

平成 22 年 6 月 1 日現在

研究種目： 基盤研究(B)  
 研究期間： 2007 ～ 2009  
 課題番号： 19300027  
 研究課題名（和文） 品質駆動型の情報源統合のための知識発見・獲得手法に関する研究  
 研究課題名（英文） Knowledge Discovery and Mining for Quality-driven  
 Information Integration

## 研究代表者

石川 佳治 (ISHIKAWA YOSHIHARU)  
 名古屋大学・情報基盤センター・教授  
 研究者番号： 80263440

研究成果の概要（和文）：本課題では、ネットワーク上に存在する大量のデータを有効に活用するため、品質に着目した情報源の統合について研究を行った。特に、P2P (peer-to-peer) ネットワークにおける情報の流通のためのトレーサブルなレコード交換のフレームワーク、ネットワーク上に存在するテキスト情報の集約および情報抽出、集約した情報を適切にユーザに提示するための配信技術の開発などを行った。

研究成果の概要（英文）：In this research, we studied on quality-driven information integration for effectively utilizing a large amount of data in the network. In particular, we developed a traceable record exchange framework for reliable information exchange in a P2P (peer-to-peer) network. In addition, we performed studies on information aggregation and information extraction from text data existing in the network and information delivery methods for appropriate information presentation to the users.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	5,400,000	1,620,000	7,020,000
2008 年度	4,400,000	1,320,000	5,720,000
2009 年度	4,700,000	1,410,000	6,110,000
年度			
年度			
総計	14,500,000	4,350,000	18,850,000

## 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：情報統合、情報の品質、情報の信頼性、テキストデータベース、文書クラスタリング、情報抽出、P2P ネットワーク、トレーサビリティ

## 1. 研究開始当初の背景

- (1) インターネット上には膨大かつ多種多様な情報源が存在しており、それらの有効活用が求められている。本研究のキーワードの一つが「信頼性」である。ネットワーク上で信頼ある情報の交換や流

通を行うためには、交換された情報の出所が後になって追跡できるトレーサビリティが重要である。しかし、従来の研究では、この問題はあまり取り組まれていなかった。

- (2) 別の観点として、膨大で不均質なネット

- ワークの情報源から、質の高い情報をどのように抽出するかという問題がある。大量の情報に対しても動作するスケラブルな手法の実現が求められている。
- (3) また、ネットワーク環境において、統合した情報をどのようにユーザに効果的に提供（配信）するかにも課題が存在する。ユーザの嗜好や状況に応じた適応的な配信技術が求められる。

## 2. 研究の目的

- (1) 上記背景(1)に関して、特に P2P (peer-to-peer) ネットワークにおける情報交換技術について研究を行う。P2P ネットワークは柔軟で自律性に富むネットワーク技術であるが、そこでの情報交換には非合法であったり、信頼性に欠ける情報であったりなど、数々の問題点が指摘されている。
- (2) 上記背景(2)に関しては、特にインターネット上に存在するテキスト情報源に着目する。テキスト情報を大量に含むウェブサイトからのユーザの要求に合った情報を抽出することや、インターネット上で配信されるニュースなどの情報を集約する手法を開発する。
- (3) 上記背景(3)に関しては、モバイル環境において移動を続けるユーザに対して適切な情報を送付する技術や、データベースに新たな情報が得られた際に、その差分情報を知らせる技術の開発などを行う。

## 3. 研究の方法

- (1) 目的(1)に関して、P2P ネットワーク環境におけるトレーサブルなレコード交換フレームワークの概念を提案した。P2P ネットワーク上に存在するピアがレコードを交換し、価値ある情報の蓄積を行うというものである。その特徴は、各ピアが保持しているレコードに対し、それがどのように得られたかを追跡できるトレーサビリティの機能が支援されていることである。トレース処理のための問合せ言語の開発や、問合せ処理実行方式の開発などを中心に研究を行う。
- (2) 目的(2)に関して、以下の研究を行う。
- ① 大規模テキスト情報源からのトピックに応じたレコード抽出：大量のテキストデータ（例：ニュース記事）の中から、たとえば企業名とその経営者名の組（レコード）を抽出する手法を開発する。ユーザが意図したトピック（例：コンピュータ関係の分野の企業）に着目して、ユーザが提供したサンプルデータをもとに抽出処理を学習する。

