

平成 26 年 5 月 26 日現在

機関番号：35311

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500142

研究課題名(和文)同期化メディアストリームコンテンツにおける高度情報統合・検索手法に関する研究

研究課題名(英文) An Advanced Information Integration and Retrieval Method of Synchronized Multimedia Streams

研究代表者

Pradhan Sujeet R (PRADHAN, SUJEET)

倉敷芸術科学大学・産業科学技術学部・教授

研究者番号：90320001

交付決定額(研究期間全体)：(直接経費) 3,900,000円、(間接経費) 1,170,000円

研究成果の概要(和文)：2次元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に対してキーワード検索手法を用いた場合に、起こり得る問題点を明らかにし、その問題が起きたときに対応できる新たな検索結果モデルを提案した。この検索モデルに必要なとする理論、特に、代数的な演算やそれらの演算における性質を形式的に定義を行い、検索モデルにおける理論上の効率や性能の保証ができる重要な成果が得られた。このデータモデル及び検索モデルの効率の面での性能、有効性について検証を行った。その研究成果をまとめた英語の論文が国際的に有力なジャーナルに掲載された。

研究成果の概要(英文)：Two dimensional retrieval model for searching semantically synchronized media streams has been proposed. The effectiveness of this theoretically-based model has also been verified through rigorous experiments. Conventional IR systems, which support partial retrieval of synchronized media streams, retrieve "atomic units", for example slides, pages and shots of underlying media streams such as presentations, electronic books, and lecture videos respectively. In contrast, this model is based upon the concept of an extended retrieval unit and thus retrieves dynamically integrated media streams comprising several atomic units both along and across media streams. The model outperforms (a) conventional systems based on atomic retrieval units, (b) single dimensional retrieval systems, which extend retrieval units over several atomic units of the same media and (c) cross-media retrieval systems, which extend atomic retrieval units across several media stream units.

研究分野：マルチメディア・データベース

科研費の分科・細目：マルチメディア・データベース

キーワード：代数検索モデル 同期化メディアストリーム 検索単位

1) 研究開始当初の背景

e - Learning や通信教育のような応用において学習教材として利用される講義ビデオ、講義資料、電子教科書等の教育的コンテンツの内容は、密接な相互関連を持っており、各素材を単にそのままの形態で発信するのではなく、互いに有機的に統合することにより、きめが細かく教育効果の高い情報を提供することができ、高い効果が望める。このために、国内外において従来から複数の教育素材を同期し提供する技術、即ち同期化メディアストリームについて様々な視点からの研究が進められている。一方、プレゼンテーション用の資料やプレゼンテーションビデオのような多数のメディアストリームを同期化させ、より効果的なコンテンツを容易に作成可能とす MPMeister「株式会社 RICOH」のようなソフトウェアが開発されつつあり、大企業からも同期化メディアストリームの重要性に注目が集まっている。しかしながら同期化メディアストリームに対するこれまでの研究やシステム開発は、low-level synchronization (複数のメディアを正確に同期化させる) 技術や、low-latency media streaming (同期化させたメディアストリームをネットワークを介して配信する際に待ち時間を少なくする) といった技術に重点が置かれており、同期化メディアストリームにおける特定の部分を容易に検索するというレベルに至っていない。同期化メディアストリームによる情報量は、爆発的に増大していることが明らかであり、今後、それらの情報から所望の部分を容易に検索できる検索手法が必要になってくる。これについて情報検索分野の観点から見ると、次の研究課題があると考えられる。

- a) 同期化メディアストリームにおける多数のコンテンツによる情報を統合し、それらのコンテンツ間における相互関連を表現するのにどんなデータモデルが適切であるか。
- b) 同期化メディアストリームコンテンツにおける情報の検索を、Web 情報の検索手段と同様な手段を用いて検索する場合は、どのような検索結果モデルが適切であるか。

2) 研究の目的

本研究は、同期化メディアストリームにおけるデータモデル・検索手法に必要とされる論理的な基盤を提供し、それを誰もが容易に活用できる同期化メディアストリーム情報検索システムの開発を目的とし、同期化メディアストリームコンテンツの急速に普及が進む時代に向けた先進的な IT 基盤技術の構築を目指すものである。

