

2017年12月4日

国立情報学研究所

パネルセッション2

モデレータ：角谷 和俊

関西学院大学

総合政策学部 メディア情報学科・教授
(社会情報学研究センター・センター長)

IDRユーザフォーラム 2017



パネルセッション・テーマ

「今、アカデミアが必要として
いるデータセットとは」

16:10~17:20

パネリスト

酒井 哲也 氏（早稲田大学）

河合 由起子氏（京都産業大学）

櫻井 一貴氏（株式会社リクルートテクノロジーズ）

原島 純 氏（クックパッド株式会社）

三澤賢祐 氏（株式会社Insight Tech）

アジェンダ

1. データ・リポジトリを利用した研究
[アカデミア] 酒井氏・河合氏
2. データ活用で役立った事例
[企業] 櫻井氏・原島氏・三澤氏
3. ディスカッション

お題:「今、アカデミアが必要と しているデータセットとは」

早稲田大学 情報理工学科

酒井 哲也

tetsuyasakai@acm.org

December 4, 2017@IDRユーザーフォーラム2017

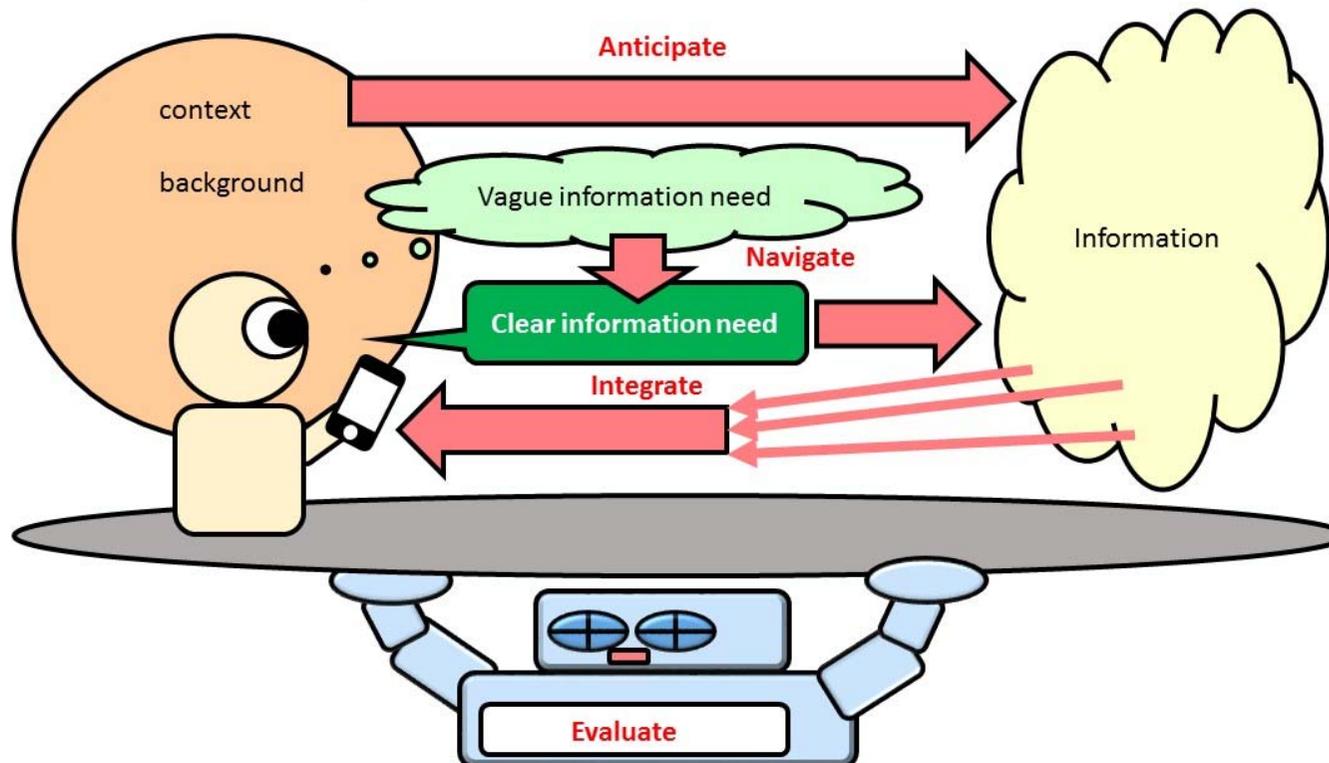
早稲田大学酒井研の守備範囲

Information
retrieval/access

natural language
processing

human computer
interaction

Mission: Easy Information Access



企業との共同研究@酒井研

- ヤフー研究所(日本) モバイル検索ログ・知恵袋・画像データなどを活用した研究 [Kadotami+17][Kato+17]
- Wider Planet (韓国) 広告データ・ショッピング履歴などを活用した研究
- Naver (韓国) モバイル検索ログなどを活用した研究 [Kawasaki+17]
- Microsoft (中国) インターン派遣による対話の研究
- Huawei (中国) ヘルプデスク対話の研究 [Zeng+17]

[Kadotami+17] Mobile Vertical Ranking based on Preference Graphs, ACM ICTIR 2017.

[Kato+17] LSTM vs. BM25 for Open-domain QA: A Hands-on Comparison of Effectiveness and Efficiency, ACM SIGIR 2017.

[Kawasaki+17] Ranking Rich Mobile Verticals based on Clicks and Abandonment, ACM CIKM 2017.