- ② ネットワーク上を配信されるテキストデータのクラスタリング：文書データのクラスタリングは、大量のテキストデータをグループ化し、大まかなサマリを構築するのに有用な手法である。本研究では、時々刻々と得られるテキストデータに対応するため、インクリメンタルなクラスタリング手法を開発する。また、ニュース記事等では新規の情報に価値が高いということを考慮し、新規性を重視するアプローチをとる。
- (3) 目的(3)に関しては、以下の研究を行う。
- ① モバイル環境におけるオンデマンド型の情報配信技術：携帯端末を保持して移動するユーザに対し、情報提供を行う場合、ユーザの状況や好みを意識することが重要となる。そのため、本研究では、モバイル P2P ネットワークにおいて適切な情報（一種の広告）をタイムリーに配信する手法を開発する。
- ② テキスト主体のデータベースからの選択的な情報配信：今日ではリレショナルデータベース上に大量のテキストデータが蓄積されていることも多い。本研究では、データベースに対して更新（主として新たなエントリの追加）が発生したときに、情報の差分を効果的に配信するための問合せ処理について研究する。ユーザはキーワード集合などで、必要とする情報を問合せとして指示しておく。新たなデータが追加されたとき、マッチする情報があればユーザに新たに配信される。

## 4. 研究成果

- (1) 研究(1)については、現時点で最も新しく包括的な成果は、下記雑誌論文[1]に集約されている。トレーサブルな P2P レコード交換システムにおける問合せ処理手法を扱ったものである。この前段階となる実績としては、コンセプトを提案した雑誌論文[7]および学会発表[10, 16]、イベント駆動型問合せの部分に着目した雑誌論文[2]がある。現在はその発展である複製を用いるアプローチについて研究を進めている（学会発表[1]）。
- (2) 研究(2)については、以下のようになる。
- ① 研究課題①については、雑誌論文[13]が最終的な実績であり、手法に関する有効性の検証などを詳細に行っている。これは、学会発表[23]の拡張版である。
- ② 研究課題②については、文書クラス

タリング手法に関する評価を行った包括的な成果が雑誌論文[9]にある。また、その発展として、クラスタの変遷を追跡する手法を雑誌論文[3]で示した。これは学会発表[15]の発展版である。また、クラスタリング結果の変遷を可視化する手法についても研究を行った。最終的な成果は雑誌論文[12]であるが、その全段階の結果は、雑誌論文[14]および学会発表[22]にある。

