

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月17日現在

機関番号：32689

研究種目：基盤研究（B）

研究期間：2009～2012

課題番号：21300034

研究課題名（和文） ソーシャルコンテンツの発展過程の分析と応用

研究課題名（英文） Analysis and Applications of Development of Social Contents

## 研究代表者

岩井原 瑞穂 (IWAHARA MIZUHO)

早稲田大学・理工学術院・教授

研究者番号：40253538

## 研究成果の概要（和文）：

本研究はソーシャルネットワークサービス（SNS）やWikiに代表されるソーシャルコンテンツから有用な情報を抽出する技術の開発を目的としている。wiki型コンテンツは多人数が1つの記事を更新することにより、バージョンが蓄積されるが、その派生過程を正確に求める手法を開発した。またSNSにおいて利用者が行うプライバシー設定の傾向を分析し、適切な設定を推薦する手法を開発した。さらにコンテンツのアクセス制御について効率的な手法の開発を行った。

## 研究成果の概要（英文）：

The aim of this project is to develop technologies for extracting useful information from social contents. Wiki-style contents are edited by many contributors. We developed an algorithm to accurately reconstruct derivation of contents from edit histories. We also analyzed tendencies of users' privacy settings, and utilized these findings in recommending appropriate settings to users. We further developed efficient access control methods for contents.

## 交付決定額

（金額単位：円）

	直接経費	間接経費	合計
21年度	4,600,000	1,380,000	5,980,000
22年度	3,200,000	960,000	4,160,000
23年度	2,800,000	840,000	3,640,000
24年度	2,900,000	870,000	3,770,000
年度			
総計	13,500,000	4,050,000	17,550,000

## 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：データベース，コンテンツ，ソーシャルメディア，構造化文書，セキュリティ，プライバシー

## 1. 研究開始当初の背景

近年急増している，個人同士のインタラク  
ションを伴い成長してゆくソーシャルコン

テンツに着目し，その発展過程を分析するこ  
とにより個人間の関わりにより生じる情報  
を集約し有効活用すること，およびプライバ

シーを考慮した安全なソーシャルコンテンツ利用技術を本研究の主題とする。ソーシャルコンテンツはブログやソーシャルネットワークサービス(SNS), wikiに代表される共同編集型コンテンツなどがあり, いずれも大量の情報蓄積がなされ, 重要な情報基盤となりつつある。本研究では特に SNS と共同編集型コンテンツの二形態に着目する。SNS は利用者が指定した Friend リンクを元に, 新たな知人・友人の発見を助けるものであるが, 一般利用者の他にも社内や業種, 医療など目的に特化した SNS が数多く設置されつつある。一方, 共同編集型コンテンツは多人数が1つの記事を更新し, 最新のバージョンが保たれるとともに, 過去のバージョンがすべて保存され, 利用者が互いに差分を確認したり, 更新を取り消して過去のバージョンに戻す操作が可能になっている。大規模な共同編集型コンテンツの例としてオンライン百科事典の Wikipedia があり, 他にも多様なコンテンツが構築され, 数千を超えるバージョンを有しているコンテンツも多い。本研究では, ソーシャルコンテンツの利活用に関するテーマについて研究を行う。

## 2. 研究の目的

### (1) 共同編集型ソーシャルコンテンツの履歴分析

ソーシャルメディアはブログやソーシャルネットワークサービス(SNS), wikiに代表される共同編集型コンテンツなどがあり, いずれも大量の情報が蓄積され, 重要な情報基盤となりつつある。本研究はソーシャルコンテンツから有用な情報を引き出すための一連の技術の開発を目的とする。共同編集型コンテンツは多人数が1つの記事を更新することによりバー

ジョンが蓄積される。共同編集型ソーシャルコンテンツの履歴分析として, 共同編集により生じる利用者間のインタラクションを分析し, 類型化を行う手法を開発する。

また記事の正確性および信頼性を編集履歴から定量的に評価する手法の開発を行う。さらに共同編集型コンテンツの分析を効率的に行うためのアルゴリズムの開発を行う。

### (2) 情報共有のためのソーシャルグラフのリスク分析

SNSでは, 利用者のプロフィールとして自己紹介や関心事を保持する属性リストが登録されている。SNSのプライバシー保護機能として, これらの属性や写真等の個人データを一般に公開したり, あるいは1つまたは2つ以内のFriendリンクでつながれた利用者にものみ公開するという機能がある。またさらにきめ細かい開示設定を行うために, Friendをいくつかのグループに分け, グループ単位で開示設定が可能なSNSもある。さらに管理者を置いたコミュニティを設置して, コミュニティにのみ開示可能という設定も可能である。プライバシー機能への要求からこのようにSNSの情報開示設定は高機能化が進んでいるが, 一方で煩雑化しているともいえ, プライバシー設定に注意を払わないFriendから漏洩する恐れがある。本テーマでは, まず利用者のプライバシー設定傾向を分析し, そのデータを基に, 適切なプライバシー設定を支援する手法を開発する。

