

つながる検索エンジンの開発

—コミュニケーションサービスと検索サービスの融合—

1. 背景

知識獲得のため Web 検索を行う場合、2 つの方法が考えられる。一つは、検索サービスの利用であり、もう一つは、掲示板や SNS の利用である。検索サービスは速度と網羅性の点で有効であり、SNS を利用した検索は、コミュニケーションにより質の高い情報が得られるという利点があるが、検索サービスよりも時間や手間を要する。つまり、既存の検索サービスやコミュニケーションサービスは、どれも一長一短で、全てを満たす検索手段がない。

2. 目的

本プロジェクトでは、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の双方の利点を同時に生かせる検索として、Web ページをライブで閲覧しているユーザのネットワークを構築し、閲覧ユーザの量と質に基づいたランキングを実現することで、「瞬時に欲しい情報」と「検索時に同じ興味を持つユーザからの支援」を同時に獲得できる「つながる検索エンジン」の開発を目的とする。これにより、検索による Web ページ情報の獲得だけでなく、その情報を通して「興味が同じ人」あるいは「精通している人」へも気軽につながることで、コミュニケーションを通じた情報検索を実現できる。また、単にコミュニケーションの場を作ることだけが目的ではなく、ライブ閲覧しているユーザから支援を得ることで質の高い情報にも「つながる」ことができ、情報検索のパラダイムとなることを目指す。

3. 開発の内容

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。しかしながら、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。また、検索結果のページにアクセスした際にページ内の「どこの部分なのか」、あるいはページ内の「関連する情報の意味が分からない」といった問題がある。そこで、本研究開発では、ハイパーリンク構造だけでなく、各 Web 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。また、ユーザがページ内で閲覧している位置を観測することで、検索結果のページ内で最も閲覧箇所の多い部分を注目箇所として抽出する。これにより、ページ内で多くのユーザが見ている箇所が分かるため、ページ内の重要な部分が明白となる。さらに、ページ内で分からないことがあれば、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(図1)。

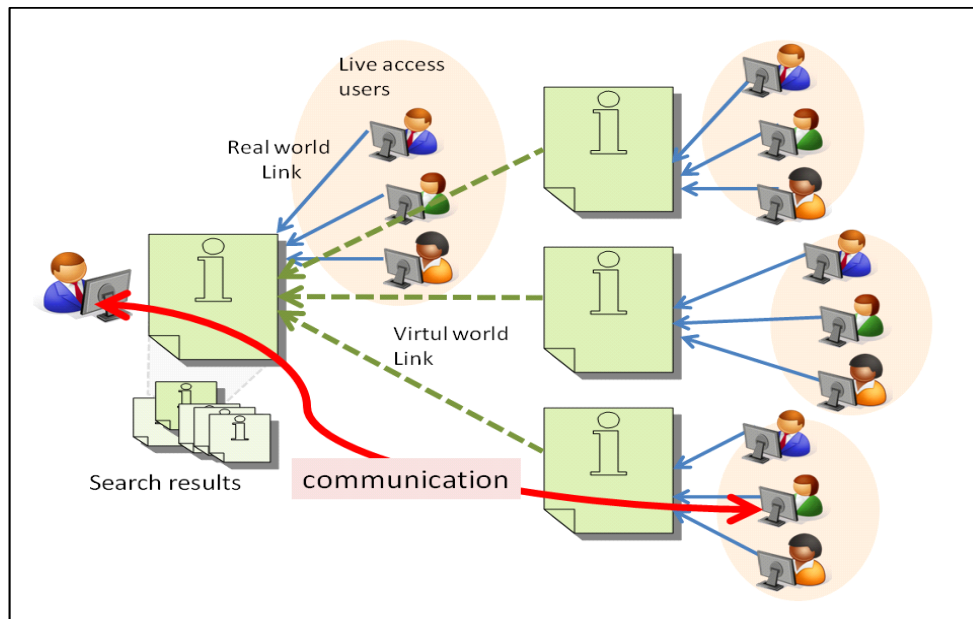


図1: 検索とコミュニケーションの融合

3. 1 サーバ構成

サーバの構成を図 2 に示す。提案システムのサーバ側の処理は Apache Tomcat6.0.18 と JDK1.6.0 07 を用いた Java の Servlet によって行っている。Servlet ではリクエストごとにユーザスレッドを作成するため、複数のリクエストを同時に処理が可能である。また Comet の技術を活用し、非ブロッキング I/O によってサーバ側からの非同期通信を実現している。これによりチャットログやサーバの状態の変更をリアルタイムにクライアントへ通知可能となった。また、ユーザとその閲覧ページの情報の格納はリレーショナルデータベースである PostgreSQL8.3.3 を用いて行った。チャットのログファイルはテキストファイルとして扱っている。Servlet は検索用サーブレットとコミュニケーション用サーブレットの二つから構成されている。