(3) 研究(3)については以下のような成果を得た。

① 研究課題①については、学会発表[13, 17]が最新の成果であり、継続して研究を進めている。

② 研究課題②については、学会発表[2, 11]が成果である。新たなアプローチとして、十分評価に値する研究であると考えている。

#### 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 14 件)

1. Fengrong Li, Yoshiharu Ishikawa, Query Processing in a Traceable P2P Record Exchange Framework, IEICE Transactions on Information and Systems, Vol. E93, No. 6, pp. 1433-1446, 2010 年, 査読有.
2. 飯田卓也, 李峰栄, 石川佳治, トレーサブルなP2Pレコード交換システムにおけるイベント駆動型問合せ, 日本データベース学会論文誌, Vol. 8, No. 1, pp. 95-100, 2009 年, 査読有.
3. Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Kitagawa, Querying Topic Evolution in Time Series Document Clusters, 日本データベース学会論文誌, Vol. 7, No. 3, pp. 7-12, 2008 年, 査読有.
4. 飯田卓也, 李峰栄, 石川佳治, トレーサブルなP2Pレコード交換システムPI-REXの設計, Vol. 2008, No. 88, pp. 289-294, 2008 年, 査読無.
5. 李峰栄, 飯田卓也, 石川佳治, トレーサブルなP2Pレコード交換システムにおける問合せ処理の効率化について, 情報処理学会研究報告, Vol. 2008, No. 88, pp. 97-102, 2008 年, 査読無.
6. Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Kitagawa, Querying Time-Series Document Clusters, 情報処理学会研究報告, Vol. 2008, No. 88, pp. 61-66, 2008 年, 査読無.
7. Fengrong Li, Takuya Iida, Yoshiharu Ishikawa, Traceable P2P Record Exchange: A Database-Oriented Approach, Frontiers of Computer Science in China, Vol. 2, No. 3, pp. 257-267, 2008, 査読有.
8. 李峰栄, 飯田卓也, 石川佳治, トレーサブルなP2Pレコード交換システムにおける問合せ処理, 情報処理学会研究報告, Vol. 2008, No. 56, 2008 年, 査読無.
9. Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Kitagawa, A Novelty-based Clustering Method for On-line Documents, World Wide Web Journal, Vol. 11, No. 1, pp. 1-37, 2008, 査読有.
10. 李峰栄, 石川佳治, トレーサブルなP2P情報流通のためのデータモデルの提案, 電子情報通信学会研究報告, Vol. 107, No. 131, pp. 461-466, 2007 年, 査読無.
11. Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Ishikawa, A Document Clustering Method Focusing on User Specified Time of Interest, 電子情報通信学会研究報告, Vol. 107, No. 131, pp. 411-416, 2007 年, 査読無.
12. 長谷川幹根, 石川佳治, T-Scroll: 時系列文書のクラスタリングに基づくトレンド可視化システム, 情報処理学会論文誌: データベース, Vol. 48, No. SIG 20(TOD 36), pp. 61-78, 2007 年, 査読有.
13. 張建偉, 石川佳治, 北川博之, トピックを考慮した大規模文書情報源からのレコード抽出, 情報処理学会論文誌: データベース, Vol. 48, No. SIG 14(TOD 35), pp. 107-123, 2007 年, 査読有.
14. 長谷川幹根, 石川佳治, T-Scroll: 時間的トピックの推移をとらえる可視化システム, 日本データベース学会Letters, Vol. 6, No. 1, pp. 149-152, 2007 年, 査読有.

[学会発表] (計 23 件)

1. Fengrong Li and Yoshiharu Ishikawa, Query Processing with Materialized Views in a Traceable P2P Record Exchange Framework, The 2nd Intl. Workshop on Web-based Contents Management Technologies (WCMT 2010), Jiuzhaigou, China, July 15, 2010, 査読有. (to appear)
2. Yanwei Xu, Yoshiharu Ishikawa, Jihong Guan, Efficient Continuous Top-k Keyword Search in Relational Databases, Proc. 11th Intl. Conf. on Web-Age Information Management (WAIM 2010), Jiuzhaigou, China, July 17, 2010. (to appear)

