

## 平成 22 年度 情報工学コース卒業研究報告要旨

阿草研究室 研究室	氏 名	山下 訓昭
卒業研究題目	ファイルアクセス履歴から抽出した参照関係に基づくファイル検索手法	

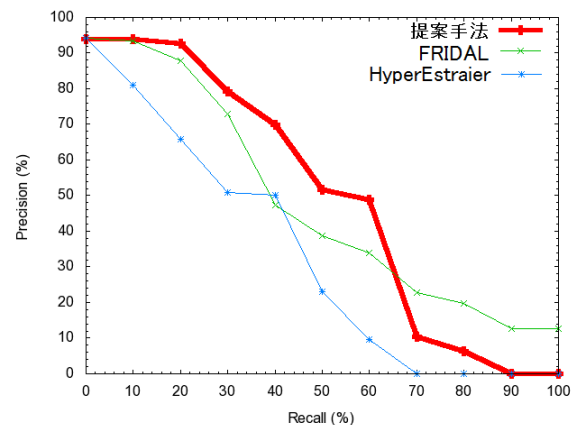
近年、ストレージの大容量化と低価格化に伴い、個人が扱う PC においてもファイルシステム内のファイル量が爆発的に増加している。現在、オペレーティングシステムのファイルシステムは階層構造で管理されているが、莫大なファイルを適切に管理することは難しい。また、関連のあるファイルであっても、プログラム、文書、画像などのカテゴリで分類し管理している場合や、ファイル名が適切でない場合は、単にディレクトリを辿るだけでは目的のファイルを探すのは困難である。

そのような背景から、ファイルシステムに対する全文検索を行うファイル検索システムが用いられてきた。しかし、これらの対象は主にテキストファイル、PDF ファイル、Microsoft Office の文書ファイル、電子メール等に限定されており、画像ファイルやデータファイルは名前にキーワードを含んでいなければ検索できない。また、検索したいファイルが全文検索の対象ファイルであっても、キーワードを含んでいなければ検索できない。しかし、このようなファイルに対する検索要求は高い。

先行研究では、頻繁に同時に使用されるファイル群に着目し、ファイル間の関連度を算出し、上述したようなキーワードを含まないファイルを検索を可能にした。ファイル間の関連度は、ファイルのオープン、クローズのアクセス履歴から抽出したファイルの使用時間を基に二つのファイルを同時に使用した時間、回数等の情報を考慮して算出する。しかし、この関連度はファイル間の関係としては弱く、多くのファイル間の関係が存在し、検索時に膨大な情報を処理することになる。

本研究の目的は、先行研究で考慮されていなかった、アクセス時の読み込みと書き込みを区別することで、同時に利用されるファイル間の関係の中でもより強い関係のみを抽出し、検索精度を向上させることである。同時に使用されるファイルの中で、ファイルへの情報の追加、変更等による書き込みアクセスが行われているファイル同士には強い関係があると考え、ファイルアクセスログから抽出したファイルの使用情報の中から書き込みアクセスが行われた情報を抽出する。このとき、ファイル間には参照被参照の関係があり、参照関係を考慮してファイル間の関連度を算出する。

研究の目的達成のために、本研究では提案手法を実装したファイル検索システムを作成した。また、被験者実験を行い、提案手法が目的を達成しているか評価を行った。実験の結果、先行研究の FRIDAL 及び従来の手法と比べて、検索結果上位において再現率、適合率が共に改善していることが確認された。また、FRIDAL と比較して、関連度の数が約 1000 分の 1、データベース使用量が約 400 分の 1 程度に減少していることが確認された。



FRIDAL との比較 (11 点平均適合率)