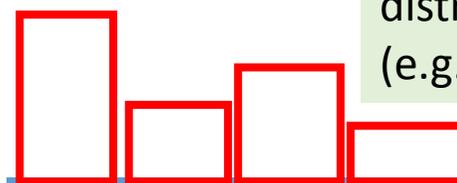
[Zeng+17] Test Collections and Measures for Evaluating Customer-Helpdesk Dialogues, EVIA 2017.

NTCIR-14 STC-3 (Chinese and English) Dialogue Quality subtask

INPUT:
a customer-helpdesk
dialogue $d \in D$

Customer post
Customer post
Helpdesk post
Customer post
Helpdesk post
Helpdesk post
Customer post
Helpdesk post

Gold distribution p^* based on
N annotators reflecting
subjective views



OUTPUT: an estimated probability
distribution p of dialogue quality score
(e.g. customer satisfaction)

NTCIR最大のタスク
Short Text Conversationでは
対話品質サブタスクを
立ち上げるが、データは
Weiboの企業公式ヘルプ
アカウントから抽出したもの

$M(d)$: how p
differs from p^*

$$meanM = \frac{1}{|D|} \sum_{d \in D} M(d)$$



とどのつまり

大量の対話データが欲しい
(タスク指向・非タスク指向)

なんでも「AIで何とかならないの」の世の中だし、みんな欲しいはず

第13回 NTCIR カンファレンス

NTCIR-13 カンファレンス

情報アクセス技術の評価

EVIA 2017

a Satellite Workshop of the NTCIR-13 Conference

2017年 12/5(火)-8(金)

学術総合センター (東京・千代田区)

国立情報学研究所 (NII)

▶参加者募集 Flyer: [\(英語版\)](#) [\(日本語版\)](#)

▶[NTCIR-13オンライン参加登録](#)

12月6日(水)は1日特別公開日です。基調講演、NTCIR-13タスク概要、NTCIR20周年招待講演に無料で参加できます。
す。:[12月6日\(水\)Free Day登録](#)

関連イベント: 12月4日(月) [IDRユーザフォーラム2017](#) (NII)

