

2.6 ビデオ情報検索

2.6.1 はじめに

大容量のビデオ情報をインターネットで提供するサービスは、高速ネットワークのインフラ整備に伴い増加の一途を辿っている。ビデオ情報にはニュースや会議、授業といったものが挙げられ、本学電子図書館においても蓄積したビデオ情報をインターネット上で提供している。これらの蓄積されたビデオ情報を提供するサービスの重要な機能の一つとして、大量のビデオ情報群の中から目的のビデオ情報を検索する機能が挙げられ、効率的で迅速な検索機能の開発が盛んに行われている。

インターネットで目的のビデオ情報を獲得する最も単純な方法は、現在提供されているテキスト情報の検索サービスの方法と同様に、全てのビデオ情報を一箇所（あるいは数箇所）に収集し集中管理する方法が考えられる。しかし、既存の収集/集中管理型の場合、ビデオ情報の検索のために付加されたテキスト情報全てを収集し管理することは可能とするかもしれないが、大容量のビデオ情報全てを一箇所に収集し管理することは不可能である。そこで、我々はビデオ情報を一箇所で集中管理することなく、目的のビデオ情報の効率的な検索を可能とする分散型のビデオ情報検索システムを提案する。

2.6.2 分散型のビデオ情報検索

我々はネットワークにおける情報検索を、『情報を蓄積しているホストがユーザーの Query を受け取って、蓄積している情報群の中から Query と一致する情報を探し出しユーザーに結果を返すこと』と定義する。ここでビデオ情報の検索で用いる Query を一般的に利用されてるテキストデータもしくは静止画像データとする。しかし、ビデオ情報はフレームといった静止画像データの集合で構成されているため、テキストデータや静止画像データを Query とする検索は、フレームの集合が表す多様な意味に対して完全に一致するビデオ情報を探し出すことを困難とする。そこで今回、ビデオ情報の検索は『Query と一致する確率（マッチングレート）の高いビデオ情報を探し出しユーザーに結果を返すこと』とする。以上の定義に基づいて、インターネットにビデオ情報を蓄積しているホストが複数存在する場合のビデオ情報の検索方法について述べる。

最も単純なビデオ情報の検索方法は、ユーザーから受け取った Query を他の全てのホストへ送信し、Query を受け取った全ホストの検索結果をユーザーへ返

信する方法である。しかし全てのホストに Query を送信し検索結果を返信することは、マッチングレートの低いビデオ情報を蓄積しているホストやマッチングレートが低くかつネットワーク遅延が大きいホストとデータの送受信を行わなければならないため、ネットワークのトラヒックと検索時間の無駄な増長を引き起こしてしまう。そこで我々は全てのホストではなくいくつかのホストを選出し、選出したホストにのみ Query を送信し検索結果をユーザーに返信することでトラヒックの軽減と検索時間の短縮を可能にするシステムを提案する。

2.6.3 ホスト選出のための各要素

前述のようにネットワークにおける分散型のビデオ情報の検索で複数のホストに Query を送信し結果を受信する場合、どのホストに検索依頼をするかが重要事項となる。我々の提案するホストの選出方法は、次の 4 つの要素を利用して選出を行う。

- ・ 検索時間
- ・ ユーザーのアクセス頻度
- ・ 存在確率
- ・ 検索方法

ここで検索時間は、『ホスト間のネットワーク遅延と保持するビデオ情報に対して検索を行う時間の合計』である。つまり、あるホストが Query を送信して Query を受信したホストが保持しているビデオ情報に対して検索を実行し、検索結果を返信し、その結果を受信するまでの時間である。ユーザーのアクセス頻度とは、一定時間内における各ホストあるいはホストに蓄積されている各ビデオ情報に対するユーザーのアクセス回数のことである。存在確率とは、ホストに蓄積されているビデオ情報の中に目的のビデオ情報と一致する情報が含まれている可能性を算出したものである。今回各ホストの存在確率は、各ホストのビデオ情報の付加情報(インデックス: テキストデータや代表的な静止画像データ)と Query を照合して算出される。検索方法とは、ホストで提供されているビデオ情報を検索する際に Query として利用できるデータ(例えばテキストデータや静止画像データ)と利用できる検索アルゴリズム等の情報である。

2.6.4 ホスト選出と Query の送信

ユーザーから Query を受信したホストは、まず自身の保持するビデオ情報に対して検索を実行する。同時に他のホストを選出し、Query は選出されたホスト

の優先順位の高い順に送信される。選出する方法は、まず検索方法の情報よりユーザーの Query が検索の際、利用可能であるホストを選出する。次に、Query と選出された各ホストのインデックスを照合して存在確率が閾値以上のものを選出する。優先順位は存在確率が高い順に決定される。ここでユーザーのアクセス頻度が閾値以上のホストが存在する場合、そのホストは選出され優先順位が上げられる。

次に Query の送信について述べる。ホストは他のホストまでの検索時間の情報を保持しており、各ホストまでの検索時間の最短経路表をダイクストラ法によって作成し保持する。最短経路表には、宛先ホストの ID(URL もしくは IP アドレス)と、宛先ホストに最短経路で到達するため次に送信すべきホストの ID が対で記されている。この最短経路表を元に選出されたホストを宛先ホストとし、その宛先ホストの対である次ホストへ Query を送信する。Query 送信の終了は、Query が目的のホストに到達した場合か、または Query 送信の際に各ホストが付加するタイムスタンプ(次に送信するホストの検索時間を加えた値)が閾値より大きい場合とする。

2.6.5 今後の課題

我々はネットワーク遅延と検索時間を削減するため 4 つの要素を元にホストを選出することを提案した。この 4 つの要素は時間軸に沿って変化するため、ホストの選出と次ホストへの Query の送信はダイナミックに行える。しかし、頻繁に更新される 3 つの要素(検索方法は他と比較して更新される頻度が少ないため除く)を随時全てのホストへ伝達することは、ネットワークのトラヒックの増加となる。頻繁に追加/削除されるビデオ情報のインデックスは存在確率の正確性を高めるために随時全ホストへ送信する必要がある、検索時間やユーザーのアクセス頻度に関しても同様なことが言える。これらの問題を解決するための新たな枠組みが必要である。例えば、インデックスと検索時間の情報伝達は全てのホストに随時送信するのではなく、意味的に分類したグループを構成し、同じグループのホストと意味的に近似するグループへのみ送信するといった方法が考えられる。また、各ホストの増減したビデオ情報の数や検索時間等の一定時間の平均値が閾値以上になった場合に送信する、といった方法が考えられる。

今後、グループ作成方法、グループ間、グループ内での各要素と Query の伝達方法の課題に取り組む予定である。