



時空間情報を含むテキストデータを を対象とした軌跡地図要約方式

石井 雄大† 岡田 龍太郎† 中西 崇文†

† 武蔵野大学データサイエンス学部データサイエンス学科

目次

- 研究背景・目的
- 提案方式
- 実験
- まとめ

目次

■ 研究背景・目的

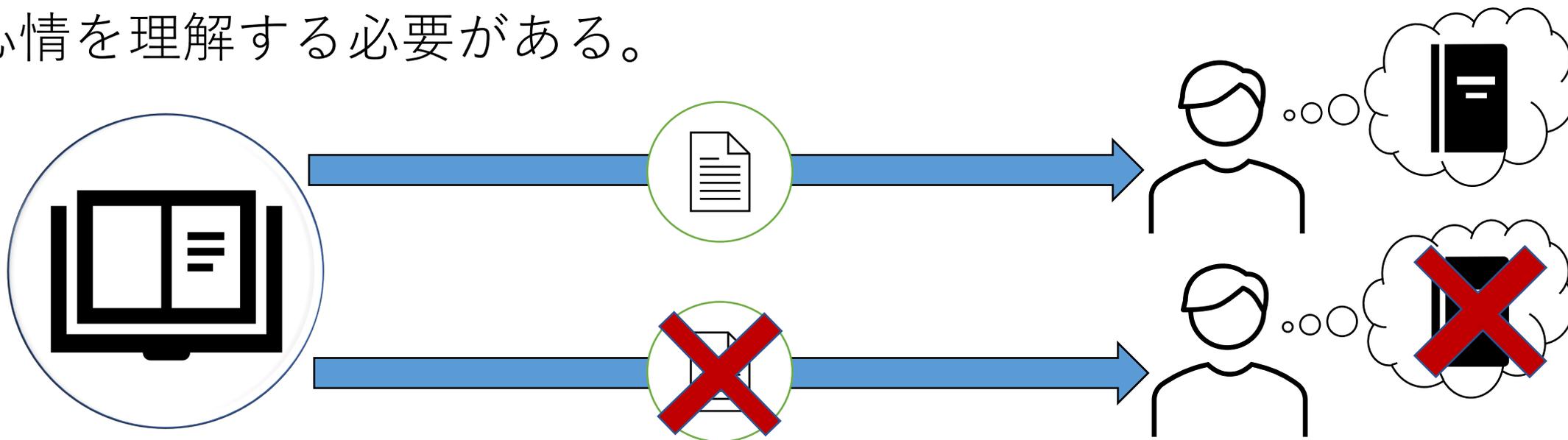
■ 提案方式

■ 実験

■ まとめ

研究背景

一般的に物語を読む際には、文章のみの情報から登場人物の心情を理解する必要がある。



