

登場人物と場所の時系列可視化による 物語の出来事の想起支援インタフェース

A Support Interface for Remembering Events in Novels by Visualizing Time-series Information of Characters and their Existing Places

西原陽子^{1*} Jiaxiu Ma² 山西良典³
Yoko Nishihara¹ Jiaxiu Ma² Ryosuke Yamanishi³

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² 立命館大学大学院情報理工学研究科

² Graduate School of Information Science and Engineering, Ritsumeikan University

³ 関西大学総合情報学部

³ Faculty of Informatics, Kansai University

Abstract: This paper proposes a support interface for remembering events in novels by visualizing time-series information of characters and their existing places. The target users of the proposed interface are people who read multiple novels in parallel. The proposed interface consists of two parts: (a) displaying of novel texts and (b) visualizing of time-series information of character names and their existing places. Given a novel text, the proposed interface extracts character names and places while the Conditional Random Fields (CRFs) is used for the extraction. The proposed interface sets the horizontal axis as the number of the sentence and the vertical axis as characters. The existing places are mapped as labelled dots on the plane given by the two axes. We experimented with participants and verified that the proposed interface could support to remember the events in novels.

1 はじめに

スマートフォンやタブレット、さらに専用の端末 (Amazon kindle など) を用い、電子的に書籍を読む機会が増えた。紙の本と異なり電子書籍は複数冊を所持していても重くないため、携帯が容易にできる。この利点から複数の書籍を並行して読み進めることも可能であり、このような読書形態は今後も増えていくことが予想される。

読書にまとまった時間を確保することが難しく、隙間時間を縫って読書を進める人も多いと考えられる。読書の間隔が空いてしまうと、読書を再開するときに既読部分の内容を忘れてしまい、読み返しをせねばならないことも出てくるだろう。特に多くの人物が登場する物語では、登場人物に関わり起こった出来事が分かりにくくなる。さらに複数の物語を並行して読み進め

ていると、既読部分の内容をより忘れがちになると考えられる。忘れた内容を思い出すための読み返しの量が増えると、読書を先に進めることが難しくなり、結果として読書を楽しむことが難しくなる可能性がある。

そこで本研究では、複数の物語を同時に読み進める人を対象とし、登場人物と場所の時系列情報を可視化することにより、物語の出来事の想起を支援するインタフェースを提案する。物語で記述された順に登場人物と、登場人物がいた場所を可視化することにより、既読部分において登場人物がどこで何をしたか、あるいは何を考えたかを想起しやすくする。これによりスムーズな読書の再開を支援することを目標とする。

長い物語を読む際の読書支援として、登場人物の特徴を示す研究や人物間の関係を示す研究などがある [1, 2]。登場人物の特徴を示す研究として、登場人物の初登場シーンを表示するインタフェースが提案されている [5]。提案インタフェースを用いて読書をするユーザが登場人物名を選択すると、その人物の初出のシーンを表示

*連絡先: 立命館大学情報理工学部
滋賀県草津市野路東 1-1-1
E-mail: nisihara@fc.ritsumeai.ac.jp

する。この研究では登場人物の特徴を失念したときに、すぐに初出のページへと飛び、内容を確認することができるので、素早い読み返しを支援できると考えられる。本研究では本文を読み返すことなく、物語の内容の想起を支援する手法を提案する。

読み返しをさせずに内容の想起を支援する研究として、登場人物間の関係を示すものがある。物語のテキストから登場人物を抽出し、家族関係、仲間関係といった登場人物間の関係を推定するものや [3], ユーザが読書しながらインタフェース上で編集した登場人物の関係を見せるものがある [6]。これらの研究では物語の進捗とともに得られる人間関係やその変化を示すことで、読書の支援を行っている。本研究では物語は登場人物に関わる出来事が順に紡がれてできると考え、既読部分の出来事を示すことにより内容の想起を支援し、読書の支援を行う。出来事はいつ、どこで、誰が、何を、どうしたにより表現されると考え、これらのうちいつ (文番号)、どこで (場所)、誰が (登場人物) の 3 つの情報を抽出し表示することで出来事を示す。

