

システム情報工学研究科修士論文概要

年 度	平成 22 年度	学位名	修士(工学)
専 攻	コンピュータサイエンス 専攻	著者氏名	吉田 光男
指導教員氏名 山本 幹雄			
論文題目			
ブログページ集合からのポスト及びコメント自動分離抽出手法の研究			
論文概要			
<p>近年のブログの普及により、ブログのコンテンツを利用するサービスや研究が盛んになってきている。ブログに限らないが、最近の Web ページは、そのページがまさに伝えたい情報(コンテンツ)と直接関係しない情報(不要部分)が多々含まれる傾向がある。このような、コンテンツ以外の不要部分は、機械的にページを処理するようなサービスにおいてノイズになることが多い。そのため、ブログのコンテンツを利用するためには、コンテンツ抽出が必要になる。さらに、ブログのコンテンツは、ポストと呼ばれるブログの書き手によるコンテンツと、コメントと呼ばれるブログの読み手によるコンテンツに大別することができる。ポストとコメントの存在はブログの特性の一つであり、ブログの特性を活用するサービスや研究では、ポスト及びコメントを別々に抽出できていることが望ましい。</p> <p>本研究では、ブログページ集合を用いることにより、ポストとコメントを自動的に分離抽出する手法を提案する。本手法は、ポストはブログページ集合全てのブログページに出現するが、コメントはいずれかのブログページにしか出現しないという点に着目し考案した。また、本手法のアルゴリズムを実装したソフトウェアを用いて実験を行い、日本語ブログサイトに対しての有効性を検証し、コンテンツをポスト及びコメントに分離できることを確認した。</p>			
審査日 平成 23 年 2 月 4 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)	山本 幹雄
副査	筑波大学 准教授	博士(工学)	天笠 俊之
副査	筑波大学 准教授	博士(工学)	宇津呂 武仁