文章読解が不得意であったり、苦手意識を持っている人に対しては、他メディアによる**内容理解の補助が有効**である。

研究目的

- 物語の内容理解を促進するメディアの創造
- 文章中の位置情報をもとに物語の展開をツアーのように可視化する軌跡地図メディアの自動生成

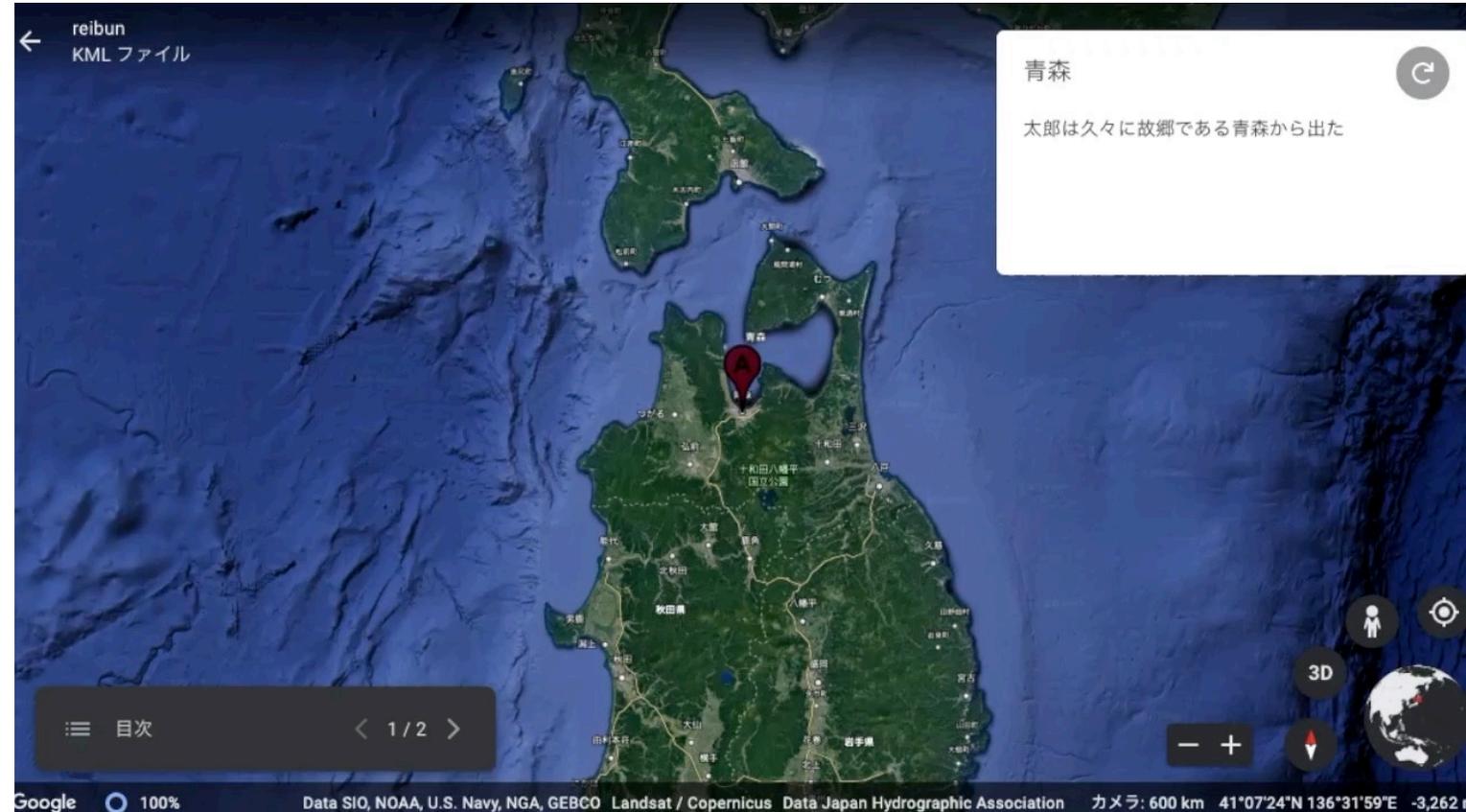


軌跡地図

物語のダイジェストとして主要人物が地点を移動する様を地図上に可視化したもの。

<例文>

太郎は久々に故郷である青森から出た。
