

多表示領域を用いた文書検索手法の提案と評価

小橋 のり子[†] 郷 健太郎[‡]山梨大学工学部[†] 山梨大学大学院医学工学総合研究部[‡]

1. はじめに

計算機の普及とディスプレイの高画質化により書類がデータ化され、ディスプレイ上で閲覧する機会が多くなった。情報共有やペーパーレス化などのメリットから書類の電子化が進んでいるためである。電子化によって書類の表示、検索が高速に行えるようになったが、ある書類の中から必要な情報を探し出すという作業はまだユーザ自身が視認によって行うことが多い[1]。このような場合、図や表を手がかりに目的の内容を検索することがある[2]。例えば読み慣れない言語の中から情報を検索するというような場合や論文の実験データから結果の補足を知りたいというような場合である。そこで、本研究では、書類内の図や表を手がかりに目的の情報を検索するというプロセスを支援するために、複数の表示領域を用いた表示順序を改良して書類内の目的の情報を効率的に検索する文書検索手法を提案する。

2. 文章閲覧法と関連研究

ディスプレイ上での書類の一般的な閲覧法は、最初から順に 1 ページずつ閲覧していく単一閲覧法である。しかしこの方法では大量の書類を閲覧する場合、最後まで閲覧するのに多くの時間を要する。また、1 枚ずつ閲覧していくため少し検索しただけでは中盤から終盤までの情報がわからないという欠点がある。

この欠点を解決するために複数カーソルを用いた走査型文字入力方式の高速化という従来研究[3]を参考にした。これは肢体障害者向けの 1 スイッチ文字入力システムを提案した研究である。文字入力には五十音表を用い、2 本のカーソルを同時に自動で走査させる。このときカーソルの配置法は 2 種類あり、等間隔配置法と隣接配置法である。等間隔配置法は 2 つのカーソルを等間隔に配置することで、列選択開始時から表の中盤の文字も選択可能になる。一方、隣接配置法は 2 つのカーソルをまとめて配置することで、目標の列を右カーソルで選択し損ねても次の移動とき左カーソルで選択することができるという特徴がある。この研究では、1 本のカーソルを用いた従来法と提案された等間隔配置法、隣接配置法の 3 つの独立変数を用い、これらの理論式にテキスト数を与えてシミュレーション実験を行った。結果は隣接配置法、等間隔配置法、従来法の順に高速であった。

以上の結果から、この手法を文章検索に応用した場合検索時間を短縮できる可能性がある。本稿ではこの従来研究から着想を得て、文章検索法を提案する。

3. 提案手法

3.1 等間隔閲覧法

等間隔閲覧法は書類を分割し、分割した始めのページから表示を行う手法である (図 1)。この手法の特徴は中盤の情報を早い段階で閲覧できるという点である。

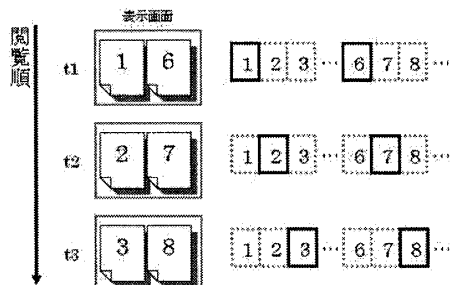


図 1: 書類 10 枚を等間隔閲覧法で表示する場合

3.2 隣接閲覧法

隣接閲覧法は右に表示していたページが次の閲覧ページで左のページに表示されるという手法である (図 2)。この手法の特徴は、一度検索し損ねても次の検索移動でもう一度見ることができるとい点である。

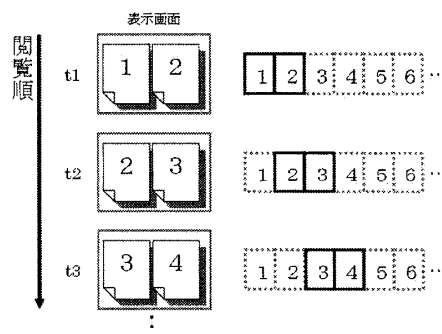


図 2: 隣接閲覧法で書類を表示した場合

4. 検索時間

単一閲覧法、等間隔閲覧法、隣接閲覧法の理論上での各平均検索時間について比較を行う。平均検索ステップ数、平均検索移動時間、平均画像閲覧時間から各手法の平均検索時間を (1)、

(2)、(3) に表す。画像枚数を L 枚、移動ボタンの 1 操作時間を m 秒、1 画像の認知時間を n 秒とすると単一閲覧法は \overline{ST}_1 、

等間隔閲覧法は \overline{ST}_2 、隣接閲覧法は \overline{ST}_3 と表される。

$$\overline{ST}_1 = \frac{L}{2}(m+n) - m \quad (1) \quad \overline{ST}_2 = \frac{L}{4}(m+2n) - m \quad (2)$$

$$\overline{ST}_3 = \frac{L}{2}(m+2n) - (m+2n) \quad (3)$$

仮に $L = 40$, $m = 0.2$, $n = 0.5$ とすると、 $\overline{ST}_1 = 13.8$ [s], $\overline{ST}_2 = 11.8$ [s], $\overline{ST}_3 = 22.8$ [s] となり、

$$\overline{ST}_3 > \overline{ST}_1 > \overline{ST}_2$$

と表される。

Document retrieval technique proposal and evaluation that uses multi display area

[†]Noriko Kobashi, [‡]Kentaro Go

[†]Faculty of Engineering, University of Yamanashi,

[‡]Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi.

