

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、最も重要な事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を重要視して適切に取っていく事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効果があるので、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」の例としては次のようなものが考えられる。

- 開発者が専門的な用語で説明をしてしまうと、発注者は十分な理解を得られない可能性がある
- 発注者が「この機能」は実装して当然だ」と想定していても、開発者は「発注者から何も言われていない」という理由で機能をしないかもしれない

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は「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」という事である。これらのドキュメントは「関係者間のコミュニケーションツールとして利用する」という事を意識して、「不適切なコミュニケーション」が発生しないように作成する事で、プロジェクトの円滑な進行に役立つツールとなる [1]。

本研究では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ため共通な仕組みができないか着目し、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。私はこれと同じような事例を探し、参考に研究し、さらに、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、誰でも自分専用の Wikipedia を運用できる [3]。以上のことから本研究は MediaWiki を利用し、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。

2. 研究の目的

本研究の目的は、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、また卒論で応用するため知識を深める事である。

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、各視点から専門用語のページを作り、作った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する。

5. 現在の進捗状況

現在の進捗状況は 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。

上位語	下位語	SVM
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	1.252179
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

この結果は「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。

上位語	下位語	SVM
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	びにゃこら太	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	ラカム	1.372444
登場キャラクター	ビエール	1.505221
登場キャラクター	ロゼッタ	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ものである。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

今後の計画は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, December 2015. <http://www.mitsue.co.jp/column/backnum/20110527.html>(2015.12.15 閲覧).
- [2] 隅田飛鳥, 吉永直樹, 鳥澤健太郎. Wikipediaの記事構造からの上位下位関係抽出. 自然言語処理, Vol. Vol.16, No. No.3, pp. pp. 3–24, December 2015.
- [3] ニコニコ大百科. Mediawiki, December 2015. http://seiga.nicovideo.jp/manga/?track=global_navi_illustr(2015.12.15 閲覧).