2 提案インタフェース

提案インタフェースの概要を示す。提案インタフェースは物語のテキストを表示する部分と、登場人物と場所の時系列情報を可視化する部分の 2 つにより構成される。提案インタフェースは Web ブラウザ上で動作し、描画には JavaScript のライブラリである D3.js を用いる。

物語のテキストを表示する部分を図 1 に示す。左メニューから読む物語を選択すると、右ウィンドウに物語のテキストが表示される。

登場人物と場所の時系列情報を可視化する部分を図 2 に示す。物語テキストの表示部の右ウィンドウ最上部にある物語のタイトルをクリックすると、時系列情報の可視化結果が表示される。可視化では横軸に文番号を取り、縦軸に登場人物の名前を取る。横軸と縦軸で張られた平面上に人物が居たと考えられる場所がプロットされ、場所名をラベルとして付与する。

可視化図を得るまでのフローチャートを図 3 に示す。入力されるものは物語のテキストである。テキストを文ごとに分割し、文から登場人物と場所表現を抽出する。続いて、文ごとに登場人物と場所表現の紐付けを行い、文番号、登場人物名、登場人物がいた場所の 3 つ組のデータを作成する。3 つ組の情報を基にし、登場人物と場所の時系列情報の可視化を行う。既読部分までの可視化を見ることにより、ユーザは物語内で起こった出来事を人物や場所の情報をヒントとして想起することができると思われる。

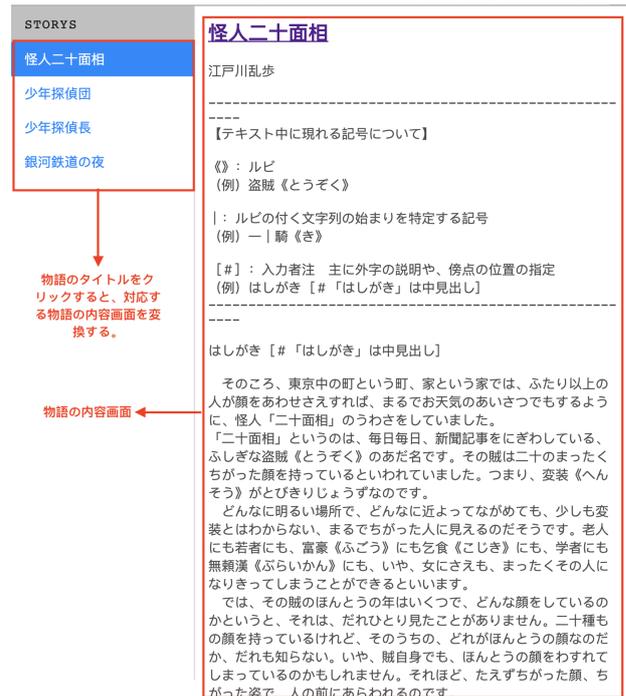


図 1: 提案インタフェースの物語テキストの表示部分

2.1 登場人物と場所表現の抽出

物語のテキストから登場人物と場所表現を抽出する方法を説明する¹。本研究では固有表現抽出に用いられる Conditional Random Fields (CRF) を用い、登場人物と場所表現を抽出する。

はじめに物語テキストを句点 (。や。) により分割し、文の集合を作る。続いて文を形態素解析し、単語と品詞の情報を得る。形態素解析には MeCab、辞書は NEologd を用いる。

続いて人手により各文に含まれる登場人物と場所表現にアノテーションを行う。アノテーションにあたっては BIO2 タグを用いる。BIO2 タグは固有表現抽出に用いられるタグであり、固有表現の先頭の形態素には B タグ、先頭以外の形態素には I タグ、固有表現以外には O タグが付与される。本研究では登場人物と場所表現をそれぞれ分けて抽出を行うため、表 1 に示す 5 種類のタグを用いる。登場人物と場所表現に対するアノテーションを行い、CRF で学習を行うための訓練データを作成する。登場人物は本文に含まれる単語やフレーズのうち、登場人物を明示するものに限定し、代名詞は含まない。同様に場所表現も本文に含まれる単語やフレーズのうち、建物や地名を明示するものに限定し、代名詞は含まない。