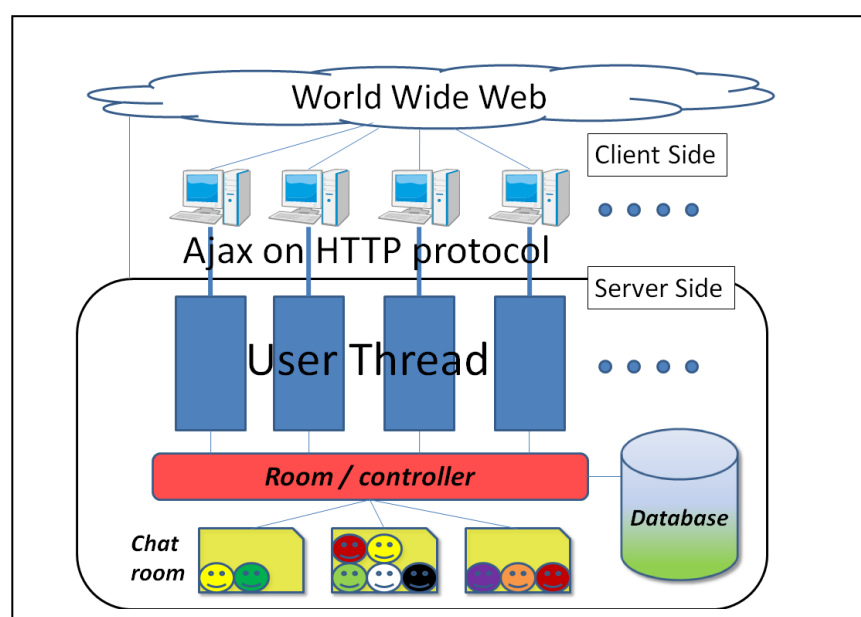


図 2: 全体の構成図

3. 2 検索機能

検索用サブルーチンではユーザからのリクエストが来ると、リクエストラインからクエリーを取り出し、Yahoo! Japan が提供するウェブ検索の API を用いて検索を行う。ウェブ検索 API からの検索結果を受け取ると、検索結果に含まれる URL ごとに閲覧者数をデータベースから取得する。データベースから取得した閲覧者数に基づき、検索結果をランキングしたものをユーザに表示している(図 3)。検索結果のタイトル横にライブでアクセスしているユーザ数を提示する。全体検索ではデータベースに登録されている URL ごとに閲覧者数を集計し、ソートを行ったものをユーザに提示する。

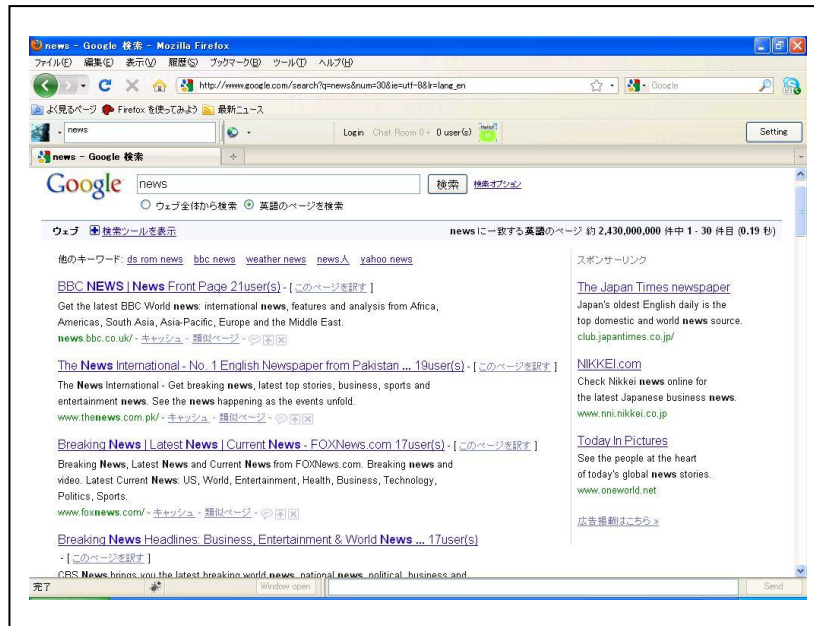


図 3: 全体検索

3. 3 コミュニケーション機能

サーバでまずユーザからのリクエストをユーザスレッドが受け取る。ユーザスレッドはリクエストが来るとリクエストラインをパースし、ユーザからのコマンドに応じてデータベース操作オブジェクトやルーム操作オブジェクトに命令を送っている。ユーザの主なデータはチャットルームのオブジェクトに格納される。チャットルームは URL、部屋名、発話ログ、入室中のユーザ情報とユーザへのレスポンスを保持している。チャットルームの操作はユーザスレッドがルーム操作クラスを用いて行っている。ユーザが同じルーム内の他ユーザにメッセージを送る際にはチャットルームの保持するユーザへのレスポンスを用いて行う。

