

観光レビューを対象とした 耳より情報提示手法の提案

☆ 阪井奎伍 甲南大学
灘本明代 甲南大学

背景

旅行へ行く前に

⋮
?時 △△神社
⋮
?時 昼食
⋮
?時 ××寺
⋮

昼食はどうしよう？



何時から開くのかな？

何を持っていこう？

どのルートで行こう？

子供は楽しめるかな？

背景

旅行先

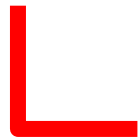


知らない土地
慣れていない土地

分からないことが
多数ある

事前に調べることによって快適な旅行が出来る

ネット検索 や 観光雑誌



➤ 公式サイト

➤ レビューサイト

など

背景（伏見稻荷大社）

公式サイト

公式サイト

基本的な情報

- 施設紹介
- 営業時間
- アクセスマップ
- イベント紹介

など

The screenshot shows the official website of Fushimi Inari Taisha. At the top, there is a navigation bar with the site name '伏見稲荷大社' and various utility links like '大社マップ', '周辺マップ', 'サイトマップ', and 'よくあるご質問'. Below the navigation bar, there is a large yellow banner with the text '基本的な情報'. Underneath the banner, there is a section titled 'これからの祭礼と行事' (Upcoming Festivals and Events) with a calendar view showing dates from 1月 to 12月. Below the calendar, there is a list of recent news items under the heading 'お知らせ' (Notice), including dates and titles like '2015年 6月20日 トピックス 2年連続!! トリップアドバイザー“行ってよかった外国人に人気の日本の観光スポット2015”第1位受賞'. On the right side of the page, there are several smaller images and text boxes, including 'ご祈禱' (Prayer), 'お守り' (Omamori), and '大社マップ' (Main Shrine Map).

