

博士論文題名: 広域ネットワーク上での分散型検索システム - 動画検索を例として -

論文内容の要旨

インターネット上で分散した大量の情報を利用するためには、検索機能が必要不可欠である。しかし、少数の検索ホストで収集し索引付けを行う集中型の検索方式では、検索ホストの処理能力とネットワーク帯域の制限から、大量の情報検索が困難となる。一方で、特定のホストが情報を収集するのではなく、利用者の検索要求を分散したホストに配信し、照合結果のみを収集する分散型検索の研究が行われており、実現が期待されている。

しかし、これまでに開発された分散型検索では、ホスト数の増加に従って、検索要求の配信と照合に必要な処理とネットワークに過度な負荷を与えてしまう。また、検索要求を配信するホスト数を TTL やネットワーク遅延等で制限する方式では、目的のホストへ到達しない可能性が高なる。そのため、検索要求を配信する際、ネットワーク遅延が小さく、各利用者の満足度が高い情報をもつホストへ選択的に配信することが重要となる。

本研究では、広域ネットワーク上で大量の情報を検索するため、検索要求を効率的に配信制御する Query Routing System (QRS) の構築および評価を目標とする。特に大容量データである動画の検索を例として、QRS の設計と実験および評価を行った。

提案する QRS は複数の QRS ホストで構成され、QRS ホストは利用者与其他の QRS ホストより検索要求の画像データを受信する。検索要求を受信すると、動画を構成するフレームと照合し、同時に他の QRS ホストを選択し、選択した QRS ホストへのみ検索要求を配信する。照合と他ホストの選択および検索要求の配信は各 QRS ホストで繰り返される。

検索要求を配信する際に QRS ホストは、各 QRS ホストまでのネットワーク遅延時間と照合時間を加算した「サービス遅延時間」、利用者が検索結果より動画を視聴するためにホストを「アクセスした回数」、視聴した動画の索引情報に含まれる「キーワードの出現回数」等を選択基準とし、階層分析法により各時点で利用者にとって最適な QRS ホストを選択する。QRS ホストの選択基準に、ネットワークの情報以外に各利用者が視聴したホストへのアクセス回数や、視聴した動画の索引情報であるキーワード等を利用することで、検索キーは利用者の好みや主観に基づいて、利用者毎で動的に配信される。

本論文では、QRS ホストを LAN 上に配置し、全体に検索要求を配信した場合と、提案方式の選択した QRS ホストに配信した場合のトラフィックを計測した。その結果、提案方式では、トラフィック総量が 9% 削減された。また、広域ネットワーク上に 100 台のホストを配置した際の配信を、サービス遅延時間のみを利用する方式と、アクセス回数も利用する提案方式でシミュレートしトラフィックを計測した。その結果、提案方式ではトラフィックが削減され、さらにアクセス回数の増加に従ってトラフィックの削減が確認された。以上の結果より、QRS によりトラフィックの負荷を軽減し、有効な配信制御が行えることを示した。

論文審査結果の要旨

本論文は、インターネットにおける大量の情報検索を可能にする基礎技術として分散型検索システムに焦点をあて、既存の検索システムの本質的な問題点を明らかにするとともに、新たな分散型検索システムを提案し、トラフィックを定量的に評価した。提案した分散型検索システムでは、ネットワークの環境と利用者毎の履歴情報を評価し、検索要求を選択的に配信制御することで、各利用者の検索結果に対する満足度を向上し、さらにネットワークトラフィックの低減化を実現し、その提案システムによるトラフィックの定量的な評価を通じてシステムの有効性を示した。本論文の成果は以下に要約される。

1. 今後増加が予想される大容量な動画像の検索に焦点をあて、特定のホストが情報を収集するのではなく、利用者の検索要求を分散したホストに配信し、照合結果のみを収集する分散型検索システムに着目した。これまでに開発された分散型検索システムでは、ホスト数の増加に従ってネットワークに過度な負荷を与え、また検索要求を配信するホスト数をネットワーク遅延等の情報で制限する方式では、目的のホストへ到達しない可能性が高くなる問題点がある。本研究の新規性は、その問題点を解決する1つの方法として、ネットワーク環境の情報以外に利用者が過去に視聴した履歴情報を評価要因とし、意思決定法である階層分析法から利用者毎に最適なホストを選択し、検索要求を配信制御する Query Routing System (QRS) を提案したことにある。また、QRS を実環境で運用し、全体に検索要求を配信した場合と、選択的に QRS ホストに配信した場合の検索に必要なトラフィックを計測し評価した。その結果、QRS によりトラフィック総量の削減が確認された。

2. 上記の実環境でのトラフィックを評価する一方で、ホスト数の増加に伴うトラフィックの削減効果を、100台のホストによる検索要求の配信制御のシミュレーションで定量的に評価した。利用者の履歴情報の1つである各ホストに対するアクセス回数は照合率の高い情報を公開しているホストに集中すると仮定し、提案方式はアクセス回数とネットワーク遅延情報を評価し配信制御を行った場合とした。また、比較方式はネットワーク遅延情報のみを評価し配信制御を行った場合とし、各々の方式に発生するトラフィックを計測した。その結果、提案方式によるトラフィックの削減が確認され、更にアクセス回数の増加に従ってもトラフィックが削減されることが示され、広域ネットワークに大量に分散した情報検索における QRS の有効性が示された。

以上のように、本論文は広域ネットワークにおける分散型情報検索システムの基礎研究として、新たな検索要求の配信制御方式を提案し、実環境でのトラフィックと、更には大規模ネットワークにおけるトラフィックの削減化を提示し、学術上だけでなく、実際のシステム構築上もその貢献度は大きいといえる。よって本論文は博士(工学)の学位論文としてふさわしいものと認める。