### (3) ソーシャルコンテンツ利用のための技術開発

さらにソーシャルコンテンツを安全に利用するためのセキュリティ技

術として、ソーシャルグラフの匿名化、ソーシャルグラフにおける情報共有モデル、共同編集型コンテンツにおける効率的なアクセス制御の各手法の開発を行う。

### 3. 研究の方法

#### (1) 共同編集型ソーシャルコンテンツの履歴分析

テキストベースのソーシャルコンテンツに対する insert や delete, append などの編集操作により、新しいバージョンが生成される。Wikipedia に代表される wiki 型共同編集コンテンツでは、ある利用者の書き込みを他の利用者が書き換えることが行われるため、個々の更新において利用者間のインタラクションが発生する。他者による記述の修正、補足、追加、置き換えといった基本操作を識別することにより、コンテンツの成長の過程を特徴づけることができる。しかしこれらの基本操作は明示的に利用者が与えることはなく、投稿されたバージョンと前後のバージョンの差分を取る必要がある。ふたつのバージョンの差分を取っても、人間にとって意味のある差分が求められるとは限らず、また2つのバージョンの差分を取る操作をバージョン集合にすると、時間がかかりすぎる短所がある。利用者の共同編集作業の過程を明らかにできるような、バージョンの派生過程を識別可能であり、かつ大規模なバージョン集合にも対応できる効率的な処理方法を開発する。

さらに Twitter を対象として、短時間で高頻度に行われる書き込みを対象として、書き込みの盛り上がりの特徴づける分析手法の開発を行う。

### 4. 研究成果

#### (1) Twitter の時系列分析

Twitter に代表される実時間ソーシャルメディアでは、出来事の時間経過を通じて多数の利用者が体験や雰囲気共有できる特性がある。まずサッカーの世界カップで記録された Twitter の書き込みを分析し、試合の盛り上がりやゴールなどの特徴的イベントに伴って現れる書き込みから容易に判別できる特徴を求め、これから利用者全体の盛り上がりやチームの支持率として数値化し、実時間で提示する方法について開発し、試合の状況によく一致して支持率が変化する結果が得られた。

#### (2) Wikipedia の編集プロセス

Wikipedia に代表される共同編集型コンテンツについて、複数の筆者が1つの記事を作成する過程において、筆者間の協調行動が記事の品質に与える影響について分析を行った。Wikipedia の記事の編集履歴を分析することにより、筆者が意見の対立するグループに分割されているかや、編集が友好的に行われているかを定量化する。そして、優秀として推薦されている記事群とそうでない記事群それぞれとの相関を学習することにより、与えられた記事の編集履歴からその記事の品質を推定する方法を開発した。従来の編集回数等による品質の指標に加え、参加している編集者間の関係をネットワーク構造としてモデル化し、対立的な編集行動を行っている割合を示す *restoreratio* という指標を求めた。この新たな指標と既存の指標を組み合わせ、記事の品質を評価することができる。実験結果により従来の手法よりも精度の良い評価結果が得られた。

#### (3) Wikipedia の編集履歴の再構築

Wikipedia の編集履歴は、バージョンの投

稿者と日付、注釈等の基本情報のみからなり、編集者は差分を繰り返し求めて依存関係を把握する必要があった。バージョンが直前のものから作成されたのか、あるいは手戻りや統合が起きたのかといった依存関係・意味的关系を表すバージョングラフの再構成が必要であるが、(a)従来の文書の差分アルゴリズムは2つの文書間であるためバージョングラフの再構成には計算コストがかかりすぎることで、(b)単に編集距離が最短のものを親バージョンとしても正しい結果が得られず、また人間にとって理解の難しい不自然な差分を生成するという問題点がある。これまで編集履歴の可視化や、編集過程での協調行動の分析、編集回数等を用いた記事の品質評価などの研究があるが、いずれも正確なバージョングラフに基づいていない。

本研究では、Wikipediaに蓄積されている記事の過去のバージョンから、バージョン間の派生関係を有向グラフとして復元する問題について、n-gramの頻度分布を比較するという効率的な手法を開発した。さらに、常に連続紙で出現するn-gramどうしをひとつにまとめる操作により、さらに精度を改善できることを示している。