背景（伏見稻荷大社）

レビューサイト

レビューサイト



膨大な情報

全てを把握することは困難



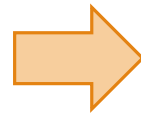
お得な情報が埋もれてしまう

- 基本
- 個別
- お行

目的

お得な情報

知ってよかった
参考になった



耳より情報

インターネット上の膨大な
レビューから耳より情報を抽出

→ 旅行の計画の手助けを！

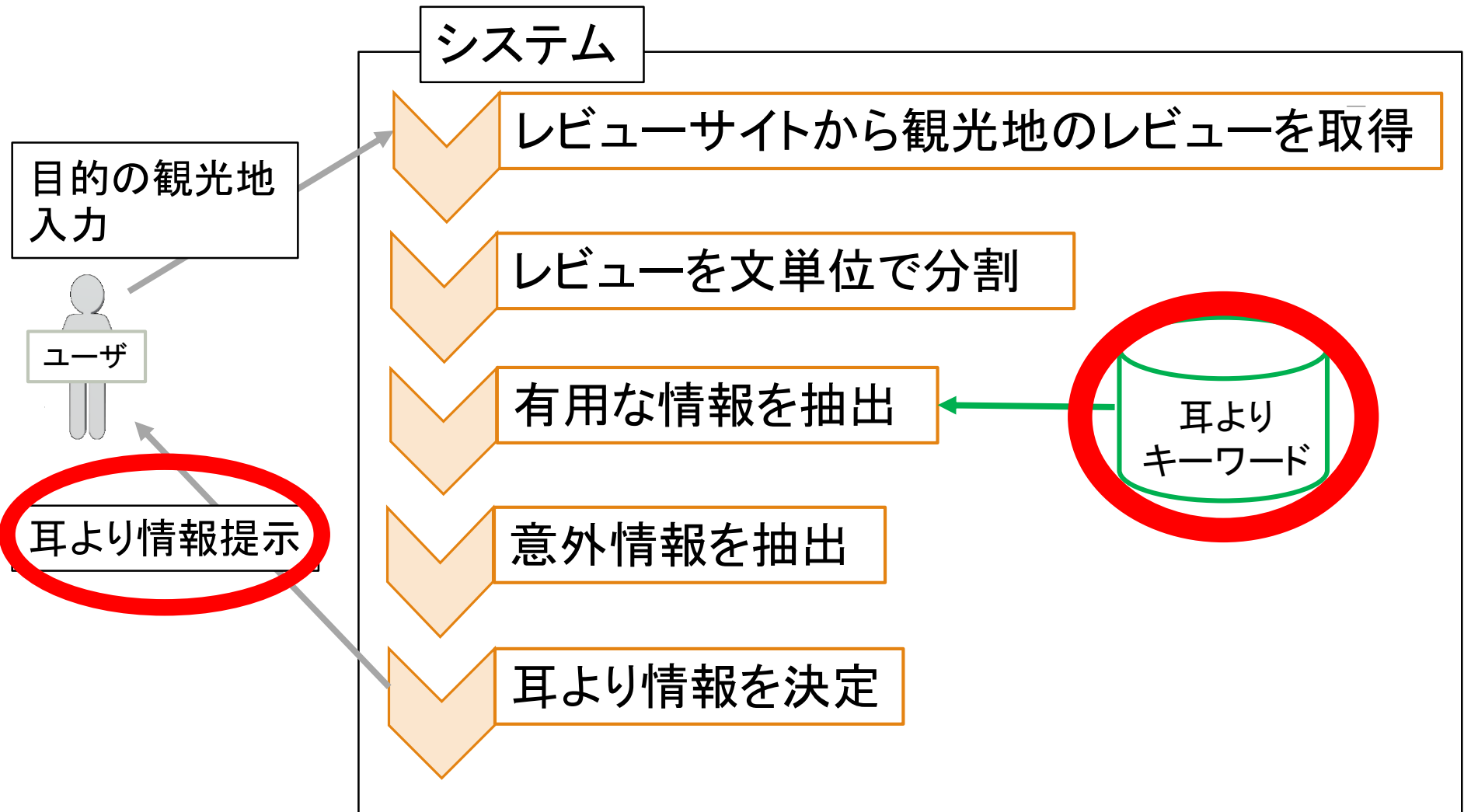
耳より情報の定義

耳よりキーワードを含む文

耳より情報 = 有用な情報 + 意外情報

一般的に広く知られていないレアな情報

全体の流れ



耳よりキーワード

実験

汎用耳よりキーワード

- 様々な観光地の有用なレビューに用いられるキーワード
- ➔ おすすめ, できる, 無料, 禁止

機械学習

特有耳よりキーワード

- 観光地特有のキーワード
- ➔ 伏見稻荷大社の場合
スズメ焼き, 足腰

汎用耳よりキーワード

| タイプ | 汎用耳よりキーワード例 |
|-------|---|
| アドバイス | 推奨, 禁止, 不安定, 利便性, 要注意, 適して, ベスト, 良心的, 避けたほうが, もってのほか |
| 時間 | 往復, 長蛇の列, 長時間, 待たずに, 空いていて, 先着順, 整理券, ツアー時間, 無制限, ランチ |
| 金銭 | %off, 無料, 割引, 参加費, 請求, 方が安い |
| 便利 | 飲み水, 必需品, レンタル, 着替え, 用意 |
| 地理 | 裏通り, 裏道, トイレ, 地図, 交通機関, アクセス |
| その他 | ベビーカー, 車椅子, 足の悪い, 接客態度 |