具体的には、次に示す 3 つの点について考察または提案を行う。

- a) 同期化メディアストリームにおけるコンテンツ間のセマンティックマッ

ピングによる情報統合法を提案し、これを表現可能とする 2 次元データモデルを提供する。

- b) 同期化メディアストリームの部分的な検索における新たな検索結果モデルを提案し、それに必要な理論的な基礎、即ち代数的な演算やそれらの演算における性質を提供する。また、このモデルでは、次のことを考慮する。(i) 同一メディアストリーム内におけるコンテキスト(ii) 異なるメディアストリーム間におけるコンテキスト
- c) 問い合わせ処理の論理的な最適化手法の提案を行う。

3) 研究の方法

- a) 2 次元データモデル：研究期間の初年度は、同期化メディアストリームにおける多種多様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。このモデルの特徴は、同期化メディアストリームにおける多数のメディア間の相対関連情報、また各メディアにおける時系列情報が 2 次元空間上で簡単に表現できるという点である。
- b) 代数的な検索結果モデル：次に、このデータモデルに基づいて構築される同期化メディアストリームにおける検索手法としては、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ることにした。しかし、この検索手法では、「情報全体を表わす 2 次元空間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、同期化メディアストリームにおける最も適切なコンテンツ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の提案を行った。具体的には、キーワードのみの入力検索に対して、最も意味があると考えられる部分とは何かを形式的に定義し、さらに、その定義による解(コンテンツ部分)を求めるために、代数的な演算の利用可能な処理法も提案した。データベース分野では、このような処理法が以前から使われていて、その理由は、数式の加工に代数的なさまざまな演算が利用できるため柔軟でありながら、実装環境を意識せず、効率良く処理するのに最も必要な機能として知られている「論理最適化」が期待できる利点があるからである。
- c) コンテンツデータベースの構築：上

記に述べたデータモデルや検索モデルを用いた試作システムの開発を行い、本手法の有効性に関する検証を行った。これに必要とされる実験用データを準備するとともに本研究の経費を利用し、1台のコンテンツ蓄積サーバを購入した。実験用データとしては、データベースやアルゴリズムといった科目における講義ビデオや電子教科書についてのデータに基づいたコンテンツデータベースを構築した。

- d) 研究成果の公開：これらの研究から得られた成果をまとめた英語の論文が国際的に有力な雑誌に掲載された。

4) 研究成果

2次元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に対してキーワード検索手法を用いた場合に、起こり得る問題点を明らかにし、その問題が起きたときに対応できる新たな検索結果モデルを提案した。その問題点とは、検索キーワードを含む部分的なメディアストリームを検索する際に、検索結果になりうるストリームが切り離せるバウンダリの正式な定義がないという点である。次に、この検索モデルに必要とする理論、特に、代数的な演算やそれらの演算における性質を形式的に定義を行った。これらの定義によって、検索モデルにおける理論上の効率や性能の保証ができるため、重要な成果だと考えられる。また、このデータモデル及び検索モデルの効率の面での性能、有効性について検証を行うために、3種類のテストコレクションを生成し、そのデータに対して実験を行い、モデルの検証を行った。テストコレクションは、実際に使われている講義スライド・講義ビデオや教科書に基づいたデータである。この実験を行うためには、コンテンツサーバ等を購入し、J A V A言語を用いてプロトタイプシステムも作成した。

本研究の目的は、同期化メディアストリームにおける検索モデルの提案、またそのモデルの有効性や効率における検証が主なものであり、その目標を達成できた。また、その研究成果をまとめた英語の論文が国際的に有力なジャーナルに掲載された。

今後、このデータモデルの問題点としては、複数のメディアを2次元空間上の表現したときに、メディアをどの順序で表示するかによって、問い合わせ結果が変わってくる可能性がある。そこで、それぞれメディアの内容自体を分析し、メディア間における相対関係を洗い出し、その結果による最も効果的な順序を自動的に生成されるアルゴリズムが望ましい。しか

し、メディアの内容分析のみにより導いたメディア間の順序が、実験結果によると、それほど効果的ではなかったことが明らかになった。そのため、メディア間の順序を決定法を今後考えていく必要がある。

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 1件)

- 1 Extending information unit across media streams for improving retrieval effectiveness, Nimit Pattanasri, Sujeet Pradhan, Katsumi Tanaka, Data & Knowledge Engineering. Volume 83, January 2013, Pages 70-92

<http://dx.doi.org/10.1016/j.datak.2012.10.003>

[学会発表](計 0件)

[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 0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

(Pradhan Sujeet R)

研究者番号：90320001

(2)研究分担者

()

研究者番号：

(3)連携研究者 ()

研究者番号：