

# ソーシャルクラウド型新世代知識情報獲得支援システムの研究開発 (102107001)

Research and development of a New Generation System for Knowledge Support on the Social Crowd

## 研究代表者

河合 由起子 京都産業大学

Yukiko KAWAI Kyoto Sangyo University

## 研究分担者

秋山 豊和<sup>†</sup>

Toyokazu AKIYAMA<sup>†</sup>

<sup>†</sup> 京都産業大学

<sup>†</sup> Kyoto Sangyo University

研究期間 平成 22 年度～平成 24 年度

## 概要

Web を利用した知識獲得手法は「検索サービス」と「掲示板・SNS」の二つに分けられる。検索サービスは速報性と網羅性が高く、掲示板・SNS は人とのコミュニケーションにより質の高い情報を得られる。本研究では、これらの利点を同時に満たすことで、瞬時に欲しい情報と豊富な知識をもつユーザの両方を同時に発見する。これにより、検索結果から得られた情報に関してその場で直接人に問合せができ、また、その質問応答ログを利用して検索精度の向上をはかれる。また、情報発信者側もそれらのログに基づいて質の高い情報へと更新できるため、情報空間全体の質の向上が期待できる。さらに、情報そのものやネットワーク構造をも大きく変える可能性がある。本研究開発では、ユーザのソーシャルネットワークを構築することで、閲覧ユーザ間のリアルタイムコミュニケーションを実現し、これまでのページリンク構造だけでなく閲覧ユーザの「量と質」に基づいたリンク構造も利用したランキングの実現を目指す。

## 1. まえがき

検索サービスが生活にとって必要不可欠となっているが、インターネットに存在する玉石混合の膨大な情報のみ対象としていた検索を、一般ユーザ自身の潜在的知識をも検索の対象とすることが近い将来に求められるであろうことは想像に難くない。これを実現する現実的な手法として、おしえて goo など不特定多数のユーザに直接複雑な問い合わせをすることで質の高い情報を獲得したり、ソーシャルネットワークサービス (SNS) など知人や趣味嗜好が似ているユーザ間で情報獲得するサービスがある。ところが、これら既存のサービスは、時間がかかるだけでなく、検索サービスに比べると格段に網羅性が低い。そこで、我々は、質の高い情報と興味か似ているユーザや豊富な知識を持つユーザを瞬時に発見することを目標として、これまでの検索サービスとコミュニケーションサービスを融合したシステムの研究開発を進めてきた。本研究では、上記のシステムをさらに発展させ、大規模分散環境上でのネットワークを構築し、ソーシャルクラウド型の知識情報獲得支援システムを構築する。最終的には、数十万規模の Web ページとユーザとを対象とした検索とコミュニケーションサービスを実装し、さらに Web ページに関するコミュニケーションログを管理・公開することで情報と人の双方の質を相乗的に向上させるクラウド型新世代検索コミュニケーションシステムの構築を目標とする。

## 2. 研究開発内容及び成果

具体的な目標として以下の 4 点を中心に開発を行った。それぞれの内容と成果を以下に紹介する。

### 2. 1 Hyper link と Social link に基づく検索

既存の検索エンジンは、PageRank や HITS アルゴリズムに代表されるように、ハイパーリンク構造を用いることで高速な検索を実現している。ページのスコアは被リンクの重みより算出され、重みはページのリンク数に基づき決定される。しかし、ページのリンク構造だけでは信頼性が

低く、検索精度向上の限界がある。本研究開発では、ハイパーリンク構造だけでなく、各ページを閲覧しているユーザに着目し、実世界でアクセスしているユーザの数と質に基づいたソーシャルリンクを構築し、仮想世界と実世界の情報構造を反映した検索を実現する。さらにページ内の疑問点に対して、リアルタイムで閲覧しているユーザに直接相談することも可能となる。この実現のために、ソーシャルリンクにおけるユーザの数と質のリンク重みへの反映方法について検討し、小規模実験から始め、大規模環境に適用し、その有効性を確認した (図 1)。「ipad」など一時的に話題になる検索キーワードでの有効性、およびニュースやロコミなどなんらかのイベントにより話題になったキーワードの順位の変化に素早く対応でき、ユーザ評価に基づいたランキングが可能であることが確認できた。

