学位請求論文要旨

協調サーチエンジンの研究

東洋大学大学院工学研究科情報工学専攻博士後期課程
4660010001 佐藤 永欣

本論文はイントラネットにおける最新情報の検索を目的とした協調サーチエンジン (Cooperative Search Engine, CSE) の開発、および研究成果をまとめたものである。

情報検索ではサーチエンジンが一般的に利用されている。一般的に使用されているサーチエンジンは、ロボットを用いて文書収集を行う集中型アーキテクチャに基づいている。これらの集中型サーチエンジンは、文書収集とインデックス作成に長い時間がかかるため、更新間隔を短縮できないという問題点がある。CSE では、各 Web サイト毎に配した局所サーチエンジンと、それを隠蔽する複数のメタサーチエンジンが協調して 1 つの大域サーチエンジンを構成する。CSE は Web サイトでインデックスを作成することで更新間隔を大幅に短縮し、数分でインデックス更新が可能となった。

更新間隔が大幅に短縮されたことにより、新鮮な情報の検索が可能となった。また、文書が出現・変更・消滅した時刻をほぼ正確に把握できるようになった。そこで、情報検索に時間の概念を導入し、どの文書がどの時点で新鮮であったかを知ることができる新鮮情報検索を提案し、実装した。

本論文は以下の 6 章より構成される。

第 1 章「はじめに」では、インターネットにおける情報検索と Web 文書の特性、特に、新鮮な情報の検索について、文書収集とサーチエンジンのアーキテクチャの問題点を踏まえた一般論を展開している。

第 2 章「関連研究」では、新鮮情報検索と協調サーチエンジンについて述べる準備として、情報検索の基礎的な概念を述べ、サーチエンジンの動作の原理を説明している。まず、情報検索とは何かを述べ、次に文書を収集し、検索を可能とする手順をのべている。次に集中型サーチエンジンとロボットによる文書収集の問題点を述べた。ついで、分散型サーチエンジンの原理を説明し、最大の問題点である大規模な分散検索の可能性について述べた。続いて文書のランキングについて述べ、新鮮情報検索の概念を説明する基礎として、最後に時制データベースについて述ている。

第 3 章「新鮮情報検索」では、CSE が提案・実現する新鮮情報検索について述べている。まず、その準備として、前章で述べた時制データベースを基礎として時間と情報検索の関係を議論し、時制情報検索を定義した。その上で Web 文書の性質と時制情報検索との適合性を議論し、新鮮情報検索を定義した。

第 4 章「協調サーチエンジン」では、まず、CSE の原理と構成を説明し、インデックス更新時、文書検索時の動作を説明し、インデックスの更新と文書の検索が可能なことを示した。次に、インデックス更新時の高速化技法を述べ、評価を示して、CSE が高速にインデックスを更新できることを示した。また、これに関連して文書収集時の高速化技法を提案・整理した。そして、開発初期の CSE の構成と検索時の応答時間の評価を示した。その後、様々な検索高速化技法を用いて検索時に発生する通信を削減することで、検索応答時間を高速化でき、多くの場合で集中型サーチエンジンに匹敵する速度で検索可能なことをこれらの評価を述べて示した。これらの高速化技法は、検索式最適化、検索結果の先読みによる応答時間の短縮、継続検索を最適化する、検索式に適合するサイトのみへ問い合わせることによる通信量の削減、キャッシュを共有することによる一部のサイトへの負荷集中の回避、永続的キャッシュを導入することによるサイト選択結果のインデックス更新後における再利用である。次に、協調サーチエンジンを実用的に使用するにあたって問題であると考えられた Location Server の耐故障性や、文書セキュリティの実現について述べ、最後に新鮮情報検索の実現について述べた。

第 5 章「実装」では、協調サーチエンジンの実装について述べた。これは協調サーチエンジンを実装するにあったっての基本方針や、検索時、更新時に通信に用いられるプロトコルの定義と解説、および、各コンポーネントの実装、その方針の説明を含む。

第 6 章「まとめ」では、本研究の成果をまとめている。

以上のように、本論文では新鮮情報検索を提案し、それを実現する協調サーチエンジンについて述べた。協調サーチエンジンは短時間でインデックス更新が可能であり、これにより、新規に作成された文書や更新された文書を直ちに発見できる。これは、これらの文書が作成・変更・消滅した時刻をほぼ正確に把握できることを意味する。協調サーチエンジンのこの特徴により、新鮮情報検索が初めて実現可能となった。新鮮な情報の検索は、イントラネットの情報を活用するためには重要な技術であり、今後の実用化・普及が期待できる。