#### (4) 構造化文書のアクセス制御

SNSではコンテンツの配信に構造化文書が用いられることが多いが、ユーザのプライバシー設定に応じた細粒度のアクセス制御が行われる必要がある。クラウドコンピューティングでは、アクセス制御ルールの適用は複数のサーバにまたがって行われる必要がある。代表的な構造化文書であるXML文書に対するアクセス制御ルールの適用を効率化するために、XML文書を格納する関係データベースにおいて、アクセス制御ルールの評価結果を

キャッシュする方式を提案した。本方式ではXML文書が一部更新されても、更新されていない部分の評価結果は再利用できるという効率化を行った。

#### (5) ロールベースアクセス制御

SNSでは、プライバシー設定に基づく細粒度のアクセス制御が必要であり、アクセス制御の定義には利用者をロールとして抽象化するロールベースアクセス制御が必要である。企業内の大規模分散環境において、ロールアクセス制御ポリシーを効率的に実現するには、大域的なポリシーを分散したノードにおいて実装する必要がある。異なる要求事項を持つ分散ノードそれぞれに適したロールアクセス制御ルールをモデル駆動で設計する手法を開発した。

#### (6) SNSにおけるプライバシー設定支援

ソーシャルネットワークサービス(SNS)は交流を活発にするソーシャルメディアとして急速に普及しているが、情報が拡散することにより思わぬ範囲までプライバシー情報が広まってしまう恐れがある。ほとんどのSNSは誰がオンラインコンテンツにアクセスできるかのプライバシー設定を提供しているが、機能は単純化されているか複雑すぎるかに分かれがちである。申請者らはこれまでにSNSユーザが自分に適した設定を行うための、プライバシー設定支援システムを開発している(SocInfo2011)。Facebookの実ユーザ66,000人のプライバシー設定のうちどの属性を一般に公開しているかの情報と、実験に参加したユーザが提供した設定をもとに、プライバシースコアと呼ばれる、設定が開放的かどうかの指標を求める。これに加え、属性間の開示の共起確率による推薦や、スコアの分布の可

視化などを行っている。さらに大規模実ユーザの分析では、性別や英語など言語間で公開傾向に大きな差が見られないことや、友人数と相関を持つ属性と持たない属性に分けられること、ユーザの公開度と活動度(どれくらい活発に投稿等を行っているかを数値化したもの)からユーザをいくつかの類型に分類できることなど、興味深い傾向が観察された。

これらの成果に加え、属性間の開示の共起確率による推薦や、スコアの分布の可視化など、有用な情報を提示することにより、ユーザのプライバシー設定を支援する手法を開発した。

#### 5. 主な発表論文等

[雑誌論文] (計 8 件)

1. Jianmin Wu, Mizuho Iwaihara, "Wikipedia revision graph extraction based on n-gram cover", Proc. Int. Workshop on Graph Data Management and Mining, WAIM 2012 Workshops, Lecture Note in Computer Science 7419, pp. 29-38, 2012 (査読有り) .
2. Toshikazu Munemasa and Mizuho Iwaihara, "Trend Analysis and Recommendation of Users' Privacy Settings on Social Networking Services," 3rd Int. Conf. Social Informatics (SocInfo2011), Lecture Note in Computer Science 6984, pp. 184-197, Oct. 2011 (査読あり).
3. 近藤誠一, 岩井原瑞穂, 吉川正俊, 小宮崇, 虎渡昌史, "異種分散環境におけるロールベースアクセス制御のモデル駆動設計手法," 情報処理学会論文誌 Vol. 52, No. 5, pp. 1882-1898, 2011年5月 (査読あり).
4. Jun Huang, Mizuho Iwaihara, "Realtime Social Sensing of Support Rate for Microblogging," Proc. 2nd Int. Workshop on Social Networks and Social Media Mining on the Web (SNSMW), Lecture Note in Computer Science 6637, pp. 357-368, April 2011 (査読有り) .
5. Se Wang, Mizuho Iwaihara, "Quality

Evaluation of Wikipedia Articles Through Edit History and Editor Groups," Proc. 13th Asia-Pacific Web Conference (APWeb 2011), Lecture Note in Computer Science 6612, pp. 188-199, April 2011 (査読有り) .