訓練データを CRF で学習する際は i 番目の形態素を

¹登場人物と場所表現の抽出については発表済みの内容と同じである [4]。

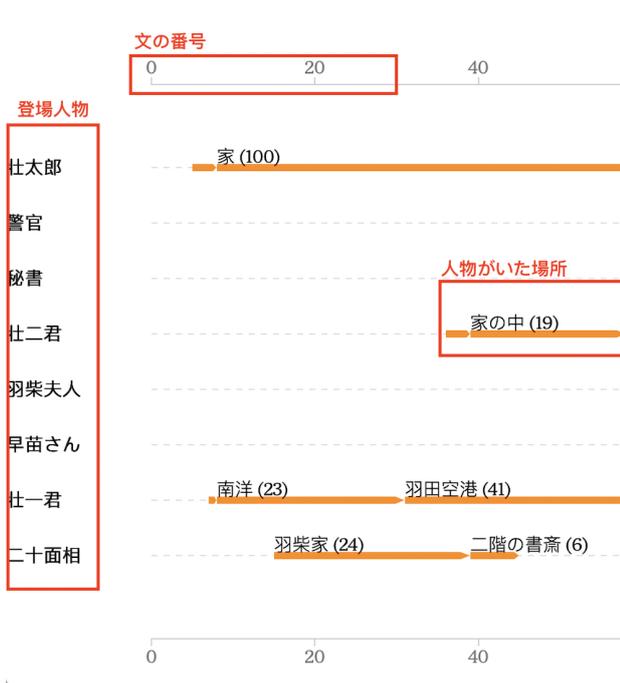


図 2: 提案インタフェースの登場人物と場所の時系列情報の可視化部分

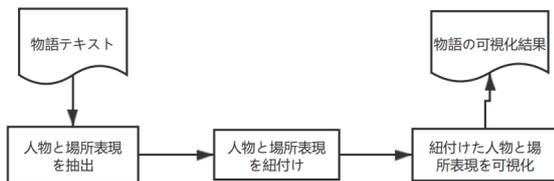


図 3: 可視化図を得るまでのフローチャート

中心とし、前後 3 つの形態素を合わせた合計 7 つの形態素に対する、形態素、品詞情報、BIO2 タグの情報をを用いる。CRF により固有表現抽出器を獲得し、物語テキストからの登場人物と場所表現の抽出を行う。

2.2 登場人物がいる場所の推定

抽出された登場人物と場所表現を紐付け、登場人物がいる場所の推定を行う。一つの文の中に登場人物の表現と場所の表現が含まれる場合、その登場人物と場所には関連があり、登場人物がその場所にいる可能性が高いと考えられる。この仮説の妥当性を示すために予備的な調査を行った。対象とした物語は「桃太郎」「赤ずきんちゃん」「銀河鉄道の夜」であった。

表 1: 人物と場所表現を示す BIO2 タグ

表現タグ	説明
B-CHAR	人物表現文字列の始まり
I-CHAR	人物表現文字列が続いている
B-POS	場所表現文字列の始まり
I-POS	場所表現文字列が続いている
O	人物と場所表現以外の形態素