全167語

特有耳よりキーワード

- 観光地特有のキーワード
SVMにより特有耳よりキーワードを判定

- ✓ 素性

数詞, 場所, 人名を除く

- ➔ レビュー中出现する名詞, 動詞, 形容詞と対象の観光地のレビュー文に対するそれらの語のTF-IDF値

- ✓ 学習 ➔ libSVM

- ✓ カーネル ➔ RBFカーネル

特有耳よりキーワード

➤ 観光地特有のキーワード

SVMにより特有耳よりキーワードを判定

教師データ

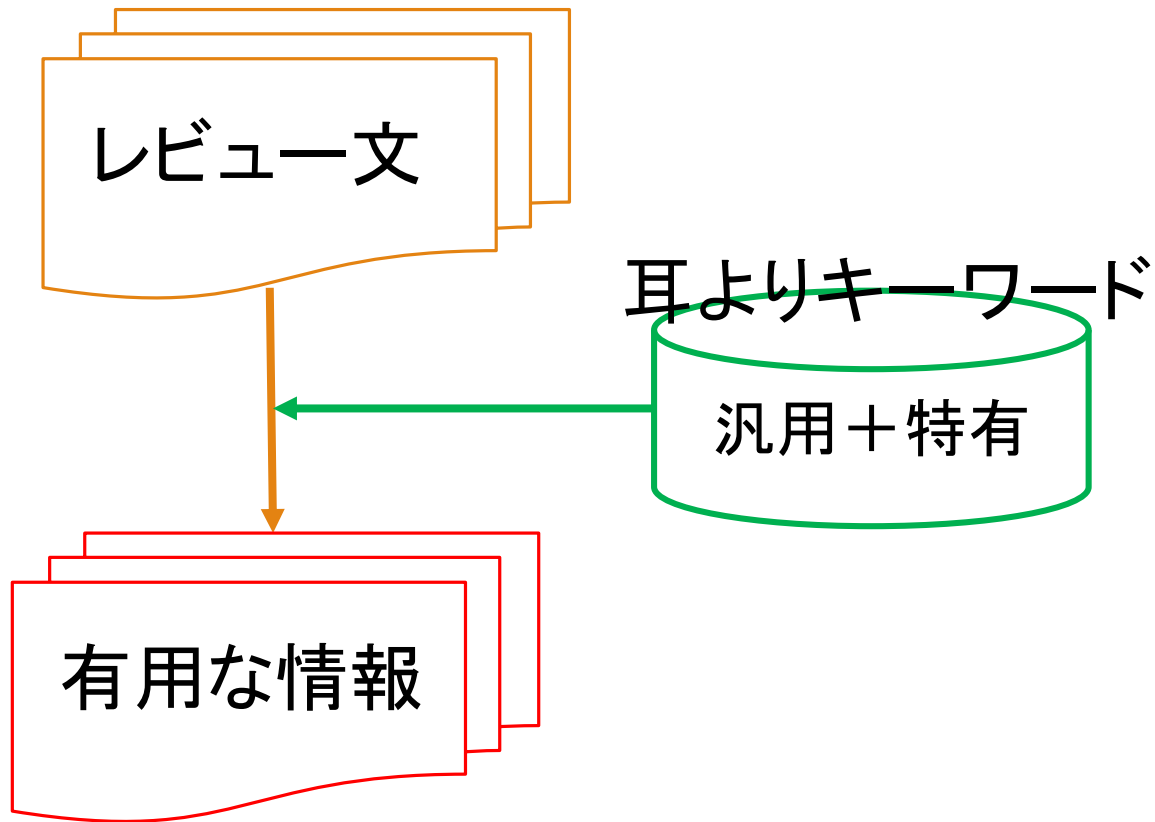
✓ 正例

