

氏 名	柿 元 俊 博
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	情 博 第 11 号
学位授与の日付	平 成 12 年 3 月 23 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	情 報 学 研 究 科 社 会 情 報 学 専 攻
学位論文題目	効率的な情報検索システムを実現するブラウジング処理法の研究

論文調査委員 (主 査) 教 授 上 林 彌 彦 教 授 石 田 亨 教 授 湯 淺 太 一

論 文 内 容 の 要 旨

本論文は、インターネット環境で利用者が適切な情報を容易に効率的に探し出すために、情報検索システムにおけるブラウジング処理を効率化するモデルとそれを実現する技術について論じた結果をまとめたもので、9章からなっている。

第1章は序論で、本研究の目的と概要が述べられている。まず本研究の背景として、最近、コンピュータネットワークの利用が拡大し、情報検索システムの重要性が高くなっており、大量情報の検索の高速化、種々雑多な情報からの検索精度の向上、マルチメディア情報への対応、利用者の拡大という課題を解決することが必要になることが述べられている。それに対応するため、従来の情報検索システムの中心であった検索キーを指定して必要な情報を絞り込む方式に対して情報そのものをブラウジングしながら欲しい情報を選択する方法を効率化することにより上記課題を解決する方法の枠組みが述べられている。

第2章は、関連研究についてまとめてあり、ブラウジングを検索に適用する技術、その構成要素として考えられる、対象情報間の関係を抽出し2または3次元空間にマッピングする技術と情報の構造の可視化技術について論じており、それぞれの技術をブラウジング処理に適用する場合の問題点を指摘している。

第3章では、情報検索システムの課題を解決するために、従来の検索手法とブラウジング処理を組み合わせた情報検索システムの構成について概要が述べられている。特に、リアルタイムに実行し、利用者の望む情報に対する観点を導入し、容易な操作を実現するための枠組みについて説明されており、情報検索システムの中でのブラウジング検索の位置付けが述べられている。

第4章では、対象情報を空間に均一に分布させ、検索者の観点を導入することを目的とする特徴表現選択について、対象情報をテキストと考え、特徴表現をキーワードとして説明している。ブラウジングの対象情報の集合と対象情報毎の特徴表現が抽出されていることを前提として、特徴表現の数を削減することにより、リアルタイム処理が実現できる。そのため、特徴表現をブラウジング処理の観点から2次元のディスプレイ空間になるべく重なりがないように均一に表示することを目的に特徴表現を選択する方法が必要になることを述べ、従来の次元削減法である多次元尺度法、特異値展開法や画像処理で利用されていた分散が最大になる特徴表現を選択する方法では、不十分なことを指摘している。新たな方法として、対象情報の集合を均等に分割する特徴表現を選択する手法を提案し、そのアルゴリズムについて述べている。

第5章では、前章で述べた特徴表現選択法により選択した特徴表現を要素にした特徴ベクトルを基に対象情報を2次元空間に分布させる方法と特徴表現選択法および2次元情報空間分布の評価法と3種類の性質の異なるテキストデータに対する評価結果が述べられている。2次元空間に分布させる手法としてリアルタイム処理の実現可能性が高く、汎用性のある Kohonen の自己組織化マップを利用し、学習範囲および最適位置の探索範囲の削減により高速化が可能なことを実験で示している。さらに、2次元情報分布空間の情報配置をキーワード検索で利用される適合率、再現率をブラウジング処理の観点から拡張し、拡張適合率、発見率という評価関数を提案し、発見率を最適にする特徴表現選択数の存在することを3種類のテキストデータで示している。さらに、特徴表現選択法の評価として、分散を最大にする特徴表現を選択する方法との比較実

験により、提案方法が優れていることを示している。

第6章では、2次元情報分布空間に時間軸などの1次元のナビゲーション軸を追加することにより、より分かり易い3次元空間を作成する方法と、各対象情報の属性情報を3次元形状により表現する方法について述べている。まだ、3次元空間の2次元ディスプレイに表示する場合の対象情報の表示の重なりを削減し、より見易くすることによりブラウジングを効率化する方法について述べている。

第7章では、テキストデータを対象に開発したブラウジング空間作成法を静止画像データに対して適用評価し、その有効性について述べている。特徴表現は色相を32等分、無彩色を3分割し計35の区画とし、値は静止画像の画素の頻度を利用し、テキストデータと同じ特徴表現選択法及び2次元分布空間作成法で作成した結果を同じ評価方法で評価し、テキストデータと同様の結果が得られることを示している。特有の処理としては、色相の各区間の配置が違和感のないようにするため、各区間の色の意味的な相関を導入することにより実現できることを示している。

第8章では、音響データに対してテキストデータと同じ手法でブラウジング空間を作成できることを示している。ここでは、特徴表現として、14の感性語のSD法による評価実験結果を利用し、評価した結果、テキストデータと同様の評価を得ることができたことを述べている。

第9章は結論で、本論文のまとめが述べられている。

論文審査の結果の要旨

本論文は、情報検索システムにおけるブラウジング処理を効率化するモデルとそれを実現する技術に関して研究した成果をまとめたものであり、得られた主な成果は次のとおりである。

(1) 情報の集合を特徴表現の類似度に従って2次元空間に均一に配置する特徴表現選択法を開発した。この方法が、従来の分散最大のものから抽出する方法より、対象データの被覆率と分布の均一性の点からテキストデータおよび静止画像データを対象にした評価で優れていることを示した。この特徴表現選択により、分布空間作成速度の向上も実現でき、高速に3次元ブラウジング空間作成が実行できることを示した。

(2) 2次元分布空間作成に利用したKohonenの自己組織化マップ(SOM)の作成速度を向上できることを示した。1,000件のデータに対してKohonenのグループが提供しているSOMPACKと比較して、4,000回の学習を約半分の時間で計算でき、量子化エラー(入力ベクトルとマップベクトルの距離の平均)も1/2以下であることを示した。

(3) 2次元情報分布空間の配置をブラウジングの観点から評価する関数を提案し、これにより、特徴表現の選択数の評価が行えることを示した。テキスト検索で性能評価に利用されてきた適合率、再現率の概念をブラウジング処理の観点から、類似度と探し回る労力を導入し、拡張適合率と発見率を提案した。さらに、3種類の性質の異なるデータで評価することにより、特徴表現の選択数と発見率の間の関係から、発見率を最適にする特徴表現選択数を決定できることを示した。

(4) 3次元ブラウジング空間で表示オブジェクトの表示の重複を避けるために、3次元ブラウジング空間の傾斜法とそれに対応したワークスルー法を開発した。これにより、表示されているデータの各種属性情報を確認し易くなり、3次元空間全体のデータの分布を把握し易くすることが出来る。

以上、本論文は、情報検索システムの課題をブラウジング処理の観点から解決できる方法を示し、実用化可能なことを示したもので、学術的にも実用性の上でも極めて有意義であるといえる。

よって、本論文は博士(情報学)の学位論文として価値あるものと認める。また、平成12年1月26日実施した論文内容とそれに関連した試問の結果合格と認めた。