太郎は1週間かけて東京観光を楽しんだ。



目次

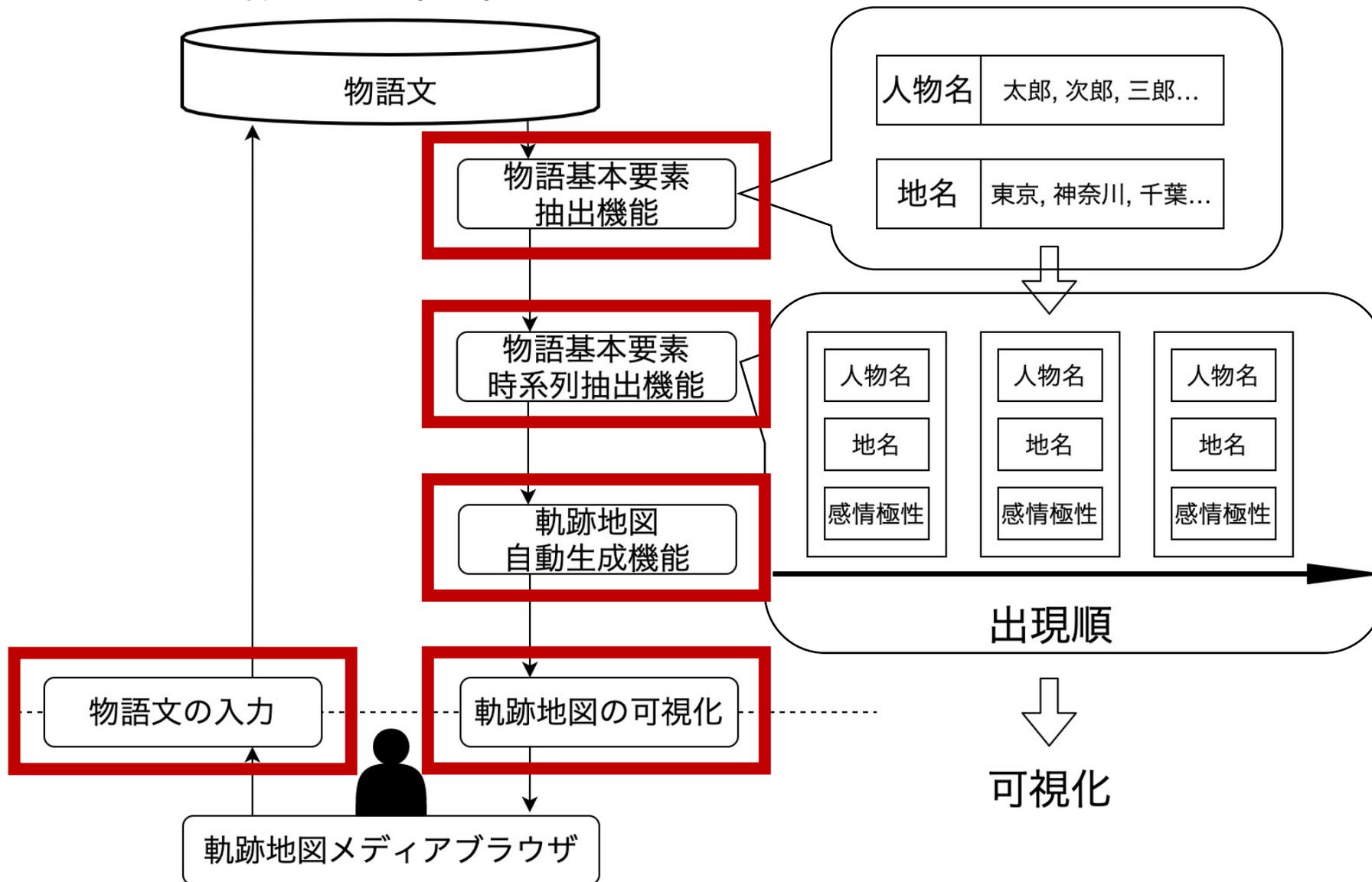
- 研究背景・目的

- 提案方式

- 実験

- まとめ

システム構成図



物語の基本要素

◆ 地名

◆ 人物名

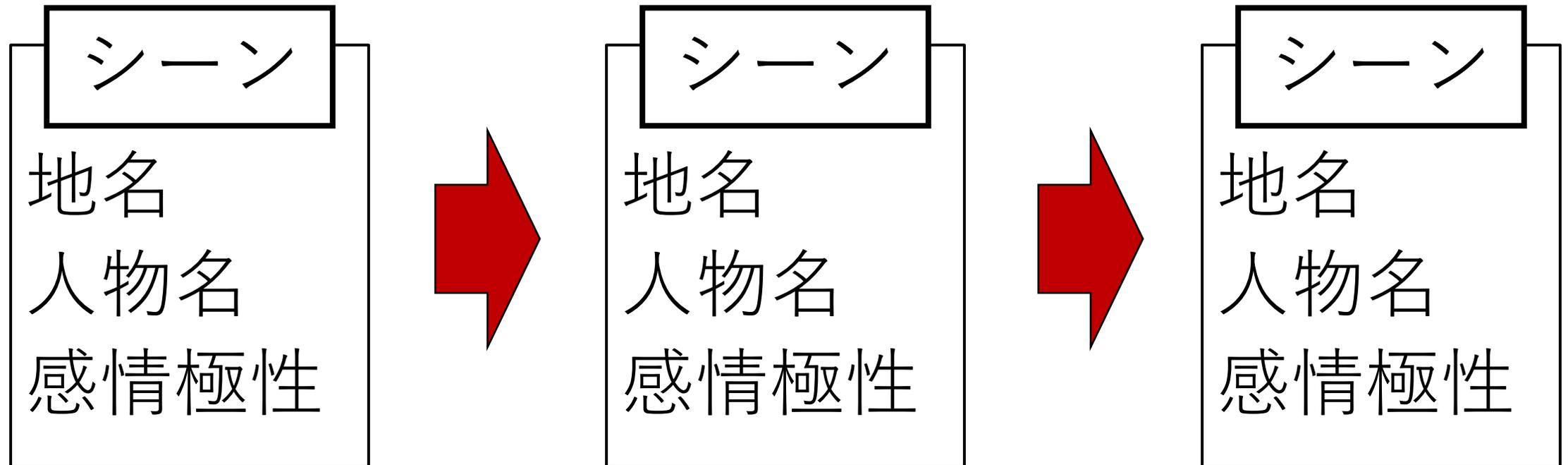
◆ 感情極性

物語の展開の理解の補助

登場人物の心情の理解の補助

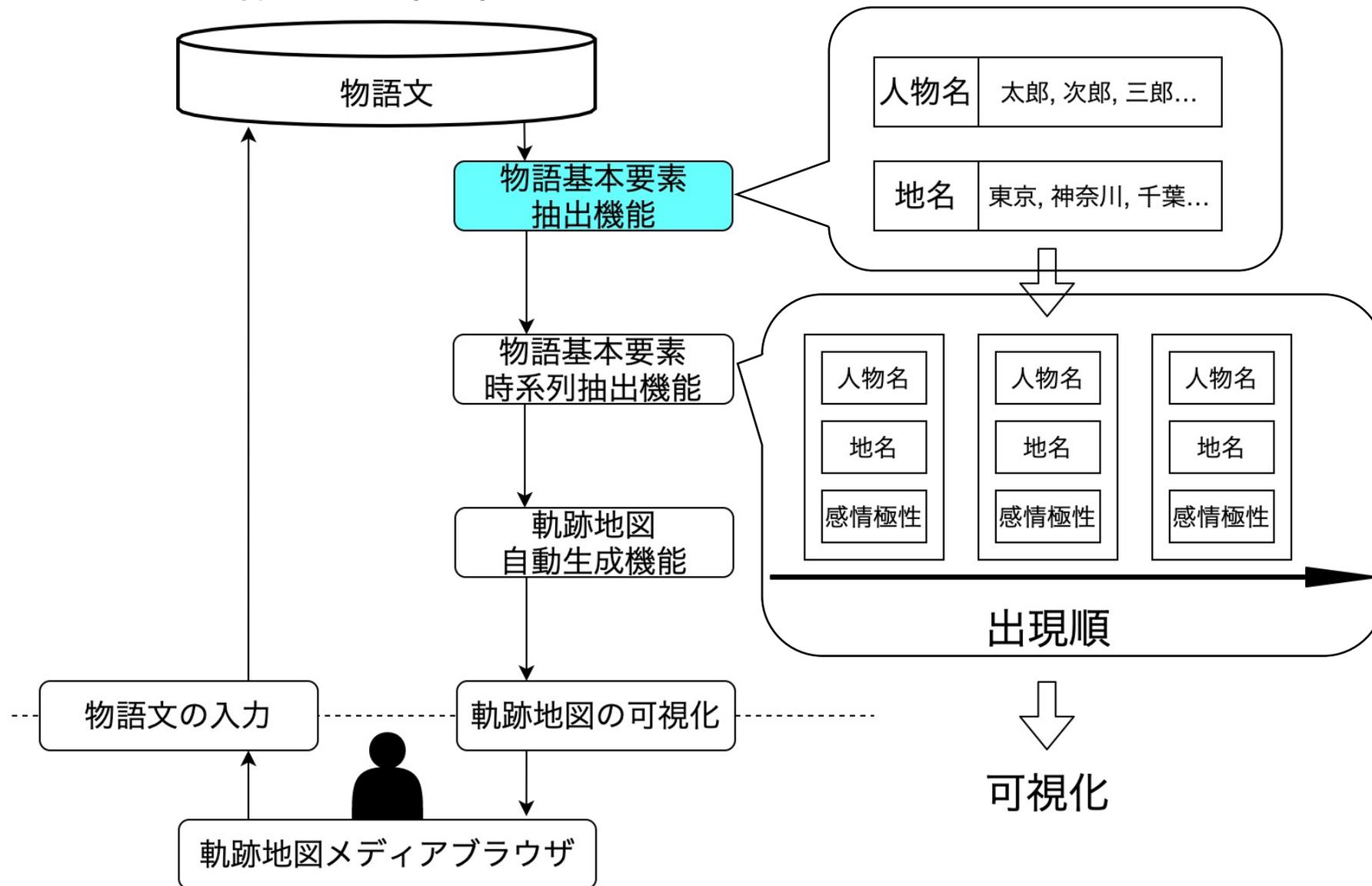
