, Int'l Con. on Ubiquitous  
Information Management and  
Communication, 2009, Seoul, South Korea
- ③ Sujeet Pradhan, A General Query Model for  
Keyword Queries: Hopes and Challenges,  
iDB2008 Forum, 2008, Kobe, JAPAN

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称 :  
発明者 :  
権利者 :  
種類 :

番号 :  
出願年月日 :  
国内外の別 :

○取得状況 (計◇件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
取得年月日 :  
国内外の別 :

[その他]  
ホームページ等

6. 研究組織

(1) 研究代表者 ( )

研究者番号 :

(2) 研究分担者 ( )

研究者番号 :

(3) 連携研究者 ( )

研究者番号 :