

論文内容の要旨

博士論文題目：メタ検索エンジンにおけるスコアの統合手法に関する研究

氏名： 鈴木 優

本研究では、複数のメディアで構成された電子文書を検索する方法として、一つのメディアだけに対応した複数の検索システムを統合したメタ検索エンジンの精度向上に関する研究を行った。メタ検索エンジンは次の五つの手順からなる。(1) 検索対象から特徴量を抽出し、(2) 複数の検索システムを用いてスコアを計算し、(3) 算出されたスコアを正規化し、(4) 正規化されたスコアを統合し、(5) 利用者に検索結果を提示する。これらの手順のうち、本研究では検索精度を向上させるために特に(3)と(4)に注目した。

本研究では、まずメタ検索エンジンの精度に最も影響すると考えられる関数、つまり評価関数についての考察を行う。複数の検索システムで算出されたスコアを統合して、一つの検索対象に対して一つのスコアを計算する。従来のメタ検索エンジンでは評価関数として相加平均や最大値、最小値などの単純な関数が用いられることが多かった。しかし、テキスト検索における一つの検索モデルである拡張ブーリアンモデルでは、多くの関数が複数の数値を統合するために用いられている。これらの関数をメタ検索エンジンに用いた場合に、検索精度がどのように変化するかを調べた。

次に、メタ検索エンジンを構成するそれぞれの検索システムは、出力するスコアの平均や分散、総和などが異なるので、統合する前に正規化する必要がある。従来のスコア正規化手法では、オブジェクト数の分布がスコアの範囲により異なるようなスコア群を適切に正規化することができない。そこで、オブジェクト数の分布を考慮したスコアの正規化手法を提案する。本提案では、ある検索対象オブジェクトのスコアの値が同じであっても、そのスコアの値周辺での検索対象オブジェクトの数が多い場合には少ない場合に比べて相対的に高い値となるべきであるとする。本提案を用いることにより、高い値を持つスコアが多い検索結果と低い値をもつスコアが多い検索結果を、それぞれ相互に統合可能なスコアへ正規化を行うことが可能となる。実験により、提案した正規化手法が有効であることを確認した。

(論文審査結果の要旨)

2003年12月26日に開催した公聴会の結果を参考に、2004年2月19日に本博士論文の審査を行った。次のとおり、提出された博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

申請者は博士論文において、電子文書を検索するための方法の一つであるメタ検索エンジンにおいて、検索精度を向上させるための方法を提案した。さらに、マルチメディア電子文書検索を実現させるために、複数の検索システムを統合する方法について提案した。次に、博士論文における成果について述べる。

(1) 複数のスコアを統合するための関数と検索精度との関係について明らかにした。複数の検索システムによる検索結果を統合するためには、それぞれの検索システムが計算するスコアを統合する必要がある。スコアの統合方法には関数の利用が考えられるが、従来のメタ検索エンジンでは関数の選択について十分考慮されていないことが多かった。実験結果から、関数の選択は検索精度に十分影響を与えることがあることが分かった。

(2) 複数のスコアを統合する前に、各々のスコアを正規化する方法についての提案を行った。複数の異なる検索システムを統合する際に、各々のスコア計算手法は異なるため、スコアの値が同じであっても異なる適合度を示す場合があると考えられる。提案手法では、シャノンの情報量の概念を援用した正規化手法を用いることによって、異なる検索システムが計算したスコアの値であっても、正規化を行うことによって、同じスコアの値が同じ適合度を示す。実験によって、実際に提案した正規化法が有効であることが分かった。

博士論文に述べられている提案手法は独創性が高く、実際の検索システムの利便性を向上することが可能である。よって、本論文は博士(工学)の学位論文として価値があるものと認める。