



# 阿部研究室 卒業研究(A/B) 内容紹介

2014年11月  
情報システム学科 阿部秀尚

## 研究内容の方針

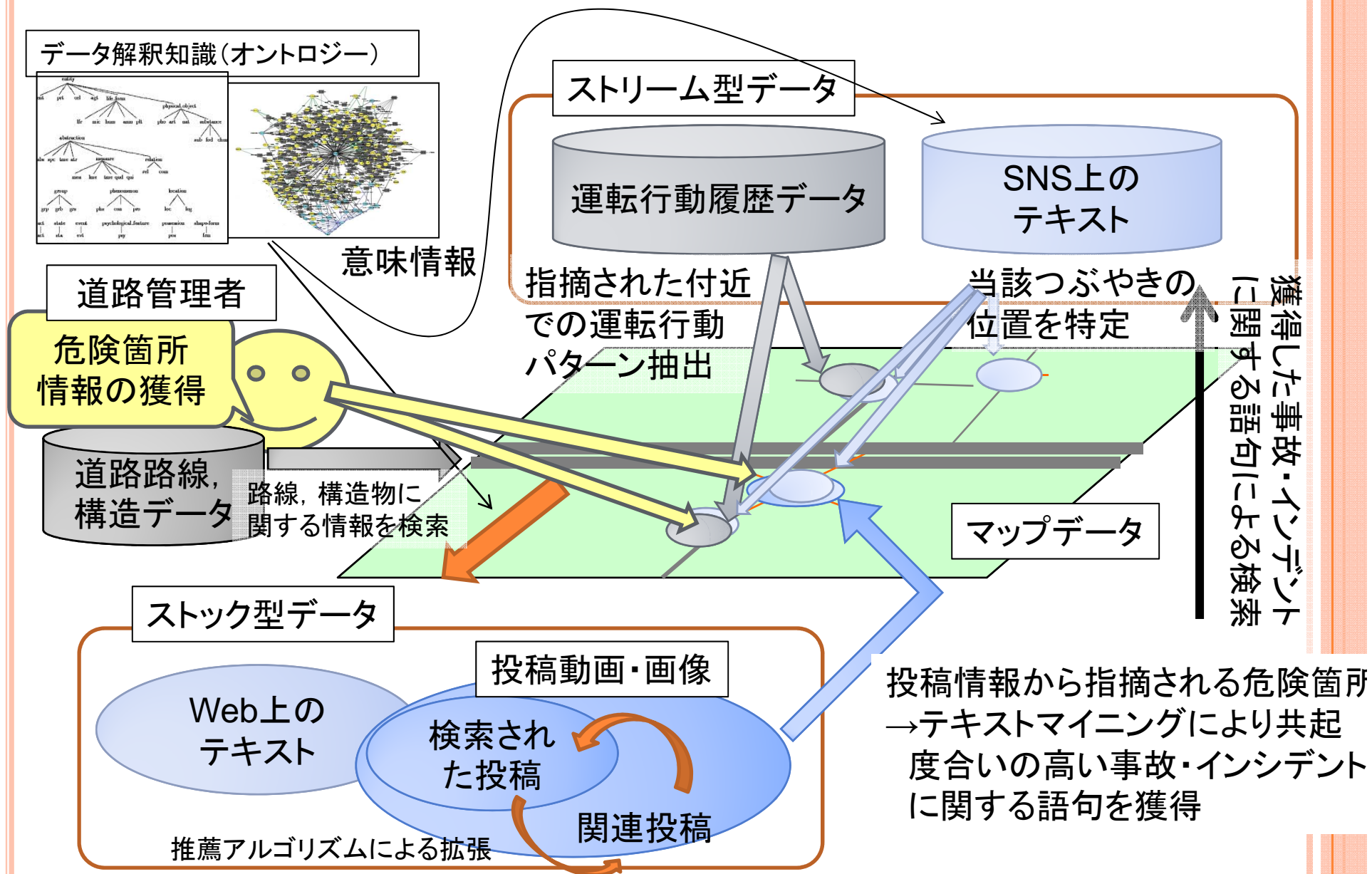
- 人手で行っている作業をデータ＋データ分析で支援する
  - データ：情報システムに蓄積されるログデータやソーシャルメディア上のテキスト
  - データ分析：データから規則性を見出す手法
    - テキスト対象だと「テキストマイニング」
    - データ全般に「データマイニング」
  - 支援：人間と全く同じではないが同じような作業を行う
    - 人間と同じ作業をある程度の正確さで疲れずに行える
    - 気が付かないような視点からの結果が得られる