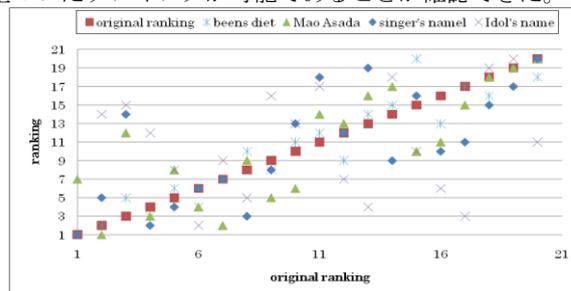


図 1: 中規模 (NTT レゾナント検索結果へのアクセスログ) データによる検索順位比較結果

### 2. 2 大規模並列分散計算環境への適用

現行のシステムは並列分散計算環境には適用しておらず、処理性能に課題がある。特に提案手法ではユーザがページを訪問している間に関連ユーザを抽出する必要があるため、リアルタイム性が要求される。そこで、今回大規模並列分散計算技術を適用した処理高速化を実現した。BLAS のサブルーチンを最適化する機能を備えたライブラリ ATLAS および GotoBLAS を用いて最適化を実施し

た。その結果、GotoBLASにより、単一ノードにおける計算性能はコア数1の場合、Ubuntuで約1/3、CentOSで約78%まで削減できた(図2)。

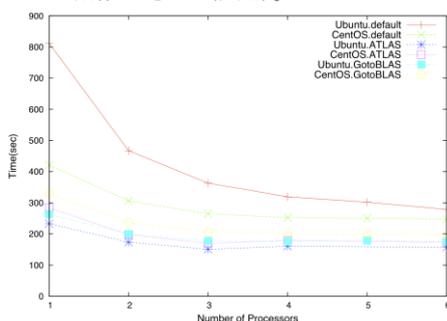


図2: ATLASとGotoBLASを用いた性能向上結果

### 2.3 コミュニケーションと検索の融合

検索結果のページにアクセスすると、そのページをライブ閲覧しているユーザとのコミュニケーションを可能にすることで、検索とコミュニケーションの融合を実現した。さらに、ページを閲覧しているユーザとのコミュニケーションを実現するだけでなく、ページの全リンク先のページを閲覧しているユーザにも通知するリンクコミュニケーション機能を実現した(図3)。これにより検索精度向上とコミュニケーション促進の相乗効果が期待できる。



図3: 拡張されたリンクコミュニケーション機能

### 2.4 既存コミュニケーションサービスとの連携

本研究開発の検索とコミュニケーションでは、教えてgooやTwitterなど既存のコミュニケーションサービスとの連携、市役所や企業の特性サイトへの適用を実装し、本システムを利用していない実世界ユーザの情報を用いた検索コミュニケーション機構を構築した。tweetデータのフォロー、フォロワーの関係性に基づいたtweetとユーザの同時発見手法(提案手法)と、既存の検索結果の直近時間順にリランキングした結果順位とのスパイアマン順位相関結果(図4)より、提案手法の有効性を確認できた。

| 指定キーワード | 被験者 | ツイート文  |       | ツイッター発信者 |       |
|---------|-----|--------|-------|----------|-------|
|         |     | 直近時間順  | 提案手法  | 直近時間順    | 提案手法  |
| 自然言語処理  | A   | 0.624  | 0.697 | 0.697    | 0.636 |
|         | B   | 0.539  | 0.321 | 0.539    | 0.891 |
|         | C   | 0.503  | 0.770 | 0.406    | 0.758 |
|         | 平均  | 0.556  | 0.596 | 0.547    | 0.762 |
| IPSJ    | A   | 0.479  | 0.830 | 0.418    | 0.436 |
|         | B   | -0.370 | 0.939 | 0.612    | 0.648 |
|         | C   | 0.164  | 0.624 | 0.418    | 0.770 |
|         | 平均  | 0.091  | 0.798 | 0.483    | 0.618 |

