

文章理解を支援するインタフェースの開発 —物語の人間関係を時系列で視覚化—

1. 背景

物語の内容を効率良く確認したい場面は数多く存在する。1 つ目は、過去に読んだ本の内容を確認したいときである。例えば、全 7 巻からなる J.K. Rowling (松岡佑子訳) のハリー・ポッターシリーズ第 7 巻「ハリー・ポッターと死の秘宝(上・下)」を読むにあたっては、第 1 巻「ハリー・ポッターと賢者の石」から第 6 巻「ハリー・ポッターと謎のプリンス(上・下)」まで計 9 冊の内容を把握していなければならない。それらを過去に読んでいたとしても、内容を忘れていたら最初から読み返さなければならず、とても時間がかかる。2 つ目は、物語から置いてきぼりにされてしまったときである。例えば、推理小説で「犯人は××だった」という表現を読んだとき、「××」が誰か分からないとその物語を楽しむことができない。しかし、そこまで読んできた文章の中から「××」が登場するシーンを探すのは時間がかかる。このように、過去に読んだ本の内容を確認したいときや物語から置いてきぼりにされてしまったとき、物語の内容を効率良く確認できると便利である。そこで、本プロジェクトでは最初から読み返すことなく物語の内容を確認できるシステムの開発を目指す。

2. 目的

本プロジェクトの目的は、大量の文章を分析して知識を抽出し、ユーザとシステムが対話を進めることで、内容の効率良い理解を可能とすることである。本プロジェクトでは物語を対象とし、登場人物の動きに注目して文の構造を解析、視覚化することで効率良く内容を確認できるシステムを開発する。ユーザは時系列で登場人物の関係が変化する様子を閲覧できる。また、ユーザが人名でない単語や表示したくない単語を選択すると、システムが結果を更新する。これにより、ユーザは難しい操作をすることなく、システムが全自動で抽出した結果を修正することが可能となる。

3. 開発の内容

本プロジェクトは Web アプリケーションとしてシステムを実装した。Web アプリケーションにしたのは、ソフトを配布したりインストールしたりする必要がなく、Windows, Mac, Linux といった OS に依存せずに利用できるようにするためである。また、動作確認は Google Chrome および Firefox で行なっている。

テキストを入力すると、システムは図 1 のように結果を表示する。左上が時系列を移動するバー、左下が登場人物の関係図、右が登場人物のリストである。ユーザは、時系列を移動するバーを動かすことで、登場人物の関係が変化していく様子を閲覧できる。ここでは、村上春樹「1Q84 book1」の結果を例に説明する。例えば、図 1(上)は 3 章までの登場人物の関係図である。ここで、シークバーを右に動かすことで任意の章までの関係図に変更できる。図 1(下)は 7 章までの登場人物の関係図である。新たに「タマル」が登場したり、「川奈天吾」と「ふかえり」につながりができたりすることが確認できる。これは当然、本文の内容とも一致している。実際、タマルは 7 章で初めて登場する人物である。また、川奈天吾とふかえりは小松を通じた知り合いの知り合いという関係だったのに対し、4 章以降で実際