情報推薦と可視化

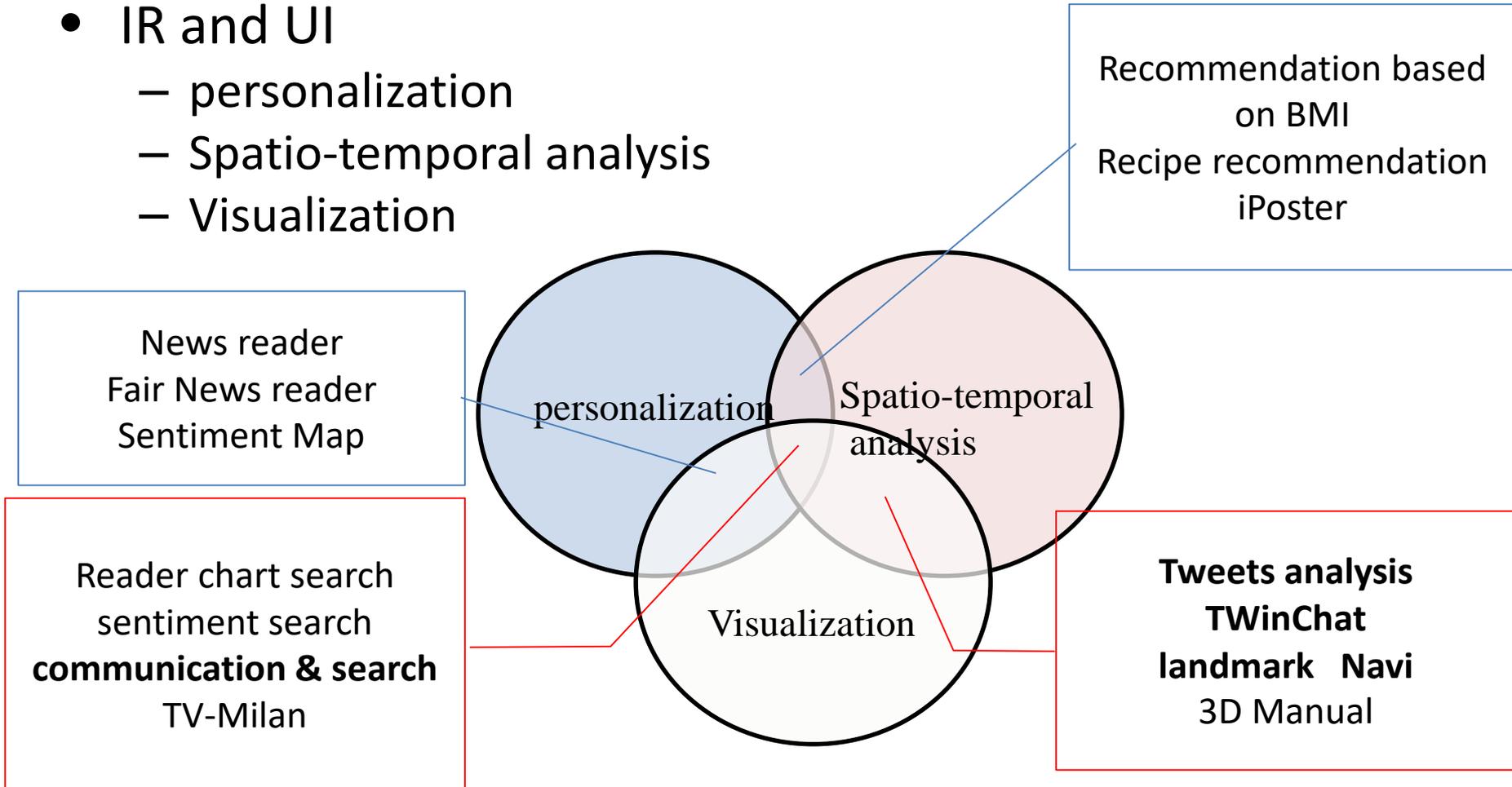
～データ利活用研究事例紹介～

河合由起子
京都産業大学



Research Topics

- IR and UI
 - personalization
 - Spatio-temporal analysis
 - Visualization

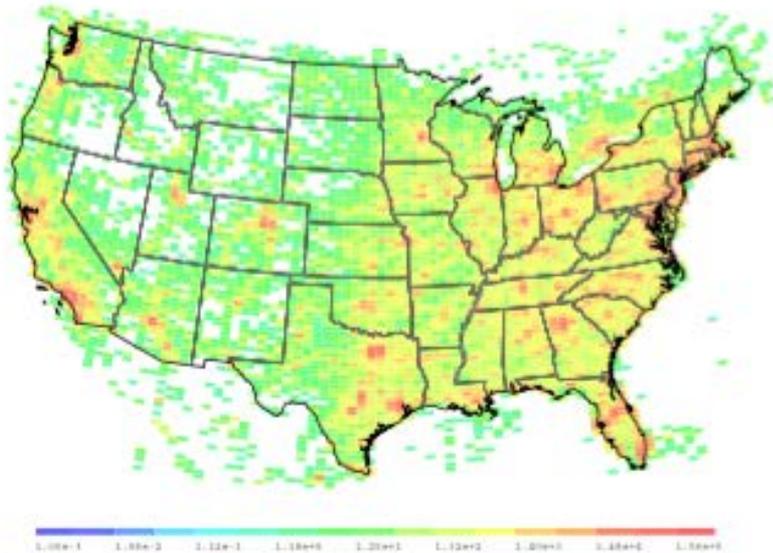
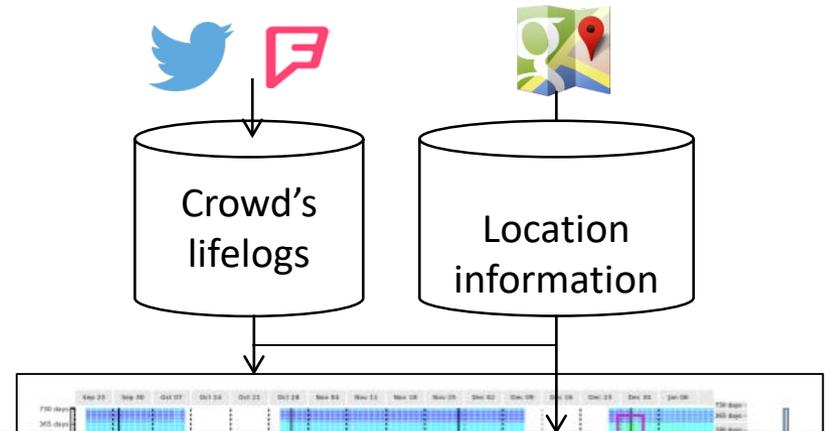


データコレクション

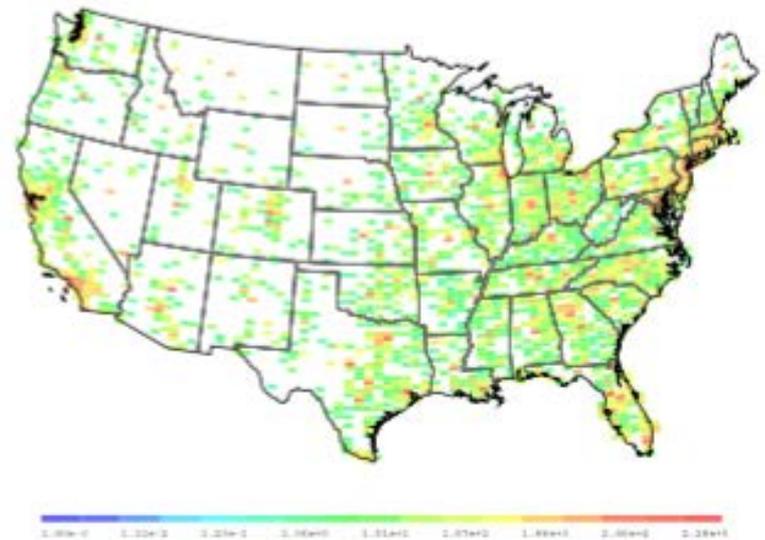
- 映像データ: 映像、音声、(closed caption)
- 地理データ: 緯度経度、建物や道名、(高さ、面積、道路幅)
- ユーザデータ
 - センサデータ: 端末ID、時間、緯度経度、(年齢、性別、MACアドレス)
 - アクセスデータ: ユーザID、時間、場所(URL)、(年齢、性別、MACアドレス)
 - ログデータ: ユーザID、クリックや入カログ、(年齢、性別、MACアドレス)
 - 生体データ: ユーザID、時間、場所、年齢、性別

SNSデータ分析事例

- ユーザの興味関心抽出
 - 過去と未来に対するユーザの興味[WWW2015]
 - 場所に対するユーザの興味抽出[KDD2015,WWW2016]