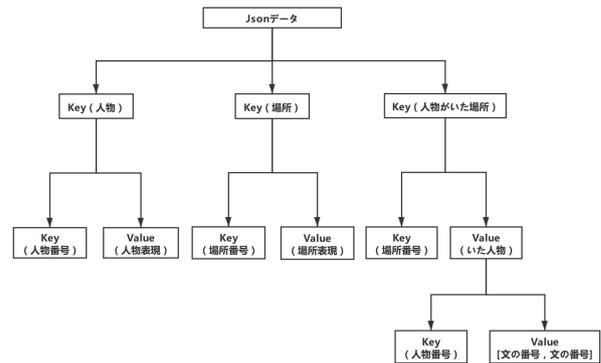


図 4: 提案インタフェースで用いる json データの形式

3 つの物語から登場人物表現のみが抽出されたのは 338 文、場所表現のみが抽出されたのは 73 文、両方が抽出されたのは 99 文、いずれも抽出されなかったのは 322 文であった。両方が抽出された 99 文について登場人物と場所の関連性を手作業にて調べたところ、71% の文で登場人物がその場所にいたことが確認された。このことから一つの文の中に登場人物と場所表現の両方が含まれる場合には、登場人物がその場所に居たと仮定し、紐付けを行うこととする。

登場人物は一文ごとに場所を移動するとは考えにくく、一定の期間、同じ場所に止まると考えられる。そこで本研究ではある文 $s(i)$ においてある登場人物 A とその登場人物 A がいた場所 B が抽出された時、別の文 $s(j)$ (ただし、 $i < j$) において登場人物 A がいた別の場所 C が抽出されるまで、登場人物 A は場所 B に滞在していたとする²。

2.3 登場人物と場所の時系列の可視化

文番号、登場人物、登場人物がいた場所の 3 つ組のデータを受け取り、登場人物と場所の時系列の可視化を行う。3 つ組の情報のデータは json 形式で作成する。図 4 に提案インタフェースで用いる json データの形式を示す。

²この仮説の検証は今後の課題とする。

3 評価実験

電子的に複数の物語を読み進めるユーザに対し、提案インタフェースが既読部分の出来事の想起を支援できるかを評価する実験を行った。

3.1 実験手順

以下の手順で実験を行った。

1. 被験者を3つのグループに分ける。それぞれグループ A, B, C とする。
2. 被験者は1日の間に、指定されたインタフェースで4つの物語の冒頭から指定された部分までを読む。
3. 1週間後に各グループの被験者は物語の出来事に関するテストを受験する。
4. 実験者は得られたテストの平均正解率を各グループで比較し、提案インタフェースの評価を行う。

3.1.1 被験者

被験者は20代の大学生、大学院生で日本人が9名、日本語能力検定試験 N1 の資格を持つ中国人が12名、合計21名であった。全員が日本語の文章を読み、理解する能力を十分に有していた。各グループに日本人を3名、中国人を4名の合計7名ずつ割り当てた。

3.1.2 使用したインタフェース

実験手順の2.で物語を読む際に使用したインタフェースについて説明をする。インタフェースは物語のタイトルと冒頭から指定された部分までのテキストを Web ブラウザを通じて表示するものであり、図1で示したものと同一ものを使った。

3.1.3 使用した物語

物語は青空文庫から数回に分けて読むと予想される長編のものを選択した。表2に実験で用いた物語の情報を示す。「少年探偵団」「少年探偵長」「怪人二十面相」「銀河鉄道の夜」の4つを用いた。提案インタフェースのユーザは隙間時間を縫って読書をする想定しており、1回の読書を20分と仮定し、平均的な読書スピードが1分間に400文字であることから積をとり、8,000文字程度で段落の切りの良い部分で区切りをいれた。その結果、読書する文字数は6,000文字から10,000文字までとなった。

表 2: 実験で用いた物語の情報

タイトル	作者	読書指定文字数
少年探偵団	江戸川乱歩	9,979
少年探偵長	海野十三	6,101
怪人二十面相	江戸川乱歩	8,907
銀河鉄道の夜	宮沢賢治	8,903