汎用耳よりキーワードとある観光地のレビュー文に対するキーワードのTF-IDF値

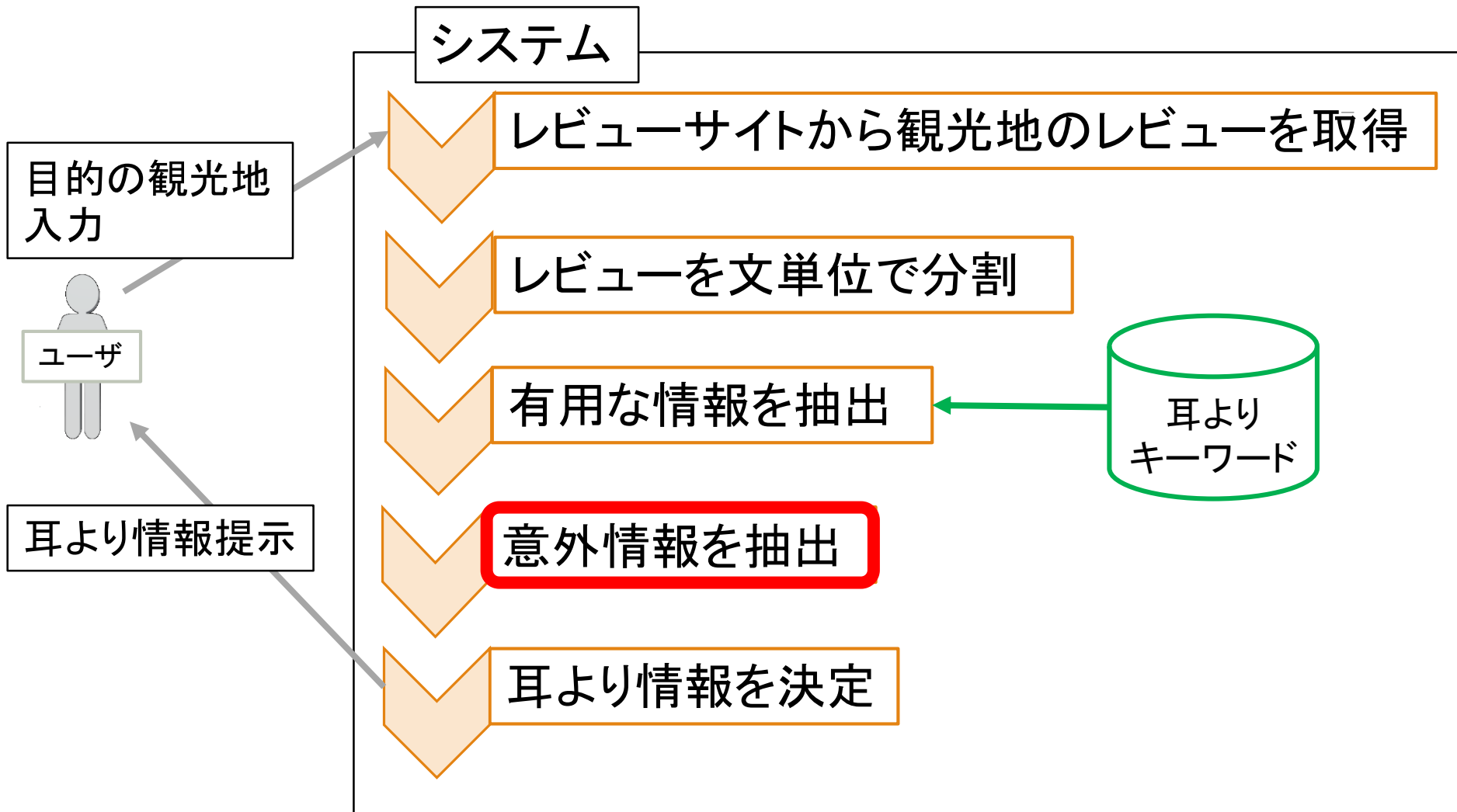
✓ 負例

汎用耳よりキーワード以外の名詞, 動詞, 形容詞とある観光地のレビュー文に対するそれらの語のTF-IDF値

有用な情報抽出



全体の流れ



意外情報

レビューにおける意外情報

あまり知られていない情報
➔ レビューにあまり書かれていない

有用な情報

話題1 話題2
話題3 話題n

話題毎にクラスタリング

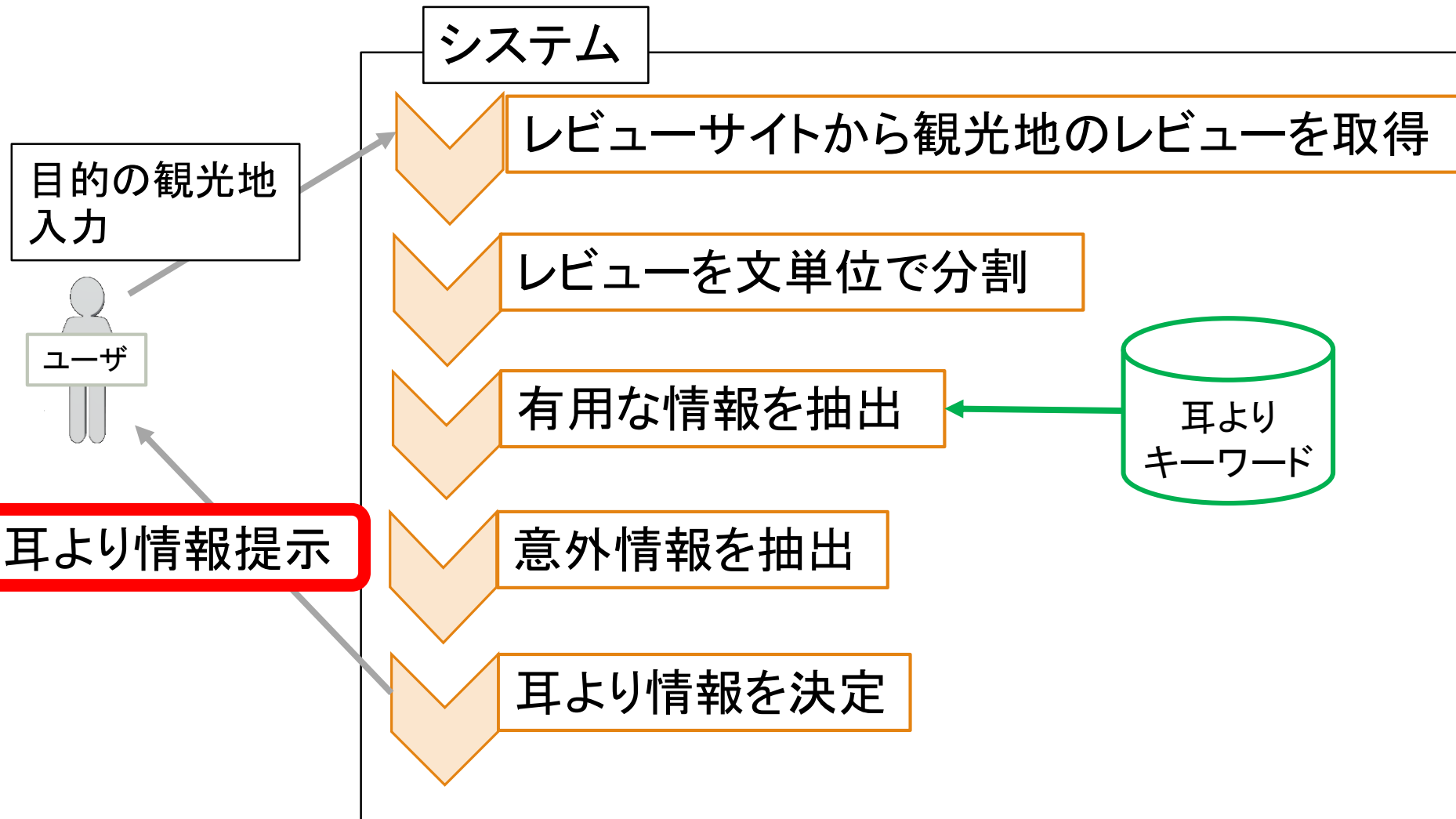
クラスタ毎に類似度を用いて
