

自己評価報告書

平成23年 3月31日現在

機関番号：14301

研究種目：基礎研究（B）

研究期間：2008～2011

課題番号：20300036

研究課題名（和文） Web上の知識資源の統合利用基盤技術に関する研究

研究課題名（英文） A Study on Fundamental Technologies for Integrated Usage of Knowledge Resources on the Web

研究代表者

吉川 正俊 (YOSHIKAWA MASATOSHI)

京都大学・大学院情報学研究所・教授

研究者番号：30182736

研究分野：データベース，情報検索，Web 情報システム

科研費の分科・細目：メディア情報学・データベース

キーワード：知識資源，Web，マルチメディア，情報抽出，情報検索

1. 研究計画の概要

Web 上で公開され質の保証がされている知識資源を統合的に利用することにより，爆発的に増大し続けている Web 上のデータからより品質の高い情報を利用者の嗜好に応じて容易に探索可能とするための基盤技術について研究を行う。検索エンジン，オントロジデータ，ドメインデータの結果など多様なデータを相互利用することにより，問合せに対して，知識ベースや検索エンジン単独では得ることができない高品質の結果を提示するための基礎研究を行う。

2. 研究の進捗状況

(1) 情報を提供するサーバ側において情報の由来(provenance)など信憑性の判断材料を提供する手法の開発

膨大な履歴情報の系統的な管理を支援するとともに，データ利用者に対して履歴情報の検索，閲覧を可能とする手段を提供する枠組みを開発した。多様な粒度のデータに対して付与された注釈を処理を通して伝播させる手法を提案し，実験により従来手法に対する空間コスト，時間コスト面での優位性を示した。

(2) 大量の RDF データを効率的に管理，検索する技術の開発

RDF のグラフトポロジ，述語の更新可能性，更新頻度，データ統計情報などに基づいたハイブリッド型スキーマを用いる格納方法を考案した。本研究の格納及び検索システムを構築する上では，SPARQL の処理のみならず，Dynamic Concise Bounded Description (DCBD)の構築を考慮した点に特色を持つ。

(3) エンティティ検索のための知識ベースを

利用した Web ページの分類

知識資源を利用した高品質な情報の探索およびその結果の効果的統合と意味的相互補完技術として，Web 検索時の目的の一つであるエンティティ検索のために知識ベースを用いることにより検索 Web ページを分類する新たな手法を提案した。本手法は，Web ページと Wikipedia 記事の類似度を計算し，YAGO を参照しながら Web ページ，エンティティ，カテゴリの間の有向グラフである PFC グラフを構築する。PFC グラフを参照することにより Web ページをカテゴリに分類する。実験により本手法の有効性を確認した。

(4) 利用者生成型事典を利用した情報検索に関する研究

Wikipedia のような Web 上の利用者生成型百科事典に対して文書間リンク構造を減衰流を用いて解析することにより，概念や事物間の関連の強度を求める手法を開発した。また，始点，終点間の経路に出現する関連を裏付ける情報を利用することにより，二つの事項に関連のある画像を精度良く検索するための手法を開発した。さらに，Wikipedia の記事とその記事における画像のキャプションの整合性，および記事タイトル，画像キャプションに対する画像の典型性，代表性を計算する手法を開発した。また，Wikipedia の編集履歴情報を利用し編集者の信頼性を計算することに

より Wikipedia 記事の信頼性を推定する手法を開発し，その手法に基づき実際に提示されている Wikipedia 記事における各部分文書の信頼度を利用者に提示するシステムを開発した。

3. 現在までの達成度

②おおむね順調に進展している。

当初目的とした,

* 情報を提供するサーバ側において情報の由来(provenance)など信憑性の判断材料を提供する手法や秘匿すべきプライバシー情報を含むコンテンツを検出する技術

* 大量の RDF データを効率的に管理, 検索する技術

* 知識資源を利用した高品質な情報の探索およびその結果の効果的統合と意味的相互補完技術, さらに, 獲得した情報を下にサーバ, クライアント双方で協調的に知識資源を拡張する技術