3.1.4 物語の内容に関するテスト

提案インタフェースでは登場人物と場所の情報を時系列で可視化することにより、物語の出来事の想起を支援するため、人物や場所に関する出来事についての設問からなるテストを作成した。提案インタフェースにより可視化された登場人物と場所の時系列情報を見ることにより、誰がどこで何をしたかを表す出来事の情報想起できたかを確認するテストを作成した。表3に実験で使用したテストの一部を示す。問題は全て物語中の出来事とつながっているものを作成した。例えば問題「ジョバンニは帰る前に何を買った？」は、物語の中で登場人物があるものを買った出来事とリンクしている。人物に関する問題は「1人の人物に関すること」「複数の人物に関すること」であり、場所に関する問題は「1つの場所に関すること」「複数の場所に関すること」とした。複数の人物に関する問題は「Aさんと場所Bにいたのは誰ですか？」が例になり、複数の場所に関する問題は「Aさんは場所Bからどこへ移動しましたか？」などが例になる。1人の人物や1つの場所の問題よりも、複数の人物や場所に関する問題は覚えていなくてはならない人物数や場所数が増えるため、問題としては難度が高くなっている。設問は1つの物語につき5問あり、合計で20問であった。20問のうち12問は記述式、残りの8問は3択の選択肢式であった。テストの受験時間は20問で5分間であった。

テストを受験する際に各グループで差をつけた。グループAはテストを受験する際に、読み返しを一切行わないように指示した。グループBは3分間の読み返しを行った上で、テストを受験するように指示した。グループCは3分間登場人物と場所の時系列が可視化された提案インタフェースを3分間使った上で、テストを受験するよう指示した。

3.2 評価方法

評価は各グループの被験者のテストの平均正解率を用いて行った。グループAとグループBの平均正解率を比較することにより、読み返しが出来事の想起支援に与える影響を明らかにする。続いて、グループAと

表 3: 評価実験用のテスト問題の一部（問題数の括弧内の数字は選択肢式の問題の数）

問題の種類	問題数	問題の例
1人の人物	9(5)	ジョバンニは帰る前に何を買った？(銀河鉄道の夜)
複数の人物	3(1)	春木少年は誰と一緒にカンヌキ山の頂上を登った？(少年探偵長)
1つの場所	5(0)	羽柴氏の宝石はどこに置いている？(怪人二十面相)
複数の場所	3(2)	桂くんはどこから明るい町のほうへかけだした？(少年探偵団)

表 4: 各グループのテストの平均正解率（カッコ内の数字はグループ A からの増減）

グループ	平均正解率
A	26%
B	36%(+10%)
C	54%(+28%)

グループ C の平均正解率を比較することにより、提案インタフェースによる登場人物と場所の時系列の可視化が出来事の想起支援に与える影響を明らかにする。

3.3 実験結果

表 4 に各グループのテストの平均正解率を示す。20 問中の正答数で正解率を出し、各グループの 7 人の平均をとった。提案インタフェースを用いるグループ C の平均正解率が最も高くなった。

問題の種類ごとの平均正解率を表 5 に示す。いずれの問題の種類においても提案インタフェースを用いるグループ C の正解率が最も高くなった。

4 考察

表 4 の全体的な平均正解率を比較すると、グループ A から B にかけての増分は+10%であった。このことから読み返しにより出来事の想起が支援されることがわかった。さらにグループ B からグループ C にかけての増分は+28%であった。この増分は読み返しによる増分よりも多かった。このことから読み返しよりも提案インタフェースを用いる方が、出来事の想起が支援されることがわかった。

続いて人物に関する問題の正解率について考察をする。「1人の人物に関する問題」と「複数の人物に関する問題」の平均正解率を比較すると、「複数の人物に関する問題」に対する方が低かった。この原因としては2つのことが考えられる。1つは、1人の人物についての情報は覚えていても、複数の人物になるとその関連を正確に把握することが難しくなったことがあると考えられる。読み返しや提案インタフェースの使用により関連を把握することができたため、グループ B と C の正解率は増加したと考えられる。もう1つの理由は、選択肢式の問題の割合が挙げられる。「1人の人物に関する問題」は9個中5個が選択肢式であったのに対し、「複数の人物に関する問題」は3個中1個が選択肢式であり、選択肢式の問題の割合がより低かった。このため相対的にグループ A の「複数の人物に関する問題」の平均正解率が低くなり、グループ B や C の増分が多くなったと考えられる。