意外情報を抽出

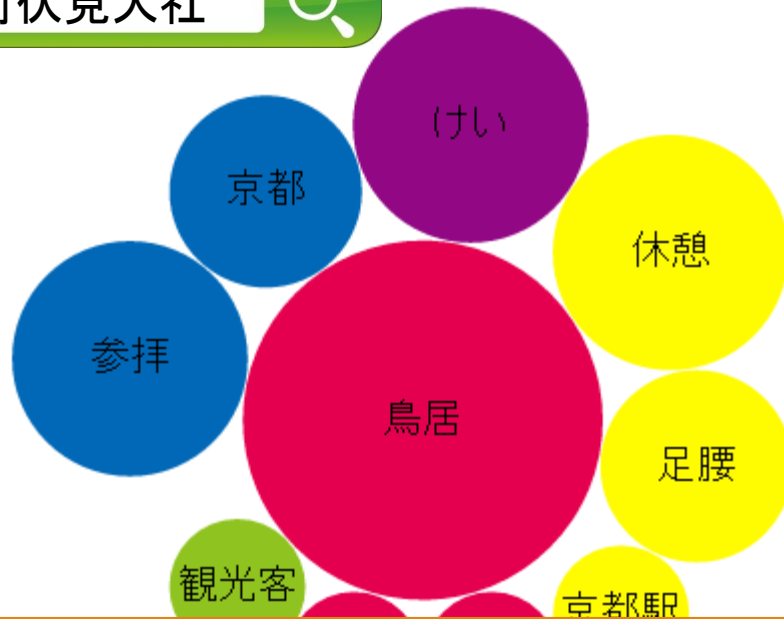
クラスタリング手法

短文にある程度適していると
考えられるRepeated Bisection法
を用いる

データ集合を繰り返し2分割し、
K-means法を実行する手法

全体の流れ



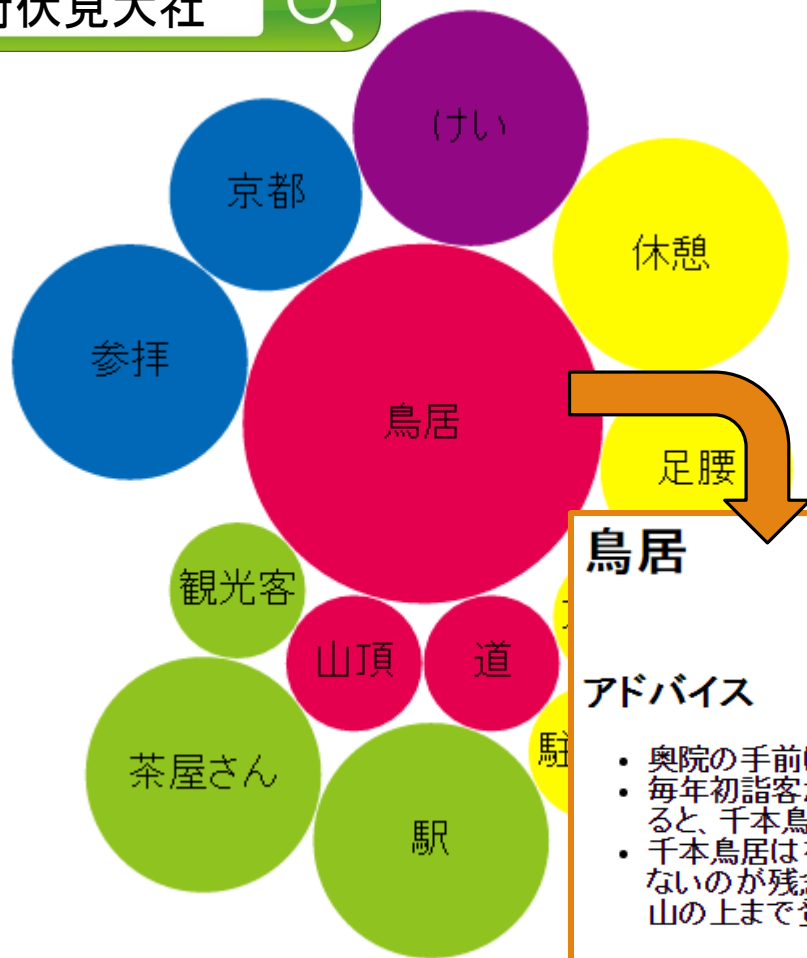


提示手法

- 話題
 - ➔ クラスタの中心ベクトルを構成する特徴語の特徴量の最も高い語
- 円のサイズ
 - ➔ 1つの話題に含まれる耳より情報の数により決定
- 色
 - ➔ 汎用耳よりキーワードのタイプとその数により決定
 - アドバイス → 赤
 - 時間 → 緑
 - 金銭 → 橙
 - 便利 → 紫
 - 地理 → 黄
 - その他 → 青



提示手法



鳥居

アドバイス

- 奥院の手前に千本鳥居があり是非みたいポイントです
- 毎年初詣客が大勢訪れる有名神社ですが、神社の建物からさらに奥、稲荷山を上ると、千本鳥居の幻想的な風景を見ることが出来ます
- 千本鳥居は有名で、人が多いときだとなかなか自分のペースで進み登ることができないのが残念ですが、人が少ないときは、少々体力に自信がなくても休憩しながら山の上まで登ってみるのがいいと思います