のうち, プライバシ情報検出技術については要素技術である人物名検索にとどまっているが, それ以外については, ほぼ達成できている。

4. 今後の研究の推進方策

当初の研究計画では, 初期知識資源としての YAGO の利用を行ったが, その後 YAGO 自身が独自の開発過程を辿ったため, Web 上のより多様な資源を対象として研究を進めている。データ品質のうち信憑性情報に関しては, 利用者生成型コンテンツ (User Generated Contents) やニュースなど多様な資源の統合利用へと展開していく予定である。また, 直列化された RDF 表現である XML の検索に関しては, 単なる検索のみではなく, 利用者支援の観点から研究を進める予定である。

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 5 件)

- ① Shin Ishida, Qiang Ma, Masatoshi Yoshikawa, "Extraction of Characteristic Description for Analyzing News Agencies", Journal of Digital Information Management, 査読有, Volume 8, Issue 6, pp. 349-354, 2010.
- ② 青戸 了, 清水 敏之, 増田 耕一, 吉川 正俊 "履歴管理システムにおける多粒度アノテーション伝播", 日本データベース学会論文誌, 査読有, Vol. 9, No. 1, pp. 64-69, 2010 年.
- ③ Takeharu Eda, Masatoshi Yoshikawa, Toshio Uchiyama and Tadasu Uchiyama, "The Effectiveness of Latent Semantic Analysis for Building Up a Bottom-up Taxonomy from Folksonomy Tags", World Wide Web Journal, 査読有, Volume 12, Number 4, pp. 421-440, 2009 年.
- ④ Umaporn Supasitthimethee, Toshiyuki

Shimizu, Masatoshi Yoshikawa and Kriengkrai Porkaew, "XSemantic: An Extension of LCA based XML Semantic Search", IEICE Transactions on Information and Systems, 査読有, Vol. E92-D, No. 5, pp. 1079-1092, 2009.

- ⑤ 米井 由美, 岩井原 瑞穂, 吉川 正俊, "XML 文書における構造の素性を用いた照応による人物検索", 日本データベース学会論文誌, 査読有, Vol. 7, No. 1, pp. 151-156, 2008 年.

[学会発表] (計 22 件)

- ① Xinpeng Zhang, Yasuhito Asano, Masatoshi Yoshikawa, "Enishi: Searching Knowledge about Relations by Complementarily Utilizing Wikipedia and the Web", The 11th International Conference on Web Information System Engineering (WISE), pp. 480-495, December 12-14, 2010, Hong Kong, China.
- ② Ling Xu, Qiang Ma, Masatoshi Yoshikawa, "Exploring Special Items in Multimedia News Based on a Stakeholder Model", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT2010), pp. 452-455, Aug.-Sep. 2010, Toronto Canada.
- ③ Yusuke Kiritani, Qiang Ma, and Masatoshi Yoshikawa, "Classifying Web Pages by Using Knowledge Bases for Entity Retrieval", Proc. 20th Int. Conf. Database and Expert Systems Applications (DEXA2009), LNCS 5690, pp. 761-768, Aug.-Sep. 2009, Linz, Austria.
- ④ Xinpeng Zhang, Yasuhito Asano and Masatoshi Yoshikawa, "Visualized Elucidations of Ranking by Exploiting Object Relations", Third International Workshop on Ranking in Databases (DBRank 2009), In Conjunction with ICDE 2009, March 29, 2009, Shanghai, China.
- ⑤ Xinpeng Zhang and Masatoshi Yoshikawa, "Querying RDF Data Using Dynamic Concise Bounded Description", Proceedings of the International Workshop on Information-Explosion and Next Generation Search 2008 (INGS2008), IEEE Computer Society, pp. 39-46, April 26-27, 2008, Shenyang, China.