

平成 23 年度 情報工学コース卒業研究報告要旨

阿草 研究室	氏 名	徳 野 達 也
卒業研究題目	デスクトップサーチシステムのための Web 検索履歴の活用手法	
<p>個人が扱うファイル量の増加に伴い、フォルダ階層によって適切にファイルを分類することは困難となっているため、ファイルシステムに対して全文検索を行うデスクトップサーチシステムが用いられてきた。しかしファイルの内容にキーワードが含まれていない画像などのファイルは検索できない。このような背景から、汎用的な全文検索手法ではなく、ユーザの活動履歴を用いる PIM(Personal Information Management) に特化した手法が提案されてきた。ユーザがいつ、どのファイルを使用したかという情報は、ファイル検索において有用であることが分かっており、ユーザの活動履歴を用いたファイル検索手法がいくつか提案されている。</p> <p>先行研究である FRIDAL では、高頻度で同時に使用されるファイル群には関係があると考え、ファイルアクセス履歴からユーザの活動を抽出し、潜在的なファイル間関係を自動的に発見・活用している。しかし、FRIDAL では発見することができないファイル間関係も存在する。本研究ではこれまでとは異なったアプローチにより、FRIDAL の拡張を試みる。</p> <p>本研究の目的は、Web 利用情報を考慮することで、FRIDAL では発見できなかった関係を発見し、デスクトップサーチシステムにおけるファイルの検索精度を向上させることである。ユーザは作業時に Web 上の文書から多くの情報を取得していることが実際に確認されており、Web 利用情報も重要な情報としてファイル検索に利用できるためである。</p> <p>本研究では、Web 利用情報の中でも Web ページ上で検索していたキーワードに着目することで、FRIDAL では考慮されていなかった Web 検索履歴を活用する手法を提案する。ユーザは作業時に参考にしたい Web ページを発見するために、作業に関係のあるキーワードで Web 検索を行い目的のページを発見する。そのため Web 検索を行った際に作業を行っていたファイルと Web 検索時のキーワードには関係がある可能性が高いと考える。</p> <p>本研究では、ファイルアクセス履歴と Web アクセス履歴を解析することで、ファイル使用情報と Web 検索情報を抽出し、ファイルと Web 検索キーワード間の関連度を求める。この関連度と Web 検索キーワードの持つ基本点によって、各ファイルに付与される Web 検索キーワードによる得点を算出する。最終結果として、FRIDAL 検索による得点に Web 検索キーワードによる得点を付加して順位を計算し、ユーザに提示する。本研究では、Web 検索キーワードの持つ基本点として、Web 検索キーワードと同じキーワードでファイルシステムを全文検索した際に得られる結果上位 N 位の平均点を用いる。</p> <p>本研究では、Firefox プラグインを実装して Web 利用情報を取得し、Samba や Windows のファイルアクセスログとともに解析することで提案手法を実現する検索システムを実装した。また被験者実験を行い、提案手法が検索精度の向上を達成しているか評価を行った。実験の結果、FRIDAL 及び従来の手法と比べ、検索結果上位において、再現率、適合率が向上していることを確認した。これにより、Web 検索キーワードと作業ファイルには深い関係があり、Web 利用履歴を考慮することはファイル検索に有用であることが分かった。</p>		