3. 4 クライアント

本システムを利用するクライアントは、プラットフォームの普及度と開発のおこないやすさの観点から Firefox3 の拡張機能として実装した 1) (図 4)。Firefox の拡張機能では、インタフェースを XML を Mozilla が独自拡張した XUL によって記述している。プログラム部は JavaScript で記述している。提案システムのインタフェースはツールバーとブラウザ下部のステータスバーの基本部分と、サイドバーとウィンドウの二種類のチャットログ表示部分から構成されている。チャット用のキャラクタは JavaScript の DOM を用いて、Web ページの HTML 内に埋め込むことで表示させている。

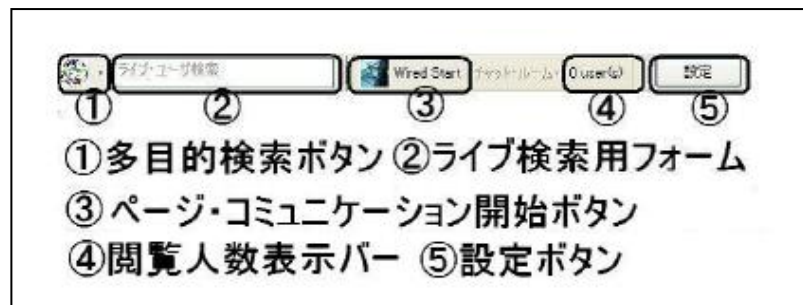


図 4: ツールバーの構成

3. 5 実装結果

今回は提案システムの検索機能とコミュニケーション機能について動作実験を行った。実行画面の遷移を図 5 に示す。

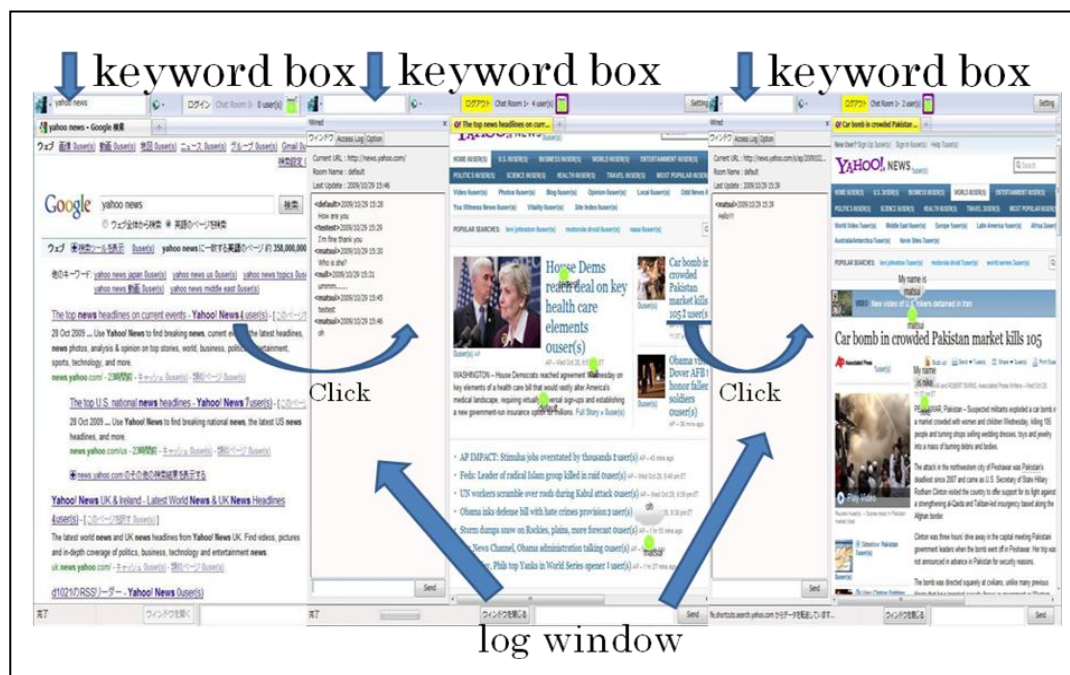


図 5: アクセス数の多い順にランキングされその後コミュニケーションを実施

図左側の Web ページは、開発したツールバーに検索キーワードを入力した結果、本検索エンジンにて検索した結果である。この検索結果は、ハイパーリンクとユーザのアクセス状況に基づきランキングしている。アクセス状況では、アクセス数とアクセスした遅延時間を考慮する。また、図中央の Web ページは、検索結果のページにアクセスしたページである。ライブで閲覧しているユーザがアバタとして表示されている。また、ペー