軌跡地図のデータ構造

軌跡地図は基本要素の組で表されるシーンの連なりで表現される。



シーンの連なりで物語を示す。

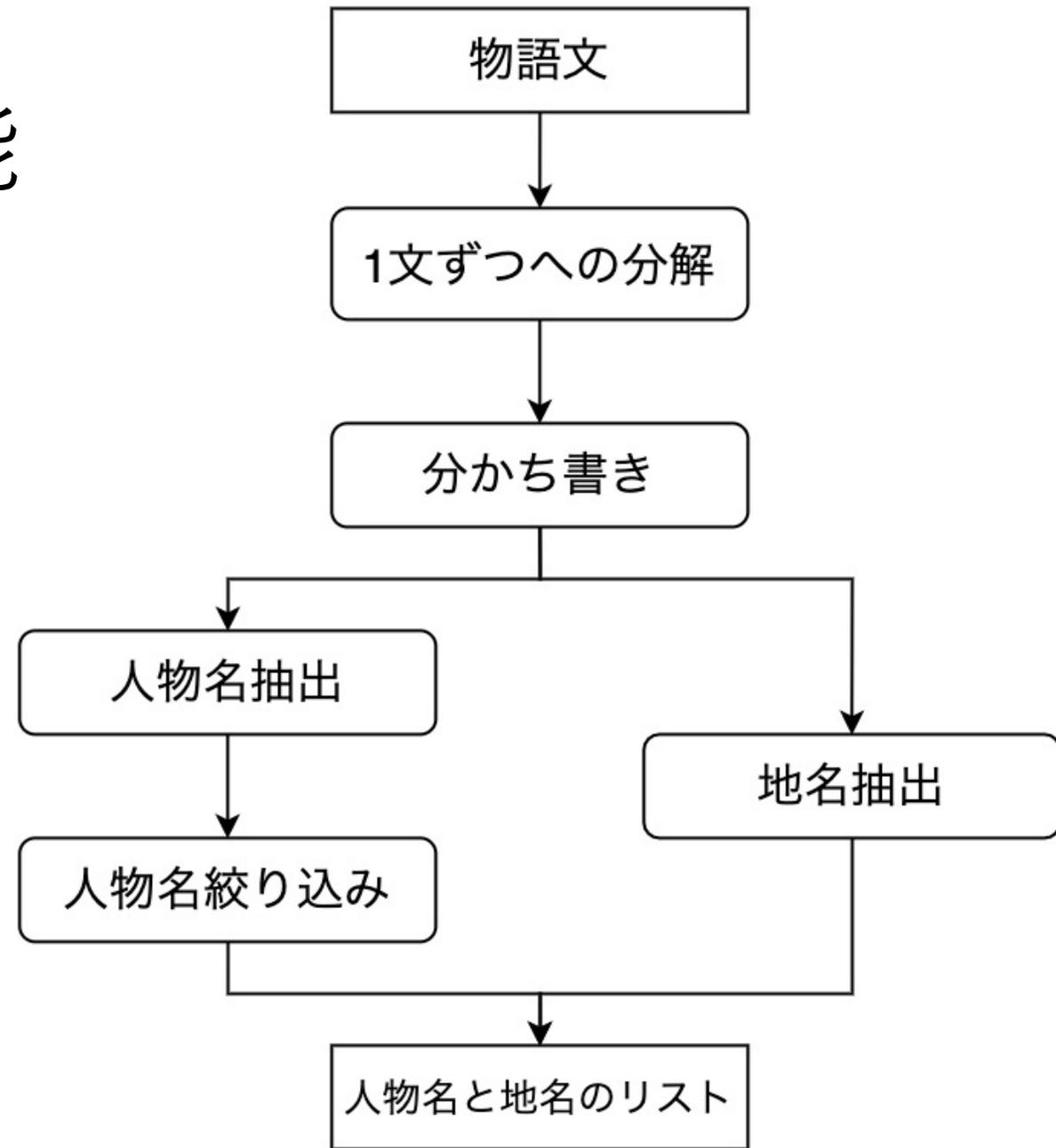
システム構成図



物語基本要素抽出機能

主に以下の5つの機能から成り立っている。

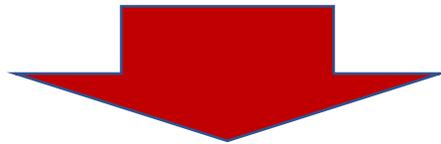
- ◆ 1文ずつへの分解
- ◆ 分かち書き
- ◆ 人物名抽出
- ◆ 人物名絞り込み
- ◆ 地名抽出



