

マイクロブログユーザのクラスタに着目したイベント手がかり語の抽出

田中 匠

昨今、電子端末の普及と情報技術の発達により、Web が情報収集のツールとして盛んに利用されている。しかし、Web 上には情報が溢れており、自分の興味に関連したイベントを網羅的に知ることは難しい。そこで本研究では、システムの利用者が「自分の興味があることに関連した過去に起きたイベント」を探す手がかりとなる「イベント手がかり語」の抽出手法を提案し、その実装を行った。

提案手法の概要を説明する。この研究は、システムの利用者が求める情報の関連情報をよく投稿しているマイクロブログユーザを集め、それらユーザの投稿を集約すると、その内容からシステムの利用者が必要としている情報が得られるという仮定に基づいている。まず、システムの利用者は、興味のある事柄を表すクエリと期間の入力を行う。次に、システムの利用者によって入力されたクエリを関連語によって拡張し、その拡張クエリを用いてマイクロブログユーザを抽出する。その後、マイクロブログユーザ群の期間中の投稿を集約することにより、指定した期間中に起こったイベントを知る手がかり語が得られる。

イベントを表す 4 つのキーワードをクエリとして用いた実験により、手がかり語としてイベントを示す言葉の出力を確認した。

提案手法による手がかり語抽出の例を挙げる。ある時期に以下のイベントが発生していたとする。

紀伊国屋書店ら 13 社が電子書籍販売推進コンソーシアムを設立し、来春には書店での電子書籍の実験販売を開始する予定。アマゾンに対抗した、連合体ジャパゾンを目指す。

このとき、システムの利用者はこのイベントの手がかりとなる語を知らないため、電子書籍というクエリだけで検索してもこのイベントにたどり着くのは難しいが、クエリとこのイベントが発生した時期をシステムの利用者が入力した、期間に含めて手がかり語抽出を行うことにより、「ジャパゾン」「コンソーシアム」「対抗」「書店」などといった手がかり語が得られ、これらの手がかり語をクエリと併用することにより、イベントの検索が容易になる。

本研究のシステムは、Web 上で動作するサービスとして実装することを目標としていた。しかし、クエリによっては実行時間がかかり長くなってしまう場合があるため、今後の課題として、実行時間の短縮と、それに伴って Web サービスへの移行が挙げられる。

(指導教員 関 洋平)