

嗜好抽出を目的とした電子書籍に対するアノテーションの分析

鈴木 啓史

インターネットの普及にともない、誰もが大量の情報を入手できるようになった一方で、その中から自分が欲しい情報を自力で探すことが難しくなった。そのため、ユーザの嗜好に基づいて必要なアイテムを推定し提示する情報推薦システムが注目されている。ユーザの好みに合ったアイテムを推薦するためには嗜好抽出が重要な技術となっている。嗜好を抽出する方法には、コンテンツに付与された情報を利用する方法がある。この方法では、ユーザがコンテンツに多数の情報を付与するため、ノイズとなる情報が生成されてしまう。これに対して、コンテンツの一部をそのまま嗜好として利用する方法では、精度の高い推薦が実現できることが先行研究により明らかになっている。本研究では、この手法を電子書籍の推薦に適用することを考えた。コンテンツの一部を取り出す方法として読書中に行うアノテーションに着目し、電子書籍で自然に読書をしながら嗜好抽出することを目指す。

本研究の目的は、情報推薦における嗜好を抽出する手法として、電子書籍に対するアノテーションが利用可能かどうかを検証することである。

実際に電子書籍に対するアノテーションを収集するために、アノテーションを記録・保存できる Web アプリケーション「あのれこ」を開発した。

「あのれこ」を使用して 12 名の大学生を対象に、それぞれ 2 冊の電子書籍に対してアノテーションをしてもらった。その後、アノテーションされた文字列中の単語、本文全体からランダムに抽出された単語、tfidf により抽出された単語それぞれに対して、「興味のある単語の割合」を比較した。

その結果、アノテーションから単語を抽出する手法の方が、本文全体からランダムに単語を抽出する手法及び tfidf による単語抽出手法よりも、興味関心のある単語を多く抽出することができた。また、アノテーションは比較的長い文章を対象に付与されることが多く、興味関心のある単語を効率よく抽出することは難しいことが明らかになった。今後の課題は、システムのアノテーション機能の改良、キーワード単位で興味関心のある箇所のアノテーションをする実験の実施が挙げられる。

(指導教員 松村敦)