1文ずつへの分解

物語文を句点ごとに区切ることで、1文ずつに分解する。

吾輩は猫である。名前はまだ無い。どこで生れたかとうんと見当がつかぬ。何でも薄暗いじめじめした所でニャーニャー泣いていた事だけは記憶している。



吾輩は猫である。

名前はまだ無い。

どこで生れたかとうんと見当がつかぬ。

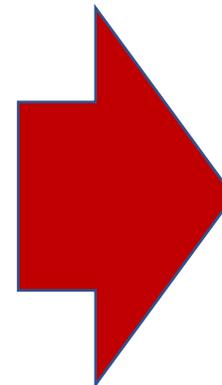
何でも薄暗いじめじめした所でニャーニャー泣いていた事だけは記憶している。

本研究では、1文を単位とする。

分かち書き

形態素解析によって単語を出現順に保持したリストを抽出する。

かず子がいてくれるから、私は伊豆へ行くのですよ。

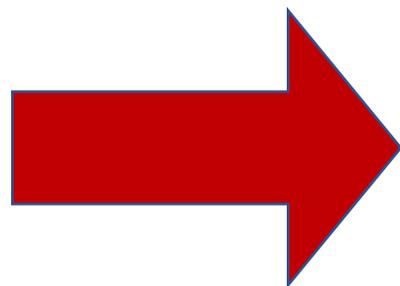


かず子
が
い
て
くれる
から
、
私
は
伊豆
へ
行く
の
です
よ
。

人物名抽出

形態素解析エンジンの品詞細分類3において「姓」または「名」を示す単語を抽出する。

かず子¹が²いてくれる³から、



かず子 カズコ かず子 名詞-固有名詞-人名-名

が ガ が 助詞-格助詞-一般

い イ いる 動詞-自立 一段 連用形

て テ て 助詞-接続助詞

くれる クレル くれる 動詞-非自立 一段・クレル 基本形

から カラ から 助詞-接続助詞

、 、 、 記号-読点

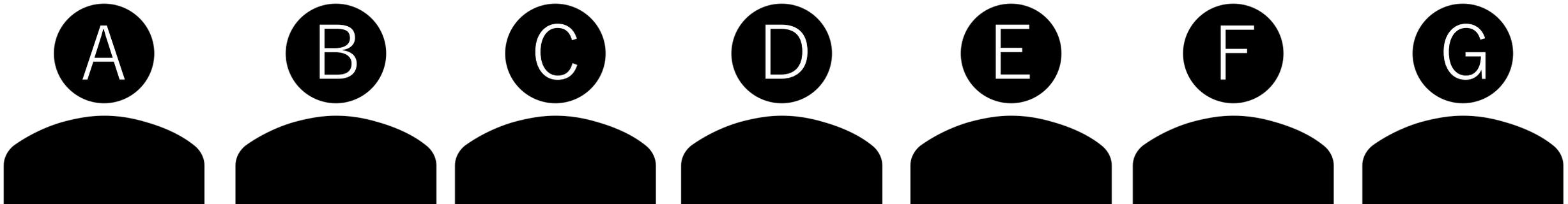
かず子¹ カズコ² かず子³ 名詞-固有名詞-人名-名