3. Yoshiharu Ishikawa, Traceable P2P Record Exchange, The 5th Korea-Japan Database Workshop, Jeju, Korea, May 29, 2010. (invited talk)
4. 飯田卓也, 石川佳治, モバイルP2P環境におけるオンデマンド型情報配信, 情報処理学会第72回全国大会, 6ZP-7, 2010年3月11日, 東京.
5. 李峰榮, 石川佳治, トレーサブルなP2Pレコード交換システムにおける実体化ビューの活用, 第2回データ工学と情報マネジメントに関するフォーラム (DEIM 2010), 2010年3月1日, 淡路島.
6. 許延偉, 石川佳治, リレーショナルデータベースにおける連続的のトップkキーワード検索手法, 第2回データ工学と情報マネジメントに関するフォーラム (DEIM 2010), 2010年3月1日, 淡路島.
7. 飯田卓也, 石川佳治, モバイルP2P環境におけるユーザの趣向を考慮した情報配信手法の提案, 第2回データ工学と情報マネジメントに関するフォーラム (DEIM 2010), 2010年2月28日, 淡路島.
8. 飯田卓也, 石川佳治, P2P環境におけるPublish/Subscribeシステムについて, 電子情報通信学会研究報告, Vol.109, No.153, pp.55-60, 2009年7月28日, 神戸.
9. Yanwei Xu and Yoshiharu Ishikawa, Efficient Incremental Top-k Keyword Search in Relational Databases, 電子情報通信学会研究報告, Vol.109, No.153, pp.1-6, 2009年7月28日, 神戸.
10. Fengrong Li, Takuya Iida, Yoshiharu Ishikawa, 'Pay-as-you-go' Processing for Tracing Queries in a P2P Record Exchange System, Proc. 14th Intl. Conf. on Database Systems for Advanced Applications (DASFAA 2009), pp.323-327, Brisbane, Australia, April 22, 2009, 査読有.
11. Yanwei Xu, Yoshiharu Ishikawa, Jihong Guan, Effective Top-k Keyword Search in Relational Database Considering Query Semantics, Proc. Intl. Workshop on DataBase and Information Retrieval and Aspects in Evaluating Holistic Quality of Ontology-Based Information Retrieval (DBIR-ENQOIR 2009), pp.172-184, Suzhou, China, April 2, 2009, 査読有.
12. 堀田孝司, 飯田卓也, 李峰榮, 石川佳治, Erlangを用いたP2Pレコード交換システムのためのシミュレータ構築, 情報処理学会第71回全国大会, 5C-1, 2009年3月11日, 滋賀県草津市.
13. 飯田卓也, 李峰榮, 石川佳治, トレーサブルなP2Pレコード交換システムにおけるイベント駆動型問合せ, 第1回データ工学と情報マネジメントに関するフォーラム (DEIM 2009), 2009年3月10日, 静岡県掛川市.
14. 李峰榮, 飯田卓也, 石川佳治, トレーサブルなP2Pレコード交換システムにおける物理構成方式に関する検討, 第1回データ工学と情報マネジメントに関するフォーラム (DEIM 2009), 2009年3月8日, 静岡県掛川市.
15. Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Kitagawa, A Query Language and Its Processing for Time-Series Document Clusters, Proc. 11th Intl. Conf. on Asia-Pacific Digital Libraries (ICADL 2008), pp.82-92, Bali, Indonesia, December 4, 2008, 査読有.
16. Fengrong Li, Yoshiharu Ishikawa, Traceable P2P Record Exchange Based on Database Technologies, Proc. 10th Asia Pacific Web Conf. (APWeb 2008), pp.475-486, Shenyang, China, April 27, 2008, 査読有.
17. 飯田卓也, 李峰榮, 石川佳治, トレーサビリティ機構を有するP2Pレコード交換システムの開発, 情報処理学会第70回全国大会, 5ZK-8, 2008年3月15日, 茨城県つくば市.
18. Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Kitagawa, Generating High Level Descriptions from Document Cluster Transitions, 情報処理学会第70回全国大会, 5A-1, 2008年3月15日, 茨城県つくば市.
19. Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Kitagawa, Summarizing Document Cluster Transitions Using Descriptive Sentences, 電子情報通信学会第19回データ工学ワークショップ (DEWS 2008), 2008年3月9日, 宮崎県宮崎市.
20. 李峰榮, 飯田卓也, 石川佳治, P2Pネットワークにおけるトレーサビリティを有するレコード交換システム機構, 電子情報通信学会第19回データ工学ワークショップ (DEWS 2008), 2008年3月9日, 宮崎県宮崎市.
21. 李峰榮, 飯田卓也, 石川佳治, トレーサビリティ機能を有するP2Pレコード交換機構, データベースとウェブ情報システムに関するシンポジウム (DBWeb 2007), 2007年11月27日 (ポスター発表), 東京.
22. Yoshiharu Ishikawa, Mikine Hasegawa,

T-Scroll: Visualizing Trend in a Time-series of Documents for Interactive User Exploration, Proc. 11th European Conf. on Research and Advanced Technology for Digital Libraries (ECDL 2007), pp.235-246, Budapest, Hungary, September 18, 2007, 査読有.

23. Jianwei Zhang, Yoshiharu Ishikawa, Hiroyuki Kitagawa, Record Extraction Based on User Feedback and Document Selection, Proc. Joint Conf. on 9th Asia-Pacific Web Conference and 8th Intl. Conf. on Web-Age Information Management (APWeb/WAIM 2007), pp. 575-585, Huang Shan, China, June 18 2007, 査読有.

## 6. 研究組織

### (1) 研究代表者

石川 佳治 (ISHIKAWA YOSHIHARU)  
名古屋大学・情報基盤センター・教授  
研究者番号：80263440

### (2) 研究分担者

該当なし

### (3) 連携研究者

該当なし