

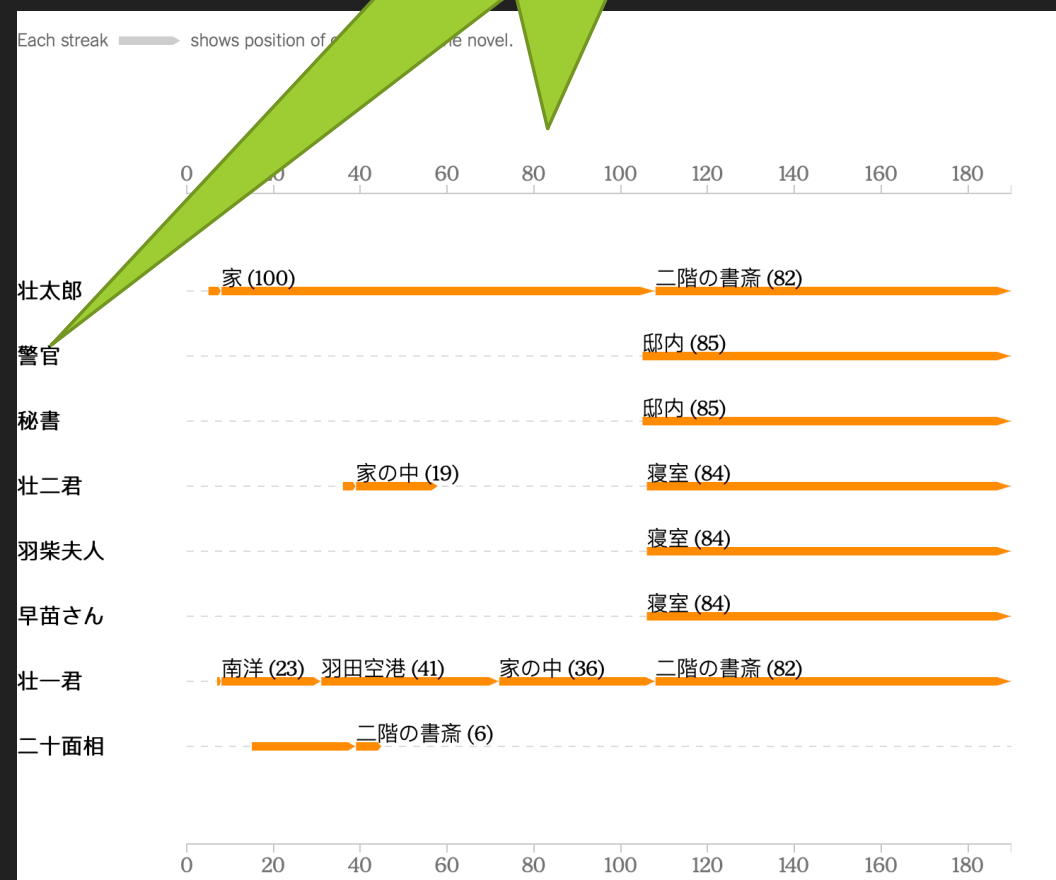
登場人物と場所の時系列可視化 による物語の出来事の想起支援 インタフェース

西原陽子・Jiaxiu Ma（立命館大学）・山西良典（関西大学）

本研究で実現したこと

横軸が文番号
縦軸が登場人物
登場人物が居たと思われる場所を平面上に記載

- 複数の物語を並行して読み進めるユーザを対象とし、登場人物と場所の時系列情報を可視化するインタフェースを提案
- 本インタフェースにより、既読部分の出来事の想起を支援
- 読み返しが不要となり、読書の再開がスムーズに行われる



参考にした可視化は世界史の年表

研究背景：電子的な読書

- 電子端末を用い、書籍を読む機会が増加
- 複数の書籍を並行し読み進める形態が増加すると考えられる
- 読書の間隔が空くと、既読部分の内容を忘れがち
 - 特に登場人物が多い物語では顕著
- 読み返しの頻度が多くなると、読書を先に進めにくくなる

研究目的

- 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
- ターゲットユーザは、複数の物語を並行して読み進める人
- 可視化により既読部分での登場人物に関わる出来事を想起しやすくする
 - 誰がどこで何をしたか、何を考えたかなど

提案インタフェースの概要

(a) 物語のテキストを表示する部分

The screenshot shows a web interface for reading a story. On the left, a sidebar titled 'STORIES' contains a list of titles: '怪人二十面相', '少年探偵団', '少年探偵長', and '銀河鉄道の夜'. The '怪人二十面相' title is highlighted in blue. A red arrow points from this title to the main content area with the text: '物語のタイトルをクリックすると、対応する物語の内容画面を交換する。' (Clicking the story title exchanges the content screen for the corresponding story). Below the sidebar, another red arrow points to the main content area with the text: '物語の内容画面' (Story content screen). The main content area displays the title '怪人二十面相' and the author '江戸川乱歩'. It includes a section for '【テキスト中に現れる記号について】' (About symbols appearing in the text) with examples for symbols like 《》, |, and [#]. The main text begins with 'はしがき [# 「はしがき」 は中見出し]' (Foreword [# 'Foreword' is a sub-heading]). The text continues with a paragraph: 'そのころ、東京中の町という町、家という家では、ふたり以上の人が顔をあわせさえすれば、まるでお天気のあいさつでもするように、怪人「二十面相」のうわさをしていました。' (At that time, in a town called 'All towns in Tokyo', in a house called 'House', whenever two or more people met, it was as if they were exchanging weather greetings, the rumors of the 'Twenty Faces' were spreading.)