品詞細分類3 ↑

人物名絞り込み

出現頻度の高い7名の人物の名前を抽出する。

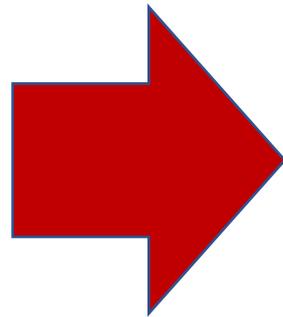
→主要人物に限定するため



地名抽出

形態素解析エンジンの品詞細分類1において「固有名詞」を示し、かつ品詞細分類2において「地域」を示す単語を抽出する。

私は伊豆へ行くのですよ



私 ワタシ 私 名詞-代名詞-一般

は ハ は 助詞-係助詞

伊豆 イズ 伊豆 名詞-固有名詞-地域-一般

へ へ へ 助詞-格助詞-一般

行く イク 行く 動詞-自立 五段・カ行促音便 基本形

の ノ の 名詞-非自立-一般

です デス です 助動詞 特殊・デス 基本形

よ ヨ よ 助詞-終助詞

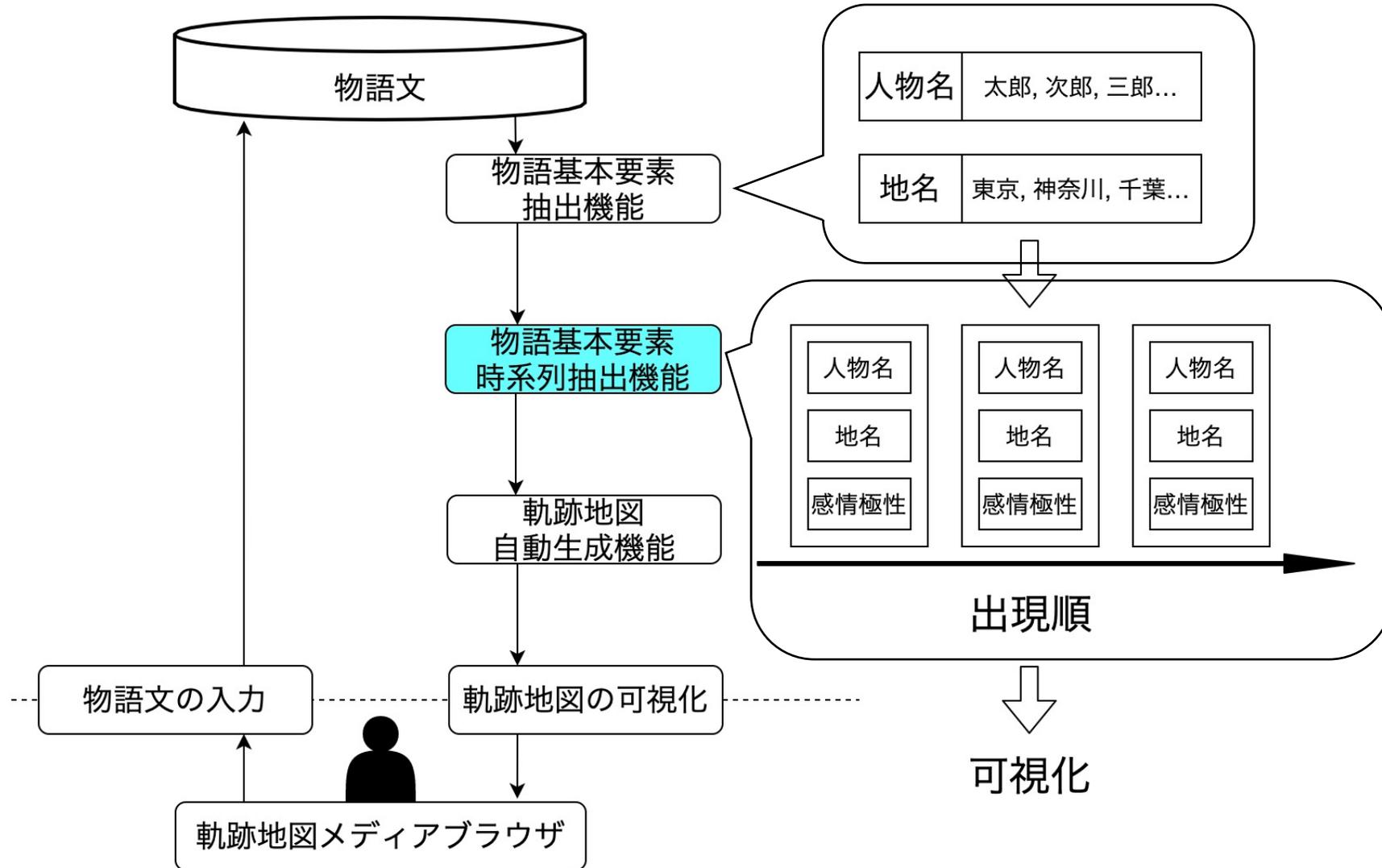
。 。 。 記号-句点

伊豆 イズ 伊豆 名詞-固有名詞-地域-一般

品詞細分類1 ↑

↑ 品詞細分類2

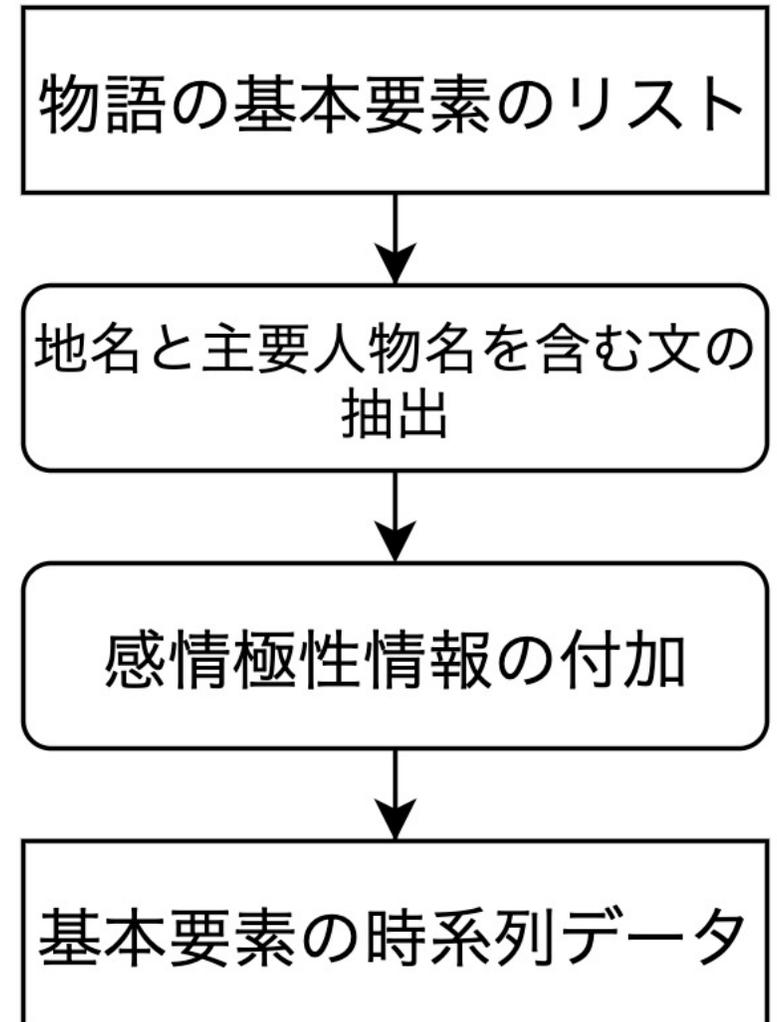
システム構成図



物語基本要素時系列抽出機能

主に以下の2つの機能から成り立っている。

- ◆地名と主要人物名を含む文の抽出
- ◆感情極性情報の付加



地名と主要人物名を含む文の抽出

- 今回は主要人物が地点を移動する様子を可視化する。
 - 「地名」と「主要人物名」の双方を含む場合のみ軌跡地図に表現する。
- 出現順に表現することで、時間情報として可視化する。