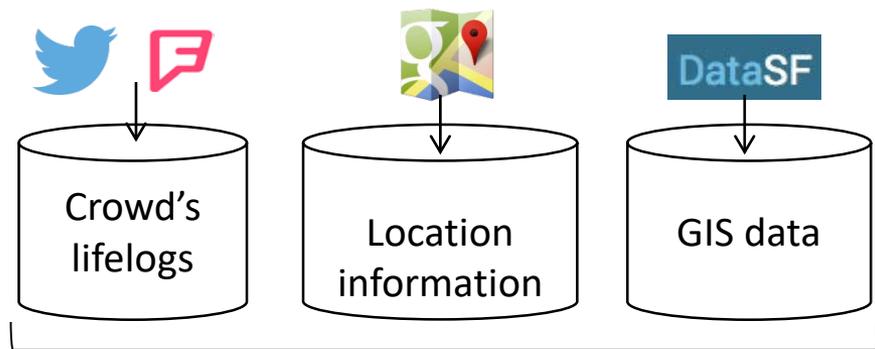
(a) Location stamp [47]



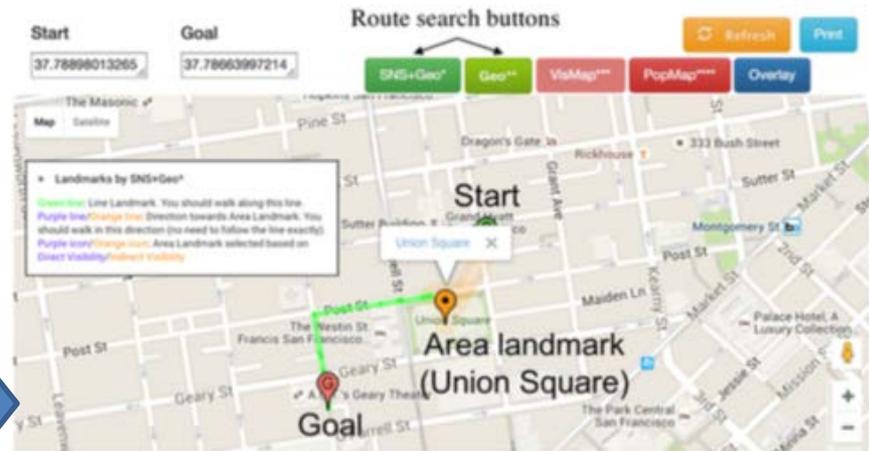
(b) Location mention [46]

地理データ分析事例

- 視認性と認知度に基づくランドマークナビゲーション
[UbiComp2016, IPSJ論文誌2016, インタラクシオン2015]



Point Landmarks Line Landmarks Area Landmarks



(a) Map

The number of the used landmarks 3

#	Category	Name
1	Area-Indirect	Union Square
2	Line	POST ST
3	Line	MASON ST

(b) Textual instructions



(c) Images of area landmark

映像データ分析事例

- 異種メディア検索

- 映像データの時空間分析および意味的分析に基づく映像、写真、地図、SNSデータ検索と融合 [IDR'17, DEIM2017, IJMDEM(2016), LBS(2016), WebDBf2016]