同一の画像が 2 度現れるため ST_3 が最も遅くなったが、画像のエラーが他の手法に比べて低い可能性もあるので、今回の実験に隣接閲覧法も加える。

5. 評価実験

5.1 実験 1

本実験の目的は単一閲覧法、等間隔閲覧法、隣接閲覧法の理論値と実測値の比較検討をすることである。実験タスクは 3 つの手法で表示させた画像群から、ディスプレイに提示された検索すべき画像を検索するというものである。このとき検索時間を計測する。提示画像は 2 枚である。画像群はキーボードのボタンを使って移動させる。戻って検索することはできない。これは理論値にこの時間が含まれていないためである。使用する画像群は世界の新聞の 1 面で 40 枚である。インタフェースは Microsoft Visual Basic 2005 で実装したものをを用いる (図 3)。被験者は一般的な視力を持った 10 人 (コンタクトレンズ、めがね使用可) が参加し、試行回数は各タスクにつき 5 回行う。



図 3: 実験 1 で使用するインタフェース

5.2 実験 2

本実験の目的は等間隔閲覧法と隣接閲覧法の有効性を確認することである。ボタン機能に前のページに戻る機能と画像群の位置確認用のスクロールバー機能が追加される。実際にこれらの独立変数を使用する場合戻る機能を必要とするので、これらの機能を含めた検索時間を計測する。実験終了後、質問紙調査を行った。実験タスク、被験者、試行回数、実験計画、実装環境は実験 1 と同様である。

5.3 各実験の独立変数と従属変数

独立変数は画像閲覧法 3 種類である。: (1) 単一閲覧法, (2) 等間隔閲覧法, (3) 隣接閲覧法である。単一閲覧法では提示画像を 2 枚表示する。これは表示面積の統制をとるためである。

従属変数は、(1) 検索時間: 検索を開始してから画像を発見するまでの時間, (2) エラー率: 実験 1 では検索画像を含んだ画面にできなかった場合、実験 2 では検索画像とは異なる画像の発見回数とした。

5.4 結果と考察

5.4.1 実験 1

平均検索時間は等間隔閲覧法、単一閲覧法、隣接閲覧法の順に速かった。平均検索時間と標準偏差を表す (図 4)。検索時間を有意水準 5% で分散分析を行った結果、有意な差は無かった。また、平均エラー率は単一閲覧法と等間隔閲覧法が同じ 6.0%、隣接閲覧法は 2.0% だった。こちらも有意差はなかった。

等間隔閲覧法と隣接閲覧法の理論上の検索時間と比較すると結果が各手法とも速くなっている。これは 1 画像あたりの閲覧時間が理論上の時間よりも短くなったためだと考えられる。

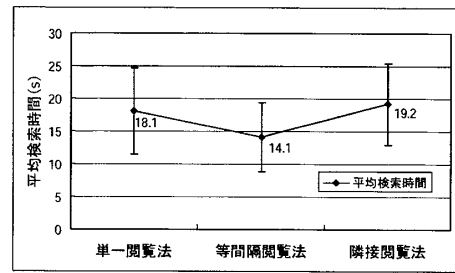


図 4: 各手法の平均検索時間と標準偏差

5.4.2 実験 2

平均検索時間は等間隔閲覧法、単一閲覧法、隣接閲覧法の順に速かった。(図 5) しかし、検索時間を有意水準 5% で分散分析を行った結果、有意な差はなかった。また、エラーは起こらなかった。

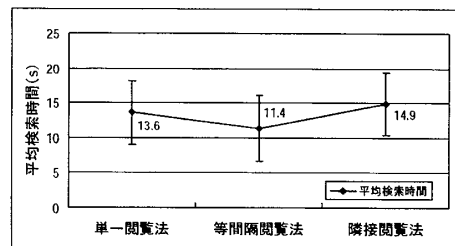


図 5: 各手法の平均検索時間と標準偏差

前の画面に戻る機能はほとんどの被験者が使用していた。その主な理由はページを移動する際に行き過ぎてしまい戻ったという理由だった。また、戻るボタンが加えられたことで安心感があったという回答もあった。このことから画像をじっくり見るというよりも、すばやく判断し次の画面に映っていたことがわかる。通り過ぎてもまた戻れるという安心感が画面の切り替えを早めた要因のひとつだと考えられる。また、各手法で感想を尋ねたところ、等間隔閲覧法において検索に集中力を要したという回答や目が疲れたという回答を得た。ステップ数は単一閲覧法に比べて約半分だが、情報量は増加したため検索時間があまり短縮されなかったと考えられる。

6. おわりに

本稿では、多表示領域を用いた文章検索手法必要性を述べ、画像検索手法を 2 種類提案した。また、理論上の数値を表した。一般的な手法も加え、3 種類の画像提示法を用いて画像検索実験を行った。今後の課題として書類間に関連のある書類を用いての有効性の確認、検索画像や提示画像の大きさの検索時間への影響の調査が挙げられる。

参考文献

- [1] 逸村 裕, 種市 淳子, 大学生のサーチエンジン情報、探索行動の分析: タイムサンプリング法を用いて. 名古屋大学付属図書館研究年報 4, pp. 1-12, 2006.
- [2] 増井 俊之, 塚田 浩二, 高林 哲, 近傍関係にもとづく情報検索システム. 情報処理研究報告会. ヒューマンインタフェース研究報告会報告, pp. 53-58, 2003.
- [3] 森 大毅, 岩淵 繁之, 粕谷 英樹, 複数カーソルを用いた走査型文字入力方式の高速化. ヒューマンインタフェース学会論文誌, Vol. 5, No. 1, pp. 135-141, 2002.