ジの左側には各アバタが会話したログを表示している。会話ログの下の入力ボックスに文字を入力すると、全アバタのWebページの左側の会話ログに入力された文字が表示される。尚、この会話ウィンドウとアバタを非表示することも可能である。さらに、特定の話題に関して会話したい場合は、チャットルームを作成し、入室することで、チャットルーム内のアバタのみと会話が可能である。さらにWebページの各リンクには、ライブで閲覧しているユーザ数が表示されている。このリンク先のページにアクセスすると、同様に会話ウィンドウとリンク先のユーザ数が表示されたWebページを取得できる(図右側のWebページ)。これにより、Webページ内の関連するリンクや興味のあるリンク先にいるユーザと再帰的につながることができ、コミュニケーションと検索の双方に相乗効果が期待できる。

4. 従来の技術(または機能)との相違

既存の検索サービスは速度と網羅性の点で有効であり、SNS を利用した検索は、コミュニケーションにより質の高い情報が得られるという利点があるが、検索サービスよりも時間や手間を要する。また、検索結果のページにアクセスした際に、ページ内の「どの部分なのか」、あるいはページ内の「関連する情報の意味が分からない」といったことが往々にしてある。

本プロジェクトで開発した検索エンジンは、下記の特徴を有する。

- ① 瞬時に検索・コミュニケーションが出来ること
- ② ライブ閲覧しているユーザの量に基づいた検索であること
- ③ 気軽に他のユーザとコミュニケーションができること
- ④ 難しい設定がいらないこと

通常の実検索にとって最も重要なことが、「速度」と「網羅性」であることを考えると、その違いは明らかである。既存の実検索エンジン・コミュニケーションツールを前記①～④に当てはめてみると、どれも一長一短で、全てを満たすシステムがない事は明確である。

5. 期待される効果

これまでの検索知識獲得支援は、検索とインタラクティブなコミュニケーションという領域で活発に研究開発されてきていたが、本プロジェクトにより情報と人との新しいソーシャルネットワークを構築し検索とコミュニケーションを融合することで、計算機と人の両方から同時に情報を獲得・提供可能となる。これにより単方向の情報発信から閲覧しているユーザからの情報発信も新たに加わり、人と計算機の双方向の情報発信が実現され、多くの人々が質の高い情報をやり取りできるようになる。その結果、ネットワークや情報通信の新たな利用方法の創出につながることが期待される。例えば、Mixi などの知り合い間で情報交換するソーシャルネットワークサービスやおしえて goo などの不特定多数に問い合わせることができる掲示板では、人に聞いてもすぐに回答をもらえる可能性は高くない。一方、検索では、速報性や網羅性は高いが、ページ内の不明な点の問合せや複雑な長文による問合せが困難である。しかしながら、開発したシステムにより、知りたい情報を検索すると内容の類似性が高いページを瞬時に発見できるだけでなく、そのページをリアルタイムで閲覧しているユーザを同時に発見できる。さらに、音声による問合せが可能になると、人を発見できる本提案システムより、ユビキタス環境での利活用が多大な効果を生むことは容易に想像できる。また、Web ページを通したコミュニケーションにより発生した新たな情報を管理することで、各ページに対するユーザ評価を分析することができ、例えばバーチャルショッピングやネットオークションなどのインタフェースが劇的に変化する可能性がある。これにより、これまでと異なる種類の E-コマースが生まれる可能性も考えられる。

6. 普及(または活用)の見通し

本プロジェクトの開発により、情報と人との新しいソーシャルネットワークを構築し検索とコミュニケーションを融合することで、計算機と人の両方から同時に情報を獲得・提供できようになった。知りたい情報を検索すると内容の類似性が高いページを瞬時に発見できるだけでなく、そのページをリアルタイムで閲覧しているユーザを同時に発見できることで、検索者や閲覧者だけでなく、Web ページコンテンツの作成者あるいは Web サイトを公開しているプロダクツにとっても有益である。よって、ビジネス上の普及の可能性は高く、また、利用者が本プロジェクトで開発した成果を導入する負荷は少なく、一般社会への普及の可能性は十分に高いと考えられる。

7. クリエータ名(所属)

河合 由起子(京都産業大学 コンピュータ理工学部)

松井 優也(京都産業大学 理学部)

(参考)関連URL

開発 HP : <http://klab.kyoto-su.ac.jp/~mito/>

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