

論文の内容の要旨

論文題目 A Study on Temporal Summarization of Web Pages
(Web ページのテンポラル要約の研究)

氏名 ヤトフト アダム ヴァディスラヴ

Web の重要性が高まっている今日、Web ページの要約に焦点を合わせた研究が必要になってきている。Web ページの最も重要な特徴の 1 つは、時間経過のなかで内容と構造が変化することである。多くの人気のある Web ページでは、絶えず情報は新しく入れ替わり、展開し、提供される。本研究の目的は、一定の期間内であるトピックに関連する一連の Web ページから、変化のあった内容を取り出して要約する方法を提案することである。取り出された要約は、一定期間の Web ページの集合の中で見出される最もポピュラーなトピックや概念を反映していなければならない。そのような時間的要約は、Web ページの内容の変化に焦点を合わせている点で、通常の多文書要約とは異なっている。

本論文で、私は Web 要約の時系列面に焦点を合わせ、単語のスコアリングとセンテンス抽出による Web ページの時系列要約の 2 つの主要なアプローチを提案した。すなわち、オンライン及びレトロスペクティブ(retrospective)なアプローチである。オンラインの時系列要約では、関連する Web ページ群について一定周期で要約を作成する。それは Web ページの集合を維持・更新する際に、一定周期毎に要約を作成する **Change Summarizer** システムとして実装されている。第 2 のアプローチであるレトロスペクティブでは、Web 集合を回顧的に捉える。時系列 Web 集合は、複数のページを取って来ることで作成され、それらは後にまとめられて要約される。単語のスコアリングの 2 つの方法もここで提案される。すなわち、スライディング・ウインドウ及び統計分析である。スライディング・ウインドウは、内容の変化のみを扱う。統計手法は時系列データマイニングのアプローチを用いて計算される単語の統計的パラメータ情報を用いる。