既読部分の
出来事を確認



読書再開



(b) 登場人物と場所の時系列情報を可視化する部分



登場人物と場所の時系列情報を得るまでの処理の流れ

1. 物語の各文から登場人物と場所表現を抽出する
2. 登場人物が居る場所を推定する
3. 文番号、登場人物、場所からなる3つ組データを用い、可視化する

1. 登場人物と場所表現の抽出

○ 固有表現抽出に用いられるCRFsを用いて抽出する

1. テキストを形態素解析 (MeCab+NEologd) し、単語と品詞情報を得る

2. 登場人物と場所表現にBIO2タグを用いアノテーションし、コーパスを作成。CRFsを用いて抽出器を得る

3. 1とは別のテキストに対し抽出器を適用し、登場人物と場所表現を抽出する

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

位置	入力単語	品詞細分類	表現タグ
i-3	これ	名詞-代名詞 一般	O
i-2	を	助詞-格助詞 一般	O
i-1	赤ずきん	名詞-固有名詞 一般	B-CHAR
i	ちゃん	名詞-接尾-人名	I-CHAR
i+1	,	記号-読点	O
i+2	ここ	名詞-代名詞 一般	O
i+3	に	助詞-格助詞 一般	O

2. 登場人物が居る場所の推定(1/2)

- 一つの文の中に登場人物Aと場所表現Bが含まれる場合、AとBには関連があり、登場人物Aが場所Bにいる可能性が高いと考えられる
- 実際の物語で上の仮説を検証したところ、登場人物Aが場所Bにいる割合は0.71であった

物語タイトル	人物	場所	人物と場所	両方がない	一致率	
赤ずきんちゃん	34	12	12	24	0.67	短編物語
桃太郎	53	17	32	24	0.75	短編物語
銀河鉄道の夜	251	44	55	274	0.71	長編物語

2. 登場人物が居る場所の推定(2/2)

- 登場人物は一定の間同じ場所に居る可能性が高いと考えられる
- ある文 $s(i)$ で登場人物Aと場所Bが抽出されたとき、別の文 $s(j)$ ($i < j$)で登場人物Aと別の場所Cが抽出されるまで登場人物Aは場所Bに居たとする



3. 文番号、登場人物、場所からなる 3つ組データを用いた可視化



提案インタフェースの評価実験概要

○実験手順

- 1. 被験者は1日の間に4つの物語を冒頭から指定された部分まで読む
- 2. 1週間後、被験者は物語の出来事に関するテストを受験する
- 3. テストの平均正解率を用いて、提案インタフェースを評価する

○被験者は21名、各グループ7名ずつ

- グループA：テストの際、ヒントなし
- グループB：テストの制限時間内で読み返しOK
- グループC：テストの制限時間内でインタフェースの使用OK



使用した物語と テストの内容

- 長編の物語を4篇使用
- 1回20分*400文字/分=8,000文字前後となるようキリの良い段落までを指定
- テストは人物に関することと、場所に関することを出題
 - 自由記述式と選択肢式

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

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）

実験結果：テストの平均正解率

- 提案インタフェースを使うグループCの正解率が最も高くなった
- グループAとBを比較すると、平均正解率は+10%
- グループAとCを比較すると、平均正解率は+28%
- 増分がより多いことから、読み返しよりも提案インタフェースの方が出来事の想起を支援できると確認

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%)

- 被験者は21名、各グループ7名ずつ
 - グループA：テストの際、ヒントなし
 - グループB：テストの制限時間内で読み返しOK
 - グループC：テストの制限時間内でインタフェースの使用OK

考察:人物に関する問題の平均正解率

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%

- 「人物」に関する問題の正解率は、一人の人物 > 複数の人物
- 一人の人物に関する出来事よりも、複数の人物に関する出来事の方が複雑で記憶に残りにくい
- 読み返しや提案インタフェースによって確認ができたため、正解率が高くなったと考えられる
- 問題の数と種類が均等でないため、今後の評価実験ではより精緻なテストを作成し、評価を行う

考察:場所に関する問題の平均正解率

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%

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)

- 「場所」に関する問題の正解率は、一つの場所 > 複数の場所
 - 一つの場所に関する問題は全て記述式であり、複数の場所に関する問題よりも難しかったと考えられる
- 全てを記述式とした場合の正解率が40%であることから、問題文に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%程度である可能性が高い
- 覚えていなければならない出来事に限定し、それを問うテストを作成し、提案インタフェースの支援の可能性について明らかにしていきたい

本発表のまとめ

- 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を可視化することにより、出来事の想起を支援するインタフェースを提案した
- 評価実験を行い、提案インタフェースは読み返しを行うよりも、出来事の想起を支援できる可能性がみえた
- 今後、覚えていなければならない重要な出来事に絞り、テストを作り直し、評価実験を改めて行う

世界史の年表の例

○ 各時代に各地域で何朝があったかを把握することで、世界全体の動きを把握することができる

- 時代 → 文番号
- 地域 → 登場人物
- 何朝 → 場所