感情極性情報の付加

日本語評価極性辞書^{[1][2]}を用いて、極性値を抽出しその平均値でネガティブであるかポジティブであるかを判定する。

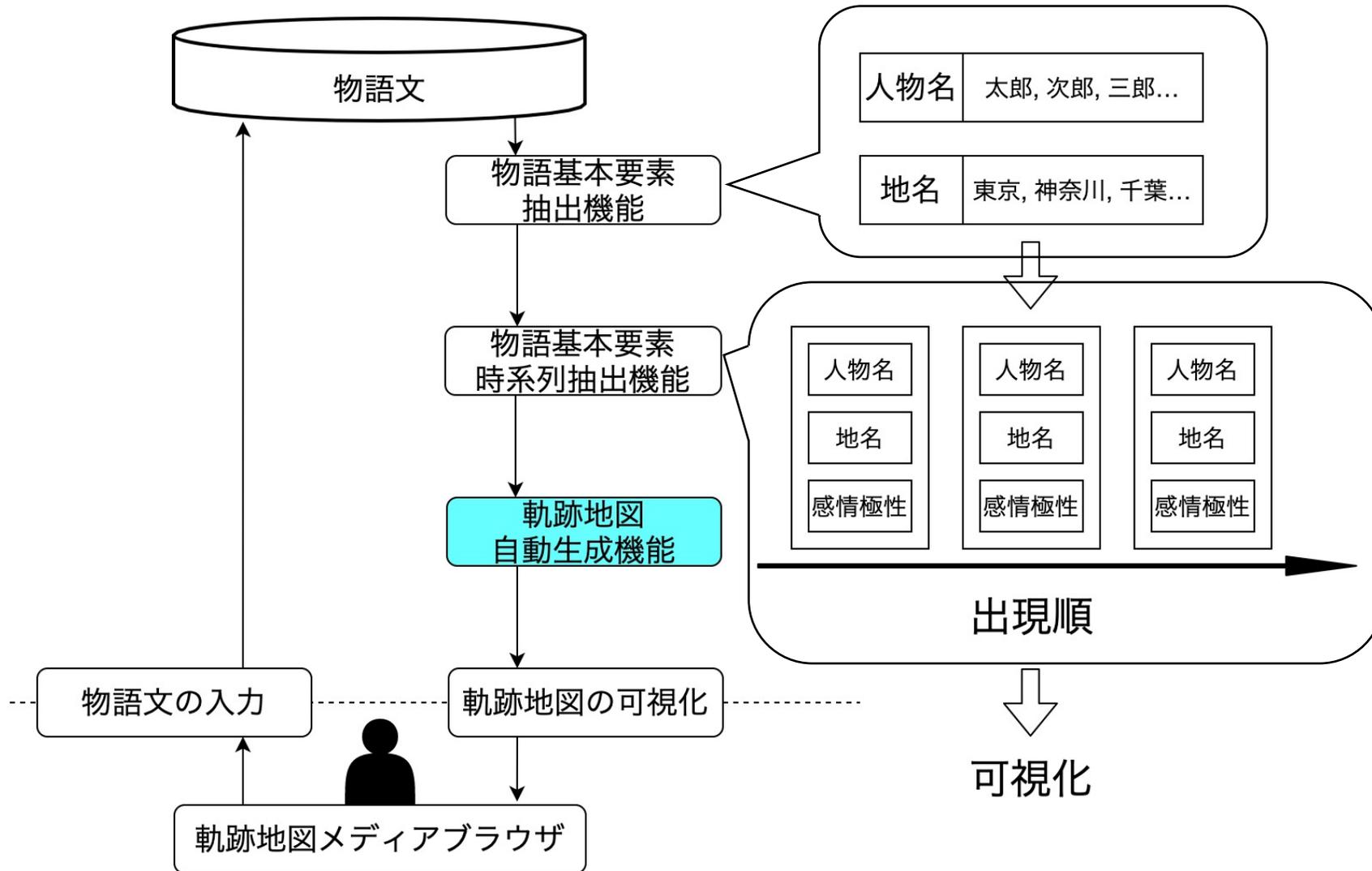
文	感情極性値
喜び というのは、今から十年以前に家出をした、長男の壮一君が、南洋ボルネオ島から、おとうさんにおわびをするために、日本へ帰ってくることでした	1.0
怪盗 の巣くつ [# 「怪盗の巣くつ」は中見出し] 賊 の手下の 美しい 婦人と、 乞食 と、赤井寅三と、 気 をうしなった明智小五郎とを乗せた自動車は、さびしい町さびしい町とえらびながら、走りに走って、やがて、代々木の明治神宮を通りすぎ、 暗い 雑木林の中にポツンと建っている、一軒の住宅の門前にとまりました	-0.71

ポジティブ：+1.0 ネガティブ：-1.0

[1]小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.3, pp.203-222, 2005.

[2]東山昌彦, 乾健太郎, 松本裕治, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第14回年次大会論文集, pp.584-587, 2008.