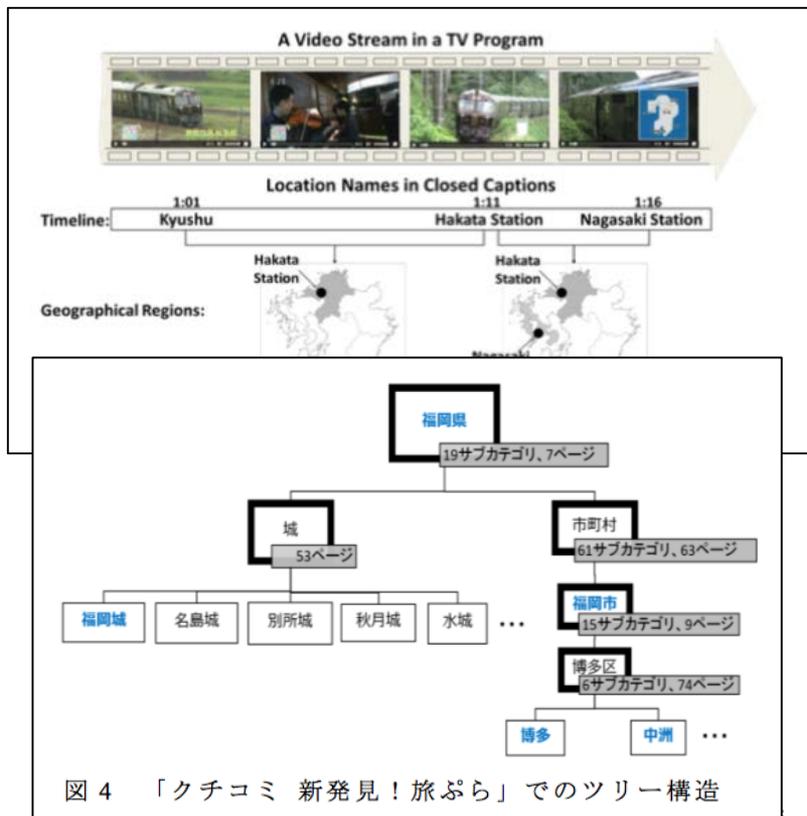


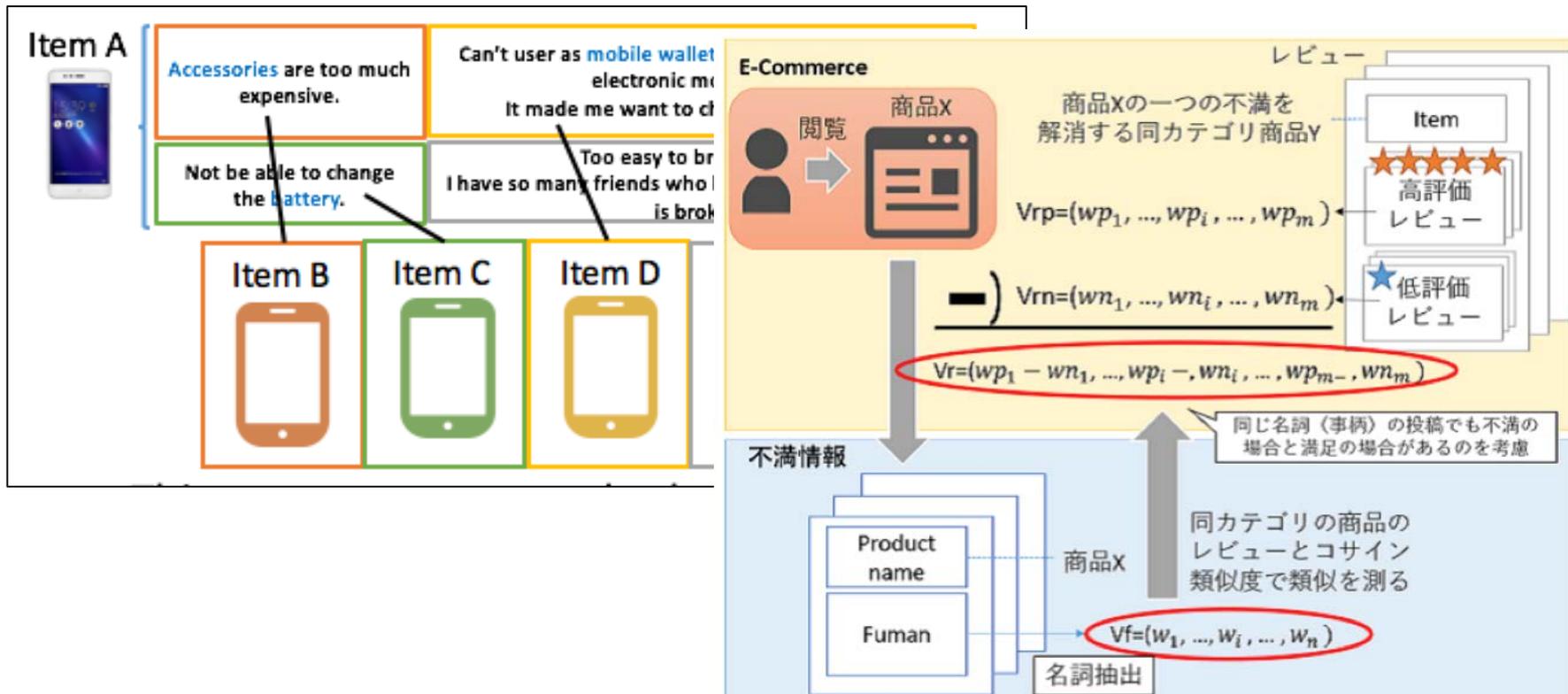
図4 「クチコミ 新発見!旅ぶら」でのツリー構造



図1 関連するタグと投稿写真と地図を同時に提示

不満データ分析事例

- E-Commerceにおける不満情報とレビューに基づいた不満解決商品推薦手法の提案[IDR'17, WebDBf2017]