→どんな作業を支援するのは、各自で設定

## 研究成果として期待する技術的テーマ

- テキストやデータの分析
- テキストやデータと連動した統計データ可視化
  - 「情報可視化」がキーワード
  - データとしては時系列テキストやLOD(Linked Open Data)を想定
- テキストマイニング手法の実装
  - 主に語句の評価指標を想定
- 時系列テキストマイニング
- 系列パターンマイニング } 難しいパターン生成などの処理は Wekaや外部プログラムを利用
- 作業意図の理解や作業意図を酌んでいないことに起因する操作パターン(使いにくい箇所)の抽出など
- 情報推薦システム

# 成果の統合計画例：道路管理情報基盤



# 交通センサスLODとソーシャルメディア検索の連携(試作)

## LOD Search (任意のSPARQL実行)

```
select ?id ?name ?value
where {
  ?id <http://zest.comp.ae.keio.ac.jp/road_census_ld/車線数> ?value.
  ?id <http://zest.comp.ae.keio.ac.jp/road_census_ld/路線名> ?name.
}
ORDER BY desc(?value)
```

XML  JSON

送信

## LOD Social Media Search

路線名: 一般国道1号

Youtu

閲覧回数

送信

一般国道1号  
一般国道174号  
大阪臨海線  
大阪臨海線  
難波境川線  
難波境川線  
広井町線  
難波境川線  
一般国道22号  
大阪臨海線  
一般国道2号  
一般国道2号  
一般国道1号  
一般国道2号  
大阪臨海線  
一般国道2号  
大阪中央環状線  
一般国道19号  
日本橋芝浦大森線  
大阪中央環状線  
一般国道19号

LOD Fetch Elapsed time: 1 sec

② 路線名をクエリとして  
ソーシャルメディアを検索  
(画面ではYoutube→Twitter)