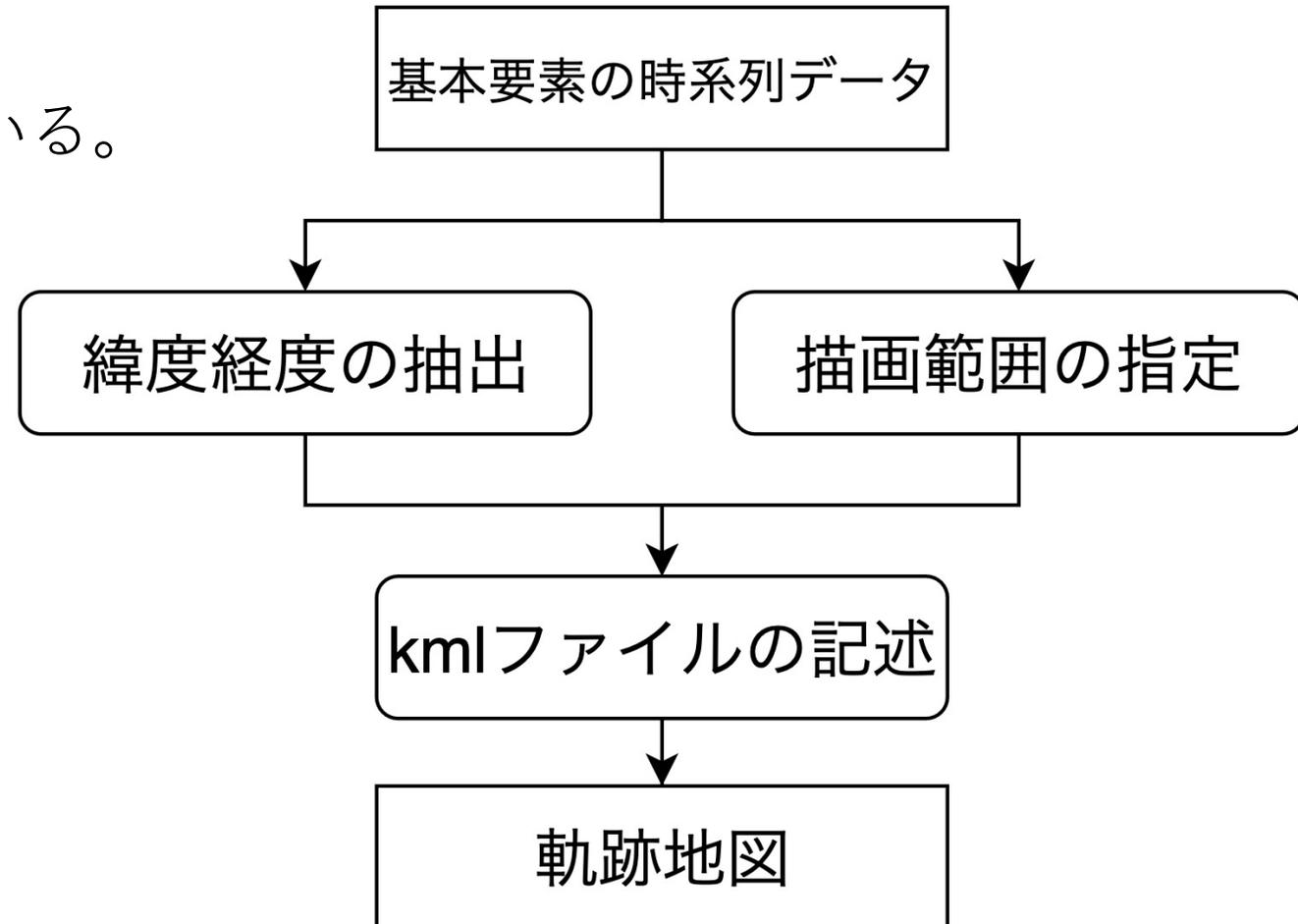
システム構成図



軌跡地図自動生成機能

主に以下の3つの機能から成り立っている。

- ◆ 緯度経度の抽出
- ◆ 描画範囲の指定
- ◆ kmlファイルの記述



緯度経度の抽出

- 地名を緯度経度に変換する機能が必要
- 今回はpythonライブラリである「Geocoder」を使用

描画範囲の指定[1/2]



内蒙古
自治区

吉林

辽宁

北朝鮮

日本海

山东

大韓民国

日本

黄海

江苏

東シナ海

描画範囲の指定[2/2]

- 描画範囲：地上からのカメラの距離
- 前実験によって描画範囲を以下のように設定する。

地名の種別	描画範囲
国	30000km
その他	600km

kmlファイル

- ◆ kml(Keyhole Markup Language)ファイル

- ・ ・ ・ 地理データと関連コンテンツを格納するためのファイル形式

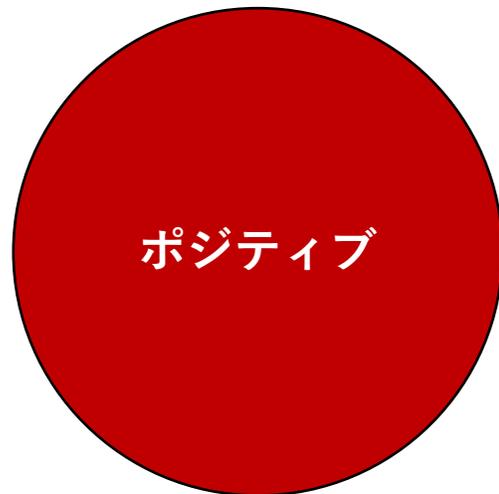
- ◆ Google earth で可視化が可能

基本要素の表現方法

- ▶ 人物は出現回数の多い順にA~Gに割り振り、アイコンとして表示



- ▶ 感情極性はそれぞれアイコンの色で表示



目次

- 研究背景・目的

- 提案方式

- 実験

- まとめ

実験

➤ 実験の目的

1. 地名と主要人物名が正しく抽出されるかの検証
2. 基本要素が軌跡地図に正しく可視化されるかの検証

➤ 実験データ

青空文庫より下記の作品を対象とした。

- ・江戸川乱歩作 「怪人二十面相」
- ・夏目漱石作 「坊ちゃん」

怪人二十面相から抽出された地名

抽出した単語における地名の割合：50%

可視化できる地名		可視化できない地名	地名以外の単語	
東京	池袋	二階	二日	三歳
日本	戸塚	岩の上	一本	二歩
東京駅	港区	国中	一日	四日
下関	王子	谷口	十人	十三日
修善寺		戸山	十時	五日
エジプト		土町	十日	四丁
代々木			五十人	三日
ロシア			三分	六分
中国			左門	日下部
鎌倉			あいだ	中
フランス			日	一網
新宿区戸塚町			れんげ	おせど

坊ちゃんから抽出された地名

抽出した単語における地名の割合：47%

可視化できる地名		可視化できない地名	地名以外		
東京	花崗	二階	五十	四丁	十間
露西亞	博多	運動場	三間	三股	八日
日本	会津	角屋	三日	十人	七日
越後	伊万里	三階	一本	五十人	十時
九州	大手町	浪の上	一日	六日	一重
高知	薬師町	古池	二日	十一	半町
延岡	神戸	温泉	海屋	大分	指金
鎌倉	土佐		菜飯	浜の	温
箱根	相良		米	平	厚
住田	鍛冶屋町		日	植	伊
宮崎	通町		驚	古賀	萩野
紀伊	小川町		折合	毘沙門	曲り

2つの作品から抽出された主要人物名

怪人二十面相	
人物名	出現回数
壮二	67
明智	64
小林	56
壮一	35
羽柴	34
中村	32
辻野	34

坊ちゃん	
人物名	出現回数
清	86
古賀	35
堀田	22
たまえ	11
馳	11
山城屋	10
勘太郎	10

怪人二十面相の軌跡地図



日本

喜びというのは、今から十年以前に家出をした、長男の壮一君が、南洋ボルネオ島から、おとうさんにおわびをするために、日本へ帰ってくることでした

目次 < 1 / 15 >

bocchan
KMLファイル

坊ちゃんの軌跡地図



日本

西洋館も日本建も全く不用であったから、そんなものは欲しくないと、いつでも清に答えた

☰ 目次 < 1 / 15 >

3D

— +

🗺️

目次

- 研究背景・目的
- 提案方式
- 実験
- まとめ

まとめ

- 文章中の位置情報をもとに物語の展開を可視化する軌跡地図の自動生成を実現した。
- 地名、人物名、感情極性の基本要素を抽出し、それぞれのシーンにおいて出現順に可視化することで読者の内容理解を補助することができる。
- 軌跡地図で可視化することで物語の場面転換をツアーのように見ることができるため、より物語のイメージ想起に繋がる。

今後の展望

- ◆基本要素抽出の精度向上
- ◆過去存在した地名への対応
- ◆観光分野への応用