レシピデータ分析事例

- レシピ推薦
 - 地域性に基づく品目に対する付け合わせマップ[IPSJ全国大会(2017)]
- ネーミング支援
 - 投稿レシピのネーミングに対する多義的コンセプト抽出[Transactions on Engineering Technologies(2014)]
- 動画レシピ推薦
 - レシピ動画の調理動作区間に基づく難易度判定とその応用[IDR'17]

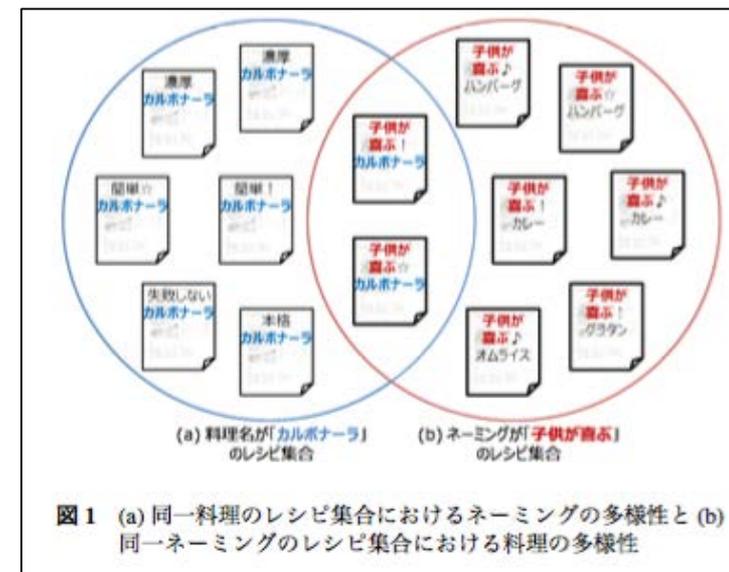
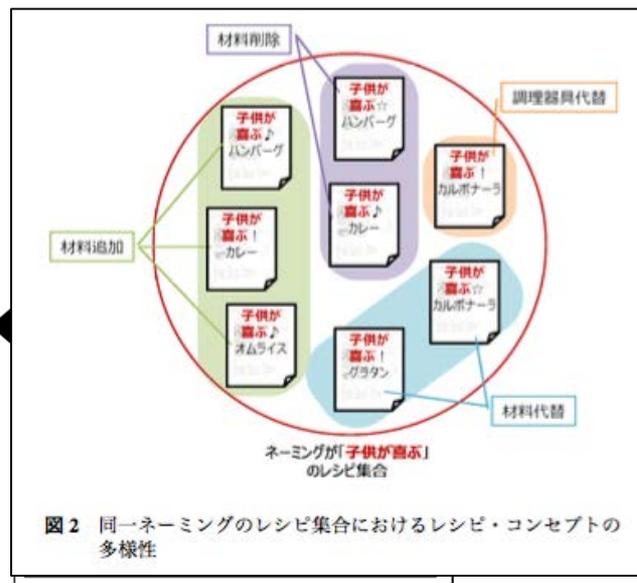


図1: 付け合わせマップのシステム概要

データコレクション

- 映像データ: 映像、音声、(closed caption)
- 地理データ: 緯度経度、建物や道名、(高さ、面積、道路幅)
- ユーザデータ
 - センサデータ: 端末ID、時間、緯度経度、(年齢、性別、MACアドレス)
 - アクセスデータ: ユーザID、時間、場所(URL)、(年齢、性別、MACアドレス)
 - ログデータ: ユーザID、クリックや入力ログ、(年齢、性別、MACアドレス)
 - 生体データ: ユーザID、時間、場所、年齢、性別
- 都市データ: 店舗情報、医療情報、事件事故

Thank you for your attention.

