

Web ページの階層的な分割手法と提示に関する研究

田崎 雄一郎

Web ページには様々な話題や情報が含まれ、それぞれのページにおける情報の量は増加している。しかし、多くの情報がページ中に含まれていたとしても、利用者が目的とする情報はページの一部分のみである場合が多い。そのため、Web ページ中に目的の情報があるかどうかも含めて、画面スクロールなどを繰り返して情報が記述されている領域を発見する作業が必要となる。

本論文では、多くの情報が含まれる Web ページの一部分を、ページのレイアウト情報を保持しつつ、あらかじめ指定した分量の情報を持ったブロックに分割し提示する手法を提案する。ブロックを単位に利用者に提示することで、利用者が目的とするページ内の領域を容易に閲覧できると考える。

提案手法の実現には、Web ページを提示に適切な大きさのブロックに分割しなければならない。ここでは HTML で記述された Web ページを、構造的な切れ目を意味する HTML タグと、記述されている情報の分量に着目して分割する。ページの構造的な切れ目には、HTML のブロックレベル要素を用い、それだけでは分割しきれなかった部分をまとめることで、ページ内の階層構造を再構成した。ブロックごとの情報の分量にはテキストの文字数を用い、階層構造を考慮しつつ適切な文字数になるようブロックの分割と結合を行う。

この手法により分割されたブロックを単位に Web ページを部分提示する応用例として、スニペットの周辺情報を、ブロックとして付随させる提示法を提案する。スニペットとは、検索エンジンが提示する、検索結果のページ中に含まれるテキスト情報である。スニペットはレイアウト情報が失われていることや、情報の分量が少ないことから、結果の比較・判断に必要な情報が不足する場合がある。

提示を行うためには、Web ページのブロック分割が部分提示に適した大きさに分割されていなければならない。ブロック分割の適切さを、利用者実験を行って評価した。この結果、300~ 500 字程度のテキスト量で分割することが、部分提示では適切なサイズであるとの知見を得た。300~ 500 字という結果は、検索エンジンが提示するスニペットよりも明らかに多く、本研究が提案するブロック単位での提示が有用であることが期待される。

今後はブロックを単位として Web ページを提示することによる、利用者の情報発見を支援するシステムを実装していきたい。

(指導教員 佐藤 哲司)