時間

- 奥社と御膳谷は稲荷山の鳥居をのぼったとこにあり、御膳谷までは30分くらいかかり、休憩なしで歩き続けるのはきついです
- 中はとても広く、一周ぐるっと回するにはかなりの時間と体力が必要ですが、鳥居をくぐって、汗を流して、ある程度上まで上ってから見える京都の風景は絶景です

地理

- みんな千本鳥居の写真を撮るので、あっちこっちで渋滞して純粹にお参りしたい人にはちょっとどうなのかなーという感じ

実験1

- 目的：訪問経験の有無による耳より情報の違い
- データセット：
 - 伏見稻荷大社 : 2337文
 - 嵐山 : 769文
 - 京都水族館 : 962文
 - 水木しげるロード : 905文
 - 横浜中華街 : 2022文
- 被験者：各観光地に対して、興味有り訪問経験有り3名
興味有り訪問経験無し3名
計6名

実験1

実験方法

- 耳より情報かそうでないか2択で判断
- 季節は考慮しないで耳より情報かどうか判断
- 訪問経験あり3名のうち2名以上が耳より情報と判断した文を訪問経験ありの正解データとする
- 訪問経験なし3名のうち2名以上が耳より情報と判断した文を訪問経験なしの正解データとする
- 訪問経験の有無別に正解データを比較する

実験1: 結果と考察

訪問経験の有無別の耳より情報の割合

| 観光地 | 訪問経験有り | 訪問経験無し |
|----------|--------|--------|
| 伏見稲荷大社 | 9.1% | 1.9% |
| 嵐山 | 6.8% | 2.9% |
| 京都水族館 | 1.2% | 6% |
| 水木しげるロード | 5.9% | 2.7% |
| 横浜中華街 | 8.4% | 3.5% |

京都水族館以外の観光地では訪問経験がある人の方が耳より情報と判断するレビューが多い

- 訪問経験がある場合、レビューの内容を理解しやすく次回の訪問時に活かすイメージがしやすいと考えられる

実験2

- 目的: 特有耳よりキーワードの有用性を測る
- データ: 実験1で用いたデータセットと訪問経験ありの正解データ
- 実験方法:
 - ✓ 伏見稻荷大社, 嵐山を教師データ
京都水族館, 横浜中華街, 水木しげるロードを評価データとする
 - ✓ 機械学習により特有耳よりキーワードと判定されたキーワードを用いて有用な情報の抽出を行い, 正解データと比較する

実験2：結果と考察

| 観光地 | 特有耳よりキーワードの有無による再現率 | |
|----------|---------------------|-----|
| | 無し | 有り |
| 京都水族館 | 50% | 75% |
| 水木しげるロード | 45.3% | 74% |
| 横浜中華街 | 32.9% | 82% |

特有耳よりキーワードを追加したことにより、全ての観光地において再現率が上がった

→ 機械学習による特有耳よりキーワード抽出はある程度できたと考える

実験2: 結果と考察

特有耳よりキーワード例

| 京都水族館 | 水木しげるロード | 横浜中華街 |
|--|--|---|
| 目玉, 作れる, うじゃうじゃ, 延長, コスト, 売り子, 重宝, 推薦, 満員, ベンチ, オオサンショウウオ | 妖怪, 出会える, 活気付いて, スタンプ, 提供, はがき, 持参, マップ, フェリー, 接待, 放映 | 詐欺, 食べ歩き, ランチタイム, 試食, 押し売り, ガイドブック, 東京メトロ, 甘栗, バイキング |

- ◆ 特有耳よりキーワードは抽出できた
- ◆ 延長や満員といった観光地特有ではないキーワードも存在した

まとめ

レビューサイトから耳より情報の抽出と提示手法を提案した

➡ 耳より情報を提示することで初めて行く
旅行先の計画を立てる手助けができる

今後の課題

- 意外情報抽出の改善
- 特有耳よりキーワード抽出の改善
- 耳より情報のパーソナライゼーション