## LOD Social Media Search

一般国道1号

### Youtube検索パラメータの設定

閲覧回数  取得投稿数: 10

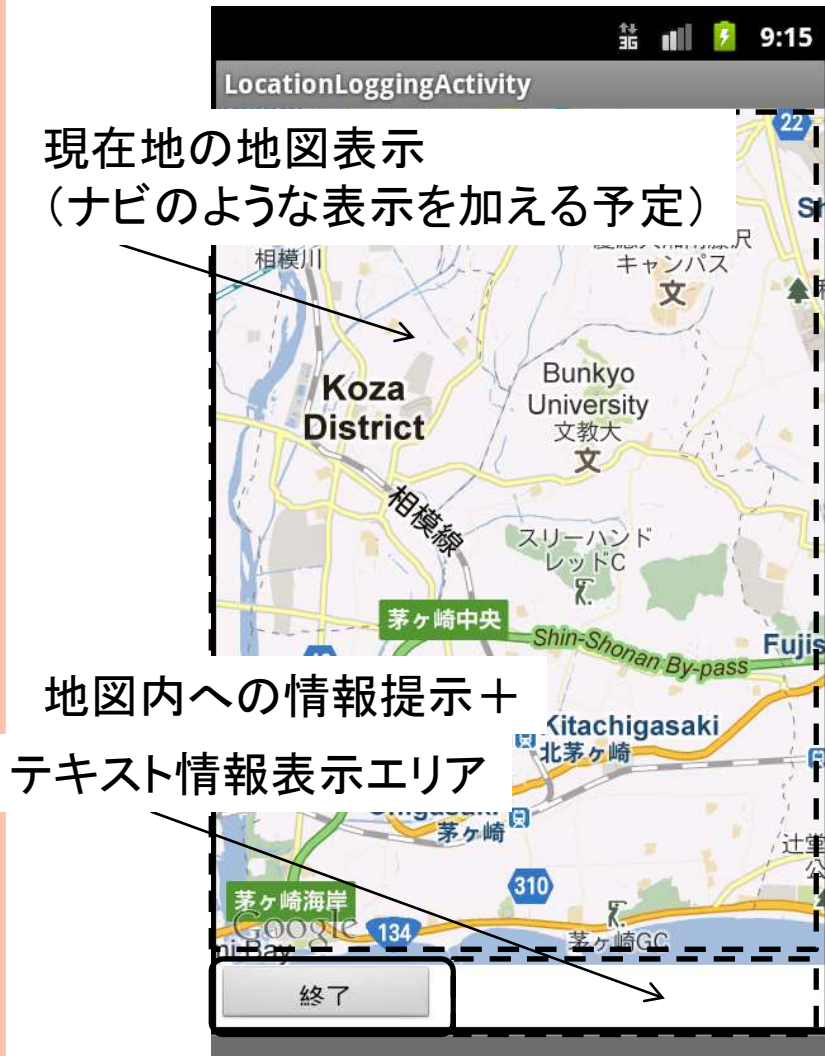
送信

	タイトル: 完全走破! 東京~大阪562.5キロ 国道1号線の旅 解説: 国道1号線の起点から終点までを撮影し、8倍速に編集したものです。 撮影日2006年08月11日~2006年08月12日。 投稿者: xMtN3y17p6E 再生回数: 19711 評価: 4.9333334 <a href="#">関連ツイート</a>
	タイトル: 国道1号 東京~大阪の旅1 解説: 東京日本橋からスタート 続き <a href="http://www.youtube.com/watch?v=6o26cCudxR0&amp;fmt=37">http://www.youtube.com/watch?v=6o26cCudxR0&amp;fmt=37</a> . 投稿者: DhaIS16frFQ 再生回数: 21159 評価: 5.0 <a href="#">関連ツイート</a>

① 交通センサスLODをSPARQL  
によって検索

# 成果の統合例：位置情報と連動した情報推薦システム

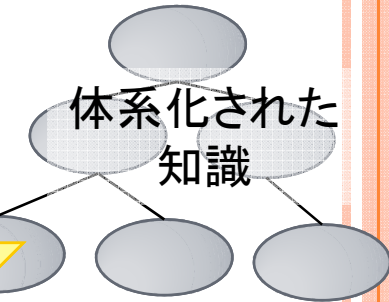
画面例：



移動履歴データ



データから得た低次パターン  
を意味づけされたデータへ



有用な情報の  
提供(推薦)

行動履歴パターンに  
基づく  
情報推薦システム

## 2013・2014年度テーマ設定

- テキスト分析 (TwitterAPIと語感極性辞書などを利用)
  - Twitter上での企業や製品の評判分析
  - 個人的な嗜好の強い対象 (アニメ, ゲーム) についての評判分析
  - テキストからの株価変動予測
- 可視化応用
  - 時系列テキスト中の盛り上がりの効果的な視覚化
- 行動履歴データの取得と利用
  - Android端末での行動履歴データ収集アプリの作成と個人の行動目的推定のための時系列データ分析

## こんな人に適しています

- テキストの分析に興味がある
  - 「お客様の声」って本当に分析されているの？と疑問に思う
  - Twitterなどでの口コミ情報を扱えるようになってみたい  
→ソーシャルリスニング(広聴), ソーシャルセンシングができる, と言えるようになりたい
- 世の中をもっと便利にしたい
  - 他人との関わりを情報システムを通すことで円滑にしたい
  - 人が便利に使える通販サイトやソーシャルメディアを考えてみたい
- とりあえず計算機に長時間, 計算させるようなことをしてみたい



# どんな技能を実践してみることができるのか ～リクエストによります～

## ○ プログラミング

- PHPによるプログラミング
- JavaやAndroidでのアプリ開発
- Webブラウザベースのアプリ開発(HTML5+JavaScript)

## ○ データ活用手法

- データ(テキスト, オープンデータなど)の収集システム
- データ分析・データマイニング手法
  - 「分析する」ということ自体が目的
- データからの情報可視化
  - ゲームだろうが, グラフィクスだろうが, 何でも良いので分かりやすく情報を提示するには?を学んで, コンピュータシステム上で実現する

# 卒業研究の進め方

- 春学期: 技術習得や論文のまとめを行ってテーマを決める
  - 4年生当初で考え付くことはすでに誰かがやっていることが多い(3年生までにそれだけ専門性がついてきた証拠)
  - 過去の研究の成果(論文)をまとめることで次の課題が見えてくる  
→ 欧米ではこれを「巨人の肩に乗る」という
  - “やってみたらこうなりました”から何を課題として解決すべきかを考えて、工夫を加えることを考える
- 秋学期: 各自(グループ)のテーマに沿って研究を進める
  - 研究全体の研究計画書を書く
  - 研究を進めるうえで必要な実験計画をたてて、実行する  
→ 1度で思うような結果は得られないのが当然
  - やったことはレポート(レポートの積み重ねが論文)としてまとめる
    - 目次は提示します