

平成 21 年 5 月 8 日現在

研究種目：特定領域研究
 研究期間：2006～2010
 課題番号：18049041
 研究課題名（和文） 情報爆発に対応するコンテンツ融合と操作環境融合に関する研究
 研究課題名（英文） Contents Fusion and Seamless Search for Information Explosion
 研究代表者 田中 克己
 京都大学・情報学研究科・教授
 研究者番号：00127375

研究分野：データベース・情報検索
 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：情報爆発、WWW、情報検索、情報統合、コンテキスト

1. 研究計画の概要

「情報爆発」に対して、真に必要なコンテンツを獲得し利活用を効果的・効率的に行うためには、種々の情報源からの多様なメディア情報を効果的に統合して利活用・管理する技術が不可欠である。本計画研究の目的は、異種コンテンツの横断的検索・統合や、コンテンツの操作環境と閲覧検索環境の効果的な統合を実現するための基盤的ソフトウェア技術を開発し、情報爆発に対処可能な新しいコンテンツ検索・融合技術を得ることである。このために、異種コンテンツの検索/融合技術、ユーザ・コンテキストの抽出に基づくコンテキストアウェアな情報検索方式、コンテンツの操作環境のシームレス融合技術の研究を実施した。

2. 研究の進捗状況

平成 20 年度までの研究の進捗状況は以下の通りである。

- (1)異種コンテンツの検索・融合技術の研究
 Web と TV 映像コンテンツを統合的にキーワードサーチしてランキングするアルゴリズムを考案し、システムの試作を行い、この結果を WWW2006 にて発表。
 従来のページ単位の検索ではなく、ページ集合を検索単位とした検索が全容検索や複数の事物の比較のための検索に有用であることに着目し、このためのランキング方式を開発し ICADL2006 や情報処理学会論文誌で発表。
 Web からの知識獲得や検索結果の信頼度評価のために、Web 検索エンジンの索引

を用いたオントロジー知識の抽出手法（情報処理学会論文誌等で発表）や、抽出知識を用いた検索結果の信頼度評価方式を開発（ECDL2007 で発表）。

Web 検索エンジンの検索結果を仮想表と見なして既存のデータベースや自然言語処理機能と統合できる Web からの知識獲得プラットフォーム EaRDB や、Office 文書等の様々な属性情報をキーワードで検索できる XML 文書検索方式を開発。

(2)ユーザ・コンテキストの抽出に基づくコンテキストアウェアな情報検索方式の研究

ユーザ位置に依存して検索語の重みを変動させユーザの検索語をインクルメンタルに自動補完する手法を開発（CONTEXT2007 等で発表）。

ユーザインタラクションにより検索精度を改善する手法として、「語ベース適合フィードバック」とそれに基づく検索結果ランキング方式を開発（DEXA2007、情報処理学会論文誌、jDB フォーラム 2008 等で発表し、データベース関連 3 学会合同の優秀若手研究者賞を受賞）。

ソーシャルブックマーク情報を用いて、既存の Web 検索エンジンの精度向上を図る方式（ACM/IEEE JCDL2007 で発表し Vannevar Bush 優秀論文賞にノミネート）Web 画像検索エンジンの精度向上を行う方式（タグ情報を用いて既存の Web 画像検索エンジンの再現率を大幅に向上、WISE2008 国際会議で最優秀論文賞受賞）を開発。

検索エンジンに投入したユーザのクエリログから、同位語や話題語を自動抽出する方式を開発（ACM WI 2008 発表）。

- (3) コンテンツの操作環境のシームレス融合技術の研究
メールから抽出された話題構造にもとづき、メールを補完する情報を自動検索する方式を開発。
ブログマイニングで得た集約情報と地図操作ソフトを統合し、各地点の特徴や訪問者の実際の体験を効果的に提示できる方式を開発(WWW2006, DEXA2006 で発表)。閲覧中の Web ページの内容から自動的に内容を補完するページを検索できる Web ブラウザを開発(WISE2007、ICADL2008 で発表)。
ユーザの複数の Web 検索行動や検索結果の同期化 / 並列化を実現するための手法を開発しその有用性を実証(WISE2008 で発表)。

3. 現在までの達成度

当初の計画以上に進展している。

(理由)

難易度の高い国際会議での論文発表と多数の受賞をしており、国際的にインパクトのある研究成果を出している。また、成果の一部が企業の新しい検索サービスや実証ラボに搭載されるなど、成果の社会への還元も順調に進んでいる。

4. 今後の研究の推進方策

当初計画との大きな変更点は無く、以下の3研究テーマに絞って研究を行う。

- (1) 異種情報の統合検索と知識検索
Web1.0 情報とWeb2.0 情報の相互連携による検索精度や検索結果の品質の向上、Web からのオントロジー知識抽出、Web 中のオブジェクト情報の集約とランキング、オブジェクトの時空間的変遷情報の抽出に関する研究。
- (2) ユーザ適応型情報検索
ユーザ行動からの検索意図や知識レベルの発見、抽出意図に基づく検索質問修正方式やユーザ理解度に基づくランキング、語ベース適合フィードバックのユーザログや検索エンジンのクエリログを用いた検索意図の抽出、ユーザ発信情報に基づく情報検索の研究。
- (3) 情報閲覧・検索・操作ソフトのシームレス融合
補完情報を自動検索する Web ブラウザのクロスメディア拡張、ユーザの「比較」行動を支援する検索技術の研究。

5. 代表的な研究成果

(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24件)

山家雄介, 中村聡史, アダムヤトフト, 田中克己: ソーシャルブックマークの特性分析とそれに基づく Web 検索結果の再ランキング手法, 情報処理学会論文誌(TOD), 1(1), 88-100(2008)

大島 裕明, 小山 聡, 田中 克己: Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見, 情報処理学会論文誌(TOD), 47, SIG19(TOD32), 98-112 (2006)

[学会発表](計24件)

Makoto Kato, Hiroaki Ohshima, Satoshi Oyama, Katsumi Tanaka: Can Social Tagging Improve Web Image Search?, Proc. WISE2008, LNCS, 5175, 235-249, Auckland, New Zealand (Sept. 2nd, 2008) (Best Paper Award)

Yusuke Yanbe, Adam Jatowt, Satoshi Nakamura, Katsumi Tanaka: Can Social Bookmarking Enhance Search in the Web?, Proc. ACM/IEEE JCDL2007, 107-116 Vancouver, Canada (June 20th, 2007) (Nominated for Vannevar Bush Best Paper Award)

Taro Tezuka, Takeshi Kurashima, Katsumi Tanaka: Toward Tighter Integration of Web Search with a Geographic Information System, Proc. WWW2006, 277-286, Edinburgh, Scotland (May 25th, 2006)

[図書](計3件)

[産業財産権]

出願状況(計0件) 取得状況(計0件)

[その他]

(1) 報道

H19 田中 克己, 中村 聡史 “Google の上” で覇権を握る手段はこれだ 最前線の研究者が議論 日経パソコンオンライン 2007/11/30

H19 田中 克己, 中村 聡史, 大島 裕明 次世代検索はグーグルの上に構築 @IT 2007/11/27

(2) ソフトウェア

H19 中村 聡史, 大島 裕明: 次世代サーチ関連のソフトウェアライブラリ整備 <http://www.dl.kuis.kyoto-u.ac.jp/i-exploration/index.html>

H18 大島 裕明, 中村 聡史, 田中 克己: SlothLib, <http://www.dl.kuis.kyoto-u.ac.jp/SlothLibWiki/>