

システム情報工学研究科修士論文概要

年 度	平成 27 年度	学位名		修士(工学)
専 攻	知能機能システム	専攻	著者氏名	井上 祐輔
指導教員氏名 宇津呂 武仁				
論文題目 ウェブ検索結果の集約における多様な話題のウェブページの選択的収集				
論文概要 <p>現代の情報社会においては、インターネットの普及により、ウェブ上に膨大な量の情報が溢れている。このような膨大な量の情報の中から、ユーザが求める情報を見つけ出すための手段としては、Google 等の検索エンジンの利用が一般的である。検索エンジン会社はユーザの検索行動支援のため、検索エンジン・サジェストというサービスを提供している。このサービスにおいては、検索者が入力した検索語のログを蓄積し、それらを用いて強い関連を持つ語が検索エンジン・サジェストとして提供されている。検索エンジン・サジェストは検索者のログに基づいて作られているため、ウェブ検索者の関心事項そのものが反映されていると考えられる。そこで、本研究では検索エンジン・サジェストをウェブ検索者の関心事項であると見なし、検索エンジン・サジェストを情報源としてウェブ検索者の情報要求観点の収集を行う。</p> <p>本研究の枠組みにおいては、ある一語に対して最大約 1,000 語のサジェストを収集することが出来る。しかし、収集されるサジェスト、および、それらを用いて収集されるウェブページでは話題が重複しており冗長であることが考えられる。そこで本研究では、検索エンジン・サジェストを情報源として収集されたウェブ検索者の情報要求観点を集約・俯瞰することを目的とする。</p> <p>特に、本研究では、トピックモデルを用いた話題集約の手法を提案する。本研究で提案する手法においては、まず、ある一語に対して最大約 1,000 語のサジェストを収集し、それらサジェストを用いてウェブページの収集を行う。そして、収集されたウェブページ集合に対してトピックモデルを適用し、トピックと呼ばれる話題のまとまりごとにウェブページのクラスタリングを行う。各ウェブページはサジェストを用いて収集されたものであるため、各ウェブページには最低一つ以上のサジェストを対応付けることが出来る。この対応付けによりサジェストの集約を行う。その結果、約 1,000 語あったサジェストを数十個程のまとまりへと集約することが出来る。ここで、サジェストの集約結果においては類似するサジェストがまとめられているため、それらサジェストにより収集されるウェブページは互いに話題が類似しており冗長である事が考えられる。そこで、本研究では、サジェストの集約結果において各トピックに属するサジェストを用いて、多様なサジェストを含むウェブページを選択的に提示する手法を提案する。本研究では、以上のサジェストの集約手法、および、ウェブ検索結果の集約におけるウェブページの選定に関して評価を行い、その有効性を示す。</p>				
審査日 平成 28 年 1 月 27 日				
審査員	(大学名 職名)	(学位)	(氏名)	
主査	筑波大学 教授	博士(工学)	宇津呂 武仁	
副査	筑波大学 教授	工学博士	丸山 勉	
副査	筑波大学 准教授	博士(情報科学)、博士(デザイン学)	星野 准一	