

上位下位関係抽出ツールを使用した語彙変換システム

PM コース 矢吹研究室 1342069 下村 渉

1. 研究の背景

プロジェクトの進行に密接に関わってくる環境的・技術的などの要因は多々あるが、最も重要なのは関係者間のコミュニケーションである。コミュニケーション重要視して適切に取っていく事で、どのような効果が期待できるか、どのような目的でコミュニケーションを取っていくのか。それは、関係者間での認識のズレ、使用の抜け・漏れ、誤解が生まれるのを防止するという目的があり、それにより、プロジェクトの円滑な進行が期待できる。発注者と開発者との間の「不適切なコミュニケーション」の例として次のようなものが考えられる。

- 開発者の専門的な用語を使った説明では、発注者は十分な理解を得られないかもしれない
- 発注者が「実装して当然」と想定している機能について、開発者は「開発者から何も言われていない」という理由で、実装する機能として想定していないかもしれない

仕様書や設計書、計画書などのドキュメントを作る目的の一つとして「関係者間のコミュニケーションツールとして利用するため」という考えがある発注者の要求を開発者が十分に汲み取り、要件として落とし込めているか、要件に過不足は無いのか、発注者と開発者の認識を合わせるためのツールであり、開発側内部では、詳細な設計や実装を行うエンジニアとのコミュニケーションツールにも利用されるだろう [1]。

発注者と開発者がそれぞれ使う専門的な用語を目的に応じた用語に変換を実現する共通な仕組みができないか着目する。用語の変換を行うために用語をタグで結び、関連用語を抽出することを考える。これと同じような事例を探し、参考に研究する。そして、2008 年の Wikipedia の記事構造から上位下位関係抽出の論文を参考にし、MediaWiki を利用する。[2] MediaWiki は現在、オープンソース (GPL) で配布されているので、MediaWiki を用いれば、自分専用の Wikipedia 運用できる [3]。目的に応じて語彙の変換を可能にするシステムを研究する。目的本研究では、MediaWiki 内にある単語の上位下位関係を抽出し、関係者間での視点や用語が異なる語彙を目的に応じた最適な語彙への変換ができるようなシステムを作ることと卒論で応用するため知識を深めるためである。プロジェクトマネジメントとの関連本研究は、発注者が発注する際に作る仕様書、設計書を開発者が十分に理解するためなどに使用する。これは発注者と開発者の不適切なコミュニケーションを避けることが出来るのでコミュニケーションマネジメントに関連づくと考えられる。

2. 研究の目的

本研究では、MediaWiki 内にある単語の上位下位関係を抽出し、関係者間での視点や用語が異なる語彙を目的に応じた最適な語彙への変換ができるようなシステムを作ることと卒論で応用するため知識を深めるためである。

3. プロジェクトマネジメントとの関連

本研究は、発注者が発注する際に作る仕様書、設計書を開発者が十分に理解するためなどに使用する。これは発注者と開発者の不適切なコミュニケーションを避けることが出来るのでコミュニケーションマネジメントに関連づくと考えられる。

4. 研究の方法

本研究では、MediaWiki のサーバーを立ち上げ、各視点からの専門用語のページを作る。作った用語のページから用語間の関連情報を抽出し、抽出した情報から用語の翻訳をする。

5. 現在の進捗状況

MediaWiki をインストールしていくつかページを作り，MediaWiki のデータを XML にダンプし上位下位関係抽出ツールで解析し抽出した．抽出したデータを少しだけ紹介する．

メインキャラクター イオ・ユークレース 1.229342

メインキャラクター オイゲン 1.229342

メインキャラクター カタリナ・アリゼ 1.229342

メインキャラクター ビィ 1.229342

メインキャラクター ラカム 1.252170

メインキャラクター ルリア 1.229342

メインキャラクター ロゼッタ 1.152684

これは「グランブルーファンタジー」というページの「メインキャラクター」の解析結果で，左から上位語，下位語，SVM を示している．次に同じページの「登場キャラクター」の解析結果の一部を記載する．

登場キャラクター イオ・ユークレース 1.349616

登場キャラクター ぴにゃこら太 0.996346

登場キャラクター オイゲン 1.349616

登場キャラクター カッタクリ 1.068125

登場キャラクター ロゼッタ 1.43511

登場キャラクター ピエール 1.50221

登場キャラクター ラカム 1.372444

登場キャラクター 帝国兵 0.874956

記載した解析結果は「メインキャラクター」とそれ以外のキャラクターの結果を交互に並べた．この結果より「メインキャラクター」で抽出されたキャラは SVM が 1.3 を超えていてそうでないキャラクターは基本下回っていることが分かるが，たまに上回っているものもある．

6. 今後の計画

専門用語を検索したときに適切な用語に転換できるよう検索システムの開発を予定している．

参考文献

- [1] 新井直人. 仕様書・設計書を作る目的の一つとして. <http://www.mitsue.co.jp/column/backnum/20110527.html>.
- [2] 隅田飛鳥鳥澤健太郎. Wikipedia の記事構造からの上位下位関係抽出. 1) 北陸先端科学技術大学院大学情報科学研究科 2) 奈良先端科学技術大学院大学情報科学研究科 3) 東京大学生産技術研究所 4) 独立行政法人情報通信研究機構, 2011.
- [3] ニコニコ大百科. Mediawiki. <http://dic.nicovideo.jp/a/mediawiki>.