続いて、場所に関する問題の正解率について考察をする。「1つの場所に関する問題」と「複数の場所に関する問題」の平均正解率を比較すると、「1つの場所に関する問題」に対する方が低かった。この原因としては先と同じく選択肢式の問題の割合が挙げられる。ただ、全てを記述式とした場合にグループ C の正解率は40%となったことから、問題文の選択肢のヒントがない状態で出来事を思い出せるのは高々40%程度となると予想される。今後、クイズの問題を改訂し評価実験を行うことで、人物に関する出来事についても同様の正解率となるかを検証し、提案インタフェースの支援可能性について明らかにしていきたい。

5 おわりに

本論文では複数の書籍を同時に読み進める人を対象とし、登場人物と場所の時系列を可視化することにより物語の出来事の想起を支援するインタフェースを提案した。提案インタフェースは物語のテキストを表示する部分と、登場人物と場所の時系列情報を可視化する部分の2つにより構成される。物語のテキストの各文から登場人物と場所表現を抽出し、文番号と登場人物、登場人物がいた場所の3つ組のデータを作成し、横軸に文番号、縦軸に登場人物名をとった2次元平面に人物がいた場所がプロットされ、場所名を線のラベルとして付与する。提案インタフェースを用いると、既読部分についての出来事の想起が支援され、スムーズな読書の再開の支援が期待される。

提案インタフェースの評価実験では被験者に4つの物語を読ませ、1週間後に物語の内容について問う制限時間付きのテストを実施し、平均正解率を比較した。テストを受験する際に何も使わないグループ A、読み

表 5: 問題の種類ごとの平均正解率 (カッコ内の数字はグループ A からの増減)

グループ	1 人	複数人	1 場所	複数場所
A	40%	14%	6%	33%
B	38% (-2%)	43% (+29%)	20% (+14%)	52% (+19%)
C	57% (+17%)	62% (+48%)	40% (+34%)	62% (+29%)
平均	45%	40%	22%	49%

返しを行うグループ B, 提案インタフェースを用いるグループ C と被験者をグループ分けし, グループごとに平均正解率を比較したところ, 提案インタフェースを用いるグループ C の平均正解率が最も高くなったことから, 既読部分についての出来事の想起を最も支援できることがわかり, 提案インタフェースの有用性を確認した. 今後, テストの問題数と回答形式を揃えた上で評価実験を再度行い, 提案インタフェースの支援の限界について評価をする.

ンコンピュータインタラクション研究会, pp.1-5, (2017).

- [6] 田中翔太郎, 岡部誠, 尾内理紀夫, 物語の内容想起支援インタフェースの開発, 第 55 回プログラミング・シンポジウム予稿集, pp.7-15, (2014).

謝辞

本研究の一部は科研費 (20K12130) によって行われました. 記して謝意を申し上げます.

参考文献

- [1] Apoorv Agarwal, Anup Kotalwar, and Owen Rambow, Automatic Extraction of Social Networks from Literary Text: A Case Study on Alice in Wonderland, Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 1202-1208, (2013).
- [2] David Elson, Nicholas Dames, and Kathleen McKeown, Extracting Social Networks from Literary Fiction, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 138-147, (2010).
- [3] 西原弘真, 白井清昭, 物語テキストを対象とした登場人物の関係抽出, 言語処理学会第 21 回年次大会, pp.628-631, (2015).
- [4] Jiaxiu Ma, 西原陽子, 山西良典, 物語内の人物と場所情報の時系列可視化による読書支援, 第 23 回インタラクティブ情報アクセスと可視化マイニング研究会, pp.9-12, (2019).
- [5] 謝涵, 西田健志, 物語の登場人物を把握しやすくするシステムの提案, 情報処理学会第 174 回ヒューマ