6. Erwin Leonardi, Sourav S. Bhowmick, and Mizuho Iwaihara, "Efficient Database-Driven Evaluation of Security Clearance for Federated Access Control of Dynamic XML Documents," Proc. Database Systems for Advanced Applications (DASFAA2010), Lecture Note in Computer Science 5981, pp. 299-306, April 2010 (査読あり).
7. 近藤 誠一, 岩井原 瑞穂, 吉川 正俊, 虎渡昌史, "異種分散環境におけるロールベースアクセス制御の定量的リスク評価," 情報処理学会論文誌, Vol. 50, No. 11, pp.2727-2739, Nov. 2009 (査読あり).
8. Tetsutaro Motomura, Mizuho Iwaihara, Masatoshi Yoshikawa, "Query Rewriting Rules for Versioned XML Documents," in Proc. 20th Int. Conf. Database and Expert Systems Applications (DEXA2009), Lecture Note in Computer Science 5690, pp.364-371, Linz, Sep. 2009 (査読有り) .

[学会発表] (計 13 件)

1. 吳 建民, 岩井原 瑞穂, "Improving revision graph extraction in Wikipedia based on supergram decomposition," DEIM2013 Forum, A4-1, 2013年3月.
2. Liming Shen, and Mizuho Iwaihara, "Trend Analysis of Privacy Settings and User Classifications in Social Network Services," DEIM2012 Forum, F10-2, Mar. 2012.
3. Xinyan Fu and Mizuho Iwaihara, "Community-Search Algorithm Based on Tree Decomposition of Graphs," DEIM2012 Forum, B6-2, Mar. 2012.
4. Jianmin Wu and Mizuho Iwaihara, "Wikipedia revision graph extraction based on n-gram distribution covering," DEIM2012 Forum, F10-1, Mar. 2012.

5. Zhe Cao, Mizuho Iwaihara, "Wikipedia version tree reconstruction by clustering revisions through keywords," IEICE Technical Report DE2011-32 (2011-8).
6. Jun Huang, Mizuho Iwaihara, "Utilizing Supporting Entities for Realtime Sentiment Classification," The 3rd Forum on Data Engineering and Information Management, F5-3, Feb. 2011.
7. Zhe Cao, Mizuho Iwaihara, "Estimating XML version trees through maximum weight branching", The 3rd Forum on Data Engineering and Information Management, E1-1, Feb. 2011.
8. Ding Xiaochen, Mizuho Iwaihara, "Evaluating Risks for Card-style Identity Management," 第9回情報科学技術フォーラム, D-018, 2010年9月.
9. Jun Huang, Mizuho Iwaihara, "Real-Time Support Rate Estimation Based on Social Sensors," 第9回情報科学技術フォーラム, D-045, 2010年9月.
10. 宗政俊一, 岩井原瑞穂. "ソーシャルネットワークキングサービスにおけるプライバシー設定の収集と集約," 第9回情報科学技術フォーラム, D-017, pp.129-130, 2010年9月.
11. Fan Zhongyi, Mizuho Iwaihara, "Utilizing Social Annotations for User's Privacy Policy Management," 電子情報通信学会 データ工学研究会報告 DE2010-16, pp.13-18, August 2010.
12. Wang Se, Mizuho Iwaihara, "Quality Evaluation of Wikipedia Articles Through Edit History and Editor Groups," 電子情報通信学会データ工学研究会報告 DE2010-17, pp.19-24, August 2010.
13. 岡野光太郎, 岩井原瑞穂, Gail-Joon Ahn, 吉川正俊, "ソーシャル・ネットワークキング・サービスの相互接続におけるリスク評価を用いたアクセス制御ポリシー設定支援," 電子情報通信学会第16回Webインテリジェンスとインタラクティブ研究会, 2009年10月.

[その他]  
研究室ホームページ  
<http://www.iwaihara-lab.org/pub/>

## 6. 研究組織

### (1) 研究代表者

岩井原 瑞穂 (MIZUHO IWAIHARA)  
早稲田大学・理工学術院・教授  
研究者番号：40253538

### (2) 研究分担者

吉川 正俊 (MASATOSHI YOSHIKAWA)  
京都大学・大学院情報研究科・教授  
研究者番号：30182736  
2009年4月から2010年3月まで  
馬 強 (QIANG MA)  
京都大学・大学院情報研究科・助教  
研究者番号：1430193420  
2009年4月から2010年3月まで  
馬 強 (QIANG MA)  
京都大学・大学院情報研究科・助教  
研究者番号：30415856  
2009年4月から2010年3月まで  
浅野 泰仁 (YASUHITO ASANO)  
京都大学・大学院情報研究科・助教  
研究者番号：20361157  
2009年4月から2010年3月まで