

うろ覚えマンガの探索支援システムの構築

渡辺 葵

現在、マンガ資料は日本の重要な資料となっている。そんなマンガ資料であるが、正確な情報を持っていれば豊富な情報源があるが、うろ覚えの情報での検索は現状困難である。そこで本研究では、うろ覚えのマンガを検索するための質問を収集、分析し、どのような探索支援を行うことが望ましいか検討し、検索支援システムを構築することを目的とした。

まず、検索ニーズの高い要素を探るため、うろ覚えの情報でマンガを検索している質問の収集、内容分析を行った。2ちゃんねるから 200 件、Yahoo!知恵袋から 100 件の質問合計 300 件の質問を対象とした。その結果、マンガに登場するキャラクターの行動とマンガのジャンルについては、双方ともに 50%以上の質問で使われていた。次いで、キャラクターの特徴についても高い検索ニーズがあることが分かった。

そこで本研究では、キャラクターの行動に着目した検索手法の提案とシステムの構築を行った。キャラクターの行動が載っているデータベースとして、Wikipedia 上のマンガ作品の記事から形態素解析ツール MeCab を用いて品詞抽出と単語分割を行い、データベースを作成した。検索には汎用連想計算エンジン GETA を用いて、類似度計算には tf-idf の内積を用いた。

評価実験には 20 作品分の質問と回答のセットを用意した。検索結果上位 1000 件での順位の逆数の平均値 MRR(Mean Reciprocal Rank)を用いて評価を行った。

結果は、動詞のみを抜き出したものでは 0.00081 と低い値になった。そのため、名詞情報を抜き出し組み合わせたところ、0.048 となり改善がみられた。さらに動詞を省いて名詞情報のみにすると 0.069 となり結果が改善された。

評価実験の結果、動詞だけではマッチングする作品が少ないこと、名詞情報を加えることで精度が向上することを確認した。

以上の結果から、キャラクターの行動を抜き出すためには動詞抽出とは別のアプローチか、係り受けなど、他の品詞との組み合わせを利用するといった改善が必要であり、動詞そのままでの検索は困難であることが分かった。また、Wikipedia にはキャラクターの属性や、概要について書かれているため、名詞での検索は精度が上がったと思われる。

今後の課題としては、上記のような改善を進めることや、質問内容分析調査の結果では、掲載年や画風などについての情報にも高い検索ニーズがあったため、そういった情報を利用した探索支援の手法の提案が挙げられる。

(指導教員 高久雅生)