IDRユーザフォーラム 2017
パネルセッション

我々がデータを公開する意義と課題 + 今後の方向性



自己紹介

櫻井 一貴 Kazutaka Sakurai

株式会社リクルートテクノロジーズ
経営企画部 広報コミュニケーショングループ
兼 アドバンスドテクノロジーラボ

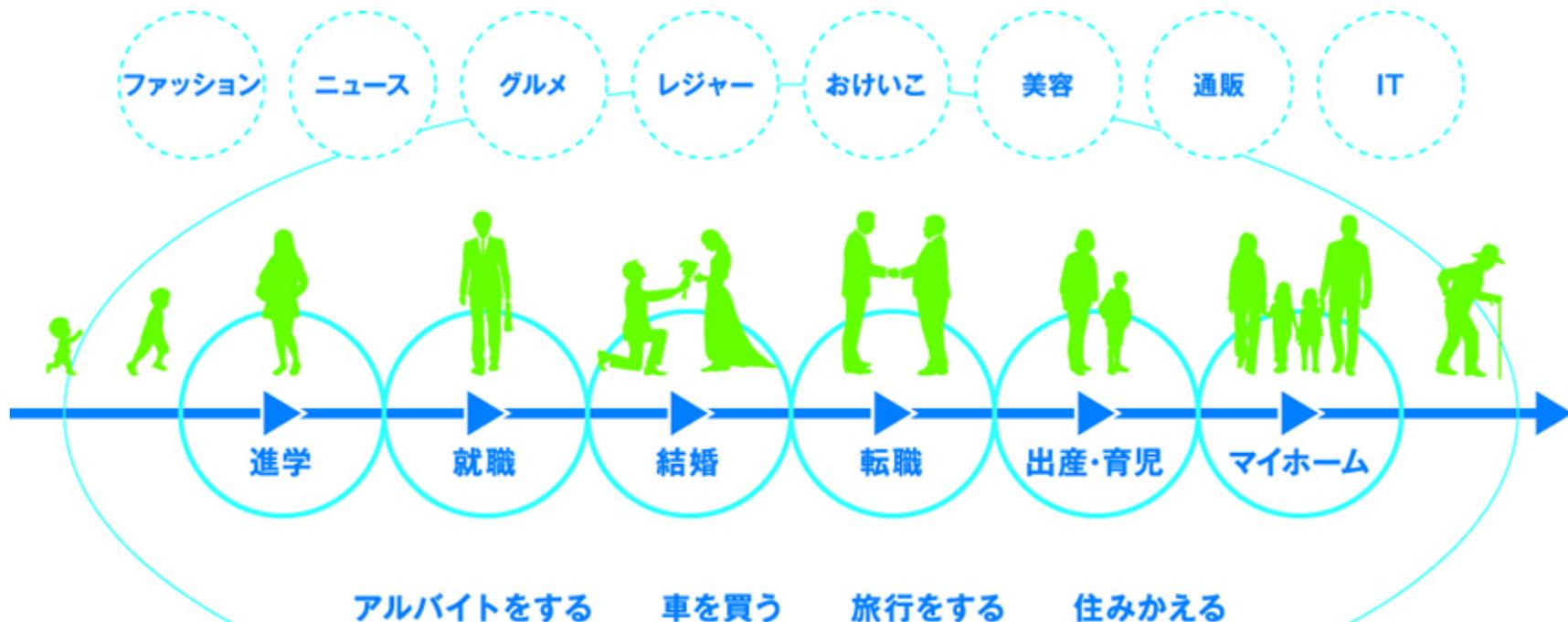
情報処理学会 DBS研究会 幹事
関西学院大学 社会情報学研究センター 客員研究員



アジェンダ

- 1 リクルートとリクルートテクノロジーズ
- 2 IDRで公開したデータセットの概要
- 3 公開する意義
- 4 データ公開に向けて感じた課題
- 5 今後に向けて

1-1 リクルートについて



私たちは、誰かと誰かを結ぶ「場」を提供しています

1-1 リクルートについて

主なビジネスドメイン



仕事



ビジネス支援



進学・教育



旅行



美容



住宅



結婚



自動車



グルメ



生活・地域情報

1-1 リクルートについて

主なビジネスドメインと代表的なサービス



リクナビ

仕事



AirREGI

ビジネス支援



スタディサプリ

進学・教育



旅行



HOT PEPPER
Beauty

美容



SUUMO
スーモ

住宅



ゼクシィ

結婚



どんなクルマと、どんな時間を。
カーセンサー

自動車



HOT PEPPER
Hot Pepper
グルメ