図4: tweet文と発信者ランキング結果に対する相関係数

3. 今後の研究開発成果の展開及び波及効果創出への取り組み

本研究課題で開発した知識獲得支援システムは、情報通信分野において技術的・学術的両面において大きな価値がある。特に、以下の2つの大きな成果があった。

#### 3.1 検索およびコミュニケーション技術

本技術は、ページやユーザの検索のみならず、センサやネットワーク機器の運用にも転用可能な技術であり、今後、抽出した関係を用いたネットワーク運用のロバスト化に適用するなど、さらなる発展が期待される。また、ユーザ

と情報のインタラクションを促進するという点で、マルチスクリーン環境などで特にその有効性が発揮できるため、平常時だけでなく震災時にも転用可能な技術であり、萌芽的な研究として災害時における検索コミュニケーションシステムの開発を開始しており、今後の発展が期待できる。

#### 3.2 大規模並列分散計算機能基盤技術

本技術のバックエンド部をJGN-Xプラットフォーム上に展開することで、効果的にリソースを利用可能となった。今後、本研究で構築したフレームワークによってSDNを活用した更なるシステムの大規模化、高速化が可能になると期待される。

#### 4. むすび

高度コンテンツ創造・分析・流通のためには、ユーザが効率よくコンテンツを理解し、より良い情報として他人へ効果的に伝えることが必要である。今回我々が開発した知識獲得支援システムは、検索サービスとコミュニケーションサービスを同時に実現するものであり、コンテンツや人を同時に検索可能にし、また、誰もが閲覧や検索を通して気軽にコミュニケーションできる技術である。本研究開発により、情報通信分野の技術的・学術的両面において1. 検索およびコミュニケーション技術、2. 大規模並列分散計算機能基盤技術の2つの大きな成果があった。今後は、通信機器にも活用することで、さらなる大規模化、高速化が期待でき、災害時支援技術への応用も期待される。

#### 【誌上发表リスト】

- [1] 松井優也、河合由起子、“人と情報の検索および相互作用を目指したソーシャルサーチシステムの研究開発”、日本ソフトウェア学会論文誌「コンピュータソフトウェア」 Vol.28 No.4(2011) pp196-205. (2011)
- [2] Toyokazu Akiyama, Yukiko Kawai, Yuya Matsui, Yoshinori Kubota and Takuya Osaki, “A Proposal for Social Search System Design”, Proc. of IEEE/IPSJ International SAINT2011, pp110-117. (July 2011)
- [3] Yuhki Shiraishi, Jianwei Zhang, Yukiko Kawai, and Toyokazu Akiyama, “Proposal of Combination System of Page-centric Communication and Search”, Proc. WOSS2012(VLDBworkshop2012), Istanbul, Turkey. (September 2012)

#### 【申請特許リスト】

- [1] 秋山豊和、河合由起子、情報検索システム及び情報検索方法、日本、2011年7月15日
- [2] 河合由起子、望月崇由、森本恭平、端末出力情報管理装置、端末装置、およびプログラム、日本、2011年11月2日

#### 【受賞リスト】

- [1] Yuya Matsui, Yukiko Kawai, Jianwei Zhang, Best Paper Award of IMECS 2011 / ICICWS 2011, “Hybrid Web Search with Social Communication Service”, March 2011.
- [2] 松井優也、青木聡、河合由起子、張建偉、最優秀インタラクティブ賞、第4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) 2012、2012年3月3日
- [3] 久保田吉徳、インターネットアーキテクチャ研究会学生研究奨励賞、“知識獲得支援システムにおける大規模ランキング計算の高速化手法の検討”、2011年3月1日

#### 【報道掲載リスト】

- [1] “gooラボ”で、同じサイトを閲覧している他のユーザとコミュニケーションできる「ペチャクチャ検索」の実証実験を開始”、日刊工業新聞、京都新聞、2011/6/2

#### 【本研究開発課題を掲載したホームページ】

<http://klab.kyoto-su.ac.jp/~mito/index-j.html>