に会っている。

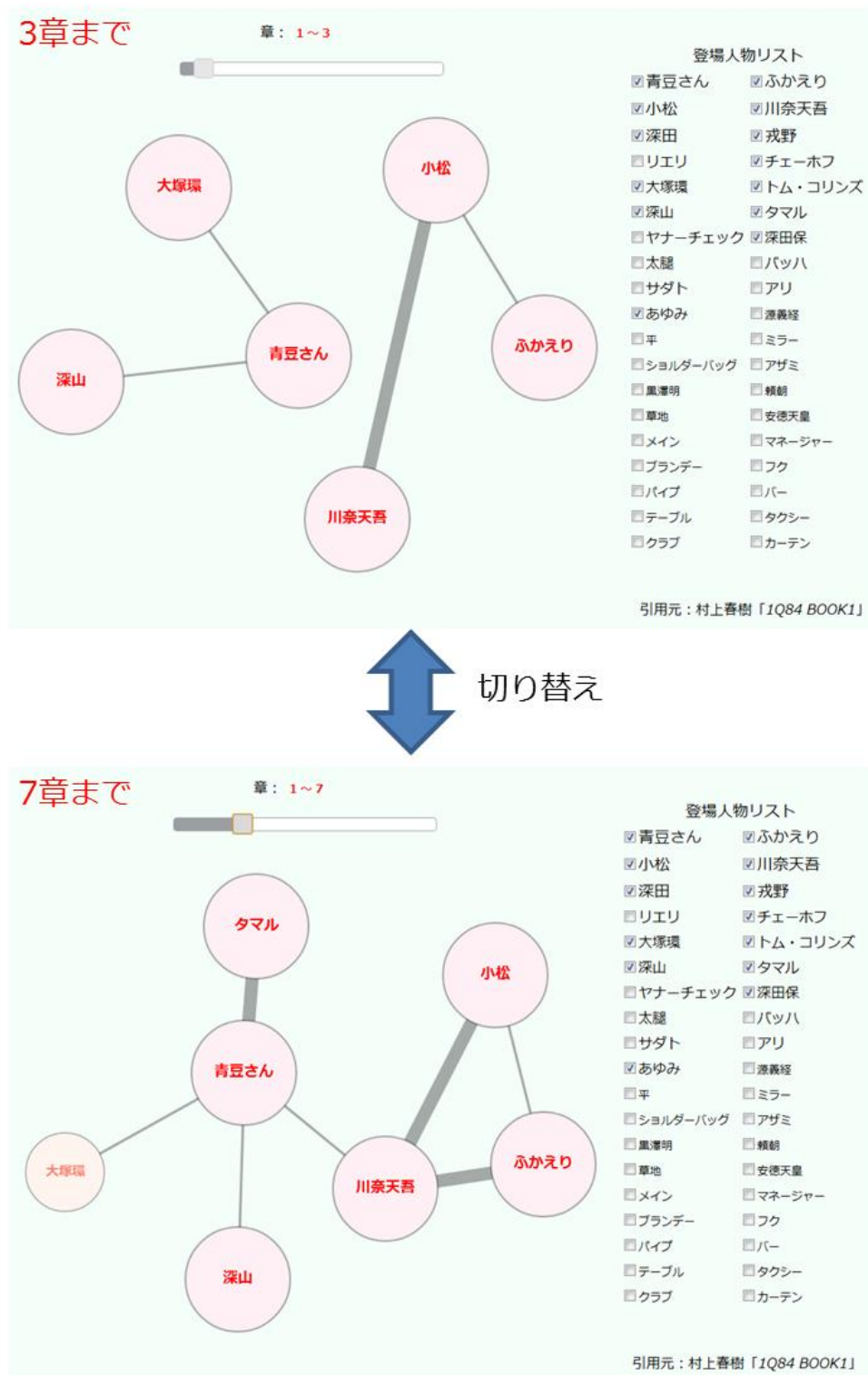


図 1:登場人物の関係図(上:3 章までの関係、下:7 章までの関係)

登場人物の関係図の各ノードはドラッグすると動かすことができる。また、ダブルクリックすると図 2 のように新たなウインドウが開き、ノードに書かれていた人名の入った本文を章ごとに閲覧できる。章の選択は、上のシークバーを動かしたり、タブをクリックしたりすること

tea.seman.cs.uec.ac.jp/cgi-bin/ui/graph/1q84/pop.py?no=62 - Google Chrome

tea.seman.cs.uec.ac.jp/cgi-bin/ui/graph/1q84/pop.py?no=62

「戒野」の関連文

←前半 シークバーで移動 後半→

1章 2章 3章 4章 5章 6章 7章 8章 9章 10章 11章 12章

13章 14章 15章 16章 17章 18章 19章 20章 21章 22章 23章

24章

- しかしそれはそれとして、あの戒野先生がふかいりの保護者になっているとは思ってもみなかったな。
- 「俺と戒野先生は昔一緒に仕事をすることがあるんだ。

また、表示するノードを任意に選択することが可能である。登場人物のリストのチェックボックスを切り替えるだけで、登場人物の関係図も更新される。この機能により、主要な登場人物だけに限って登場人物の関係を閲覧したり、システムが誤って検出した人名をユーザが容易に編集したりできる。ここで、主要な登場人物ではない「サダト」や誤って検出された「テーブル」などを非表示にした結果を図 3 に示す。このように、グラフの情報量を削減することでより見やすい登場人物の関係を構築できる。



文章を視覚化するシステムとしては、Wordle, Phrase Nets, TextArc, ThemeRiver, Docuburst などが存在する。Wordle は単語がランダムに配置されるシステムで、単語同

士の関係が分からないという問題がある。Phrase Nets はユーザが指定した関係を読みやすい形で表示できるが、X of Y といったパターンを正規表現で抽出しているため、日本語で利用するのは難しい。例えば、日本語で「XのY」を正規表現で抽出しようすると、「ベルボーイの田中」や「東京の美術館」といった助詞「の」で挟まれている表現だけでなく、「その通りだ」「なぜ殺すのだ」といった表現まで抽出してしまう。本システムではこれらの問題を解決するために、登場人物の関係を正規表現ではなく文の構造を考慮して抽出する。また、本システムのように時系列で登場人物の関係を閲覧したり、任意のノードの表示をクリック1つで切り替えたりできるものは他に存在しない。

5. 期待される効果

本システムを利用すると、過去に読んだ本の内容の想起支援が可能である。例えば、村上春樹「1Q84 book1」の内容を確認したいとき、最初から読み返すと個人差はあるものの3時間以上かかってしまう。しかし、登場人物の関係を閲覧するのにかかる時間はわずか数分である。さらに、物語から置いてきぼりをされたときの支援もできる。例えば、「××」という登場人物のことを忘れてしまった場合、本システムを利用することで「××」と関連度の強い登場人物を確認したり、「××」の入った文を章ごとに関連したりできる。このような操作は既存の電子書籍リーダーでは不可能で、より効率良く「××」という登場人物のことを理解できるようになった。

6. 普及(または活用)の見通し

今回の開発成果をもとに論文を執筆し、学会で発表したいと考えている。加えて、Webアプリケーションを一般公開し、多くの人から意見をもらってシステムを改良していきたい。また、本プロジェクトの開発成果は電子書籍リーダーと親和性が高い。電子書籍リーダーを使用して本を読む際、読んだところまでの登場人物の関係を表示できると便利である。そのため、電子書籍リーダーの追加機能として実装することも検討したい。

7. クリエータ名(所属)

田中 翔太郎(電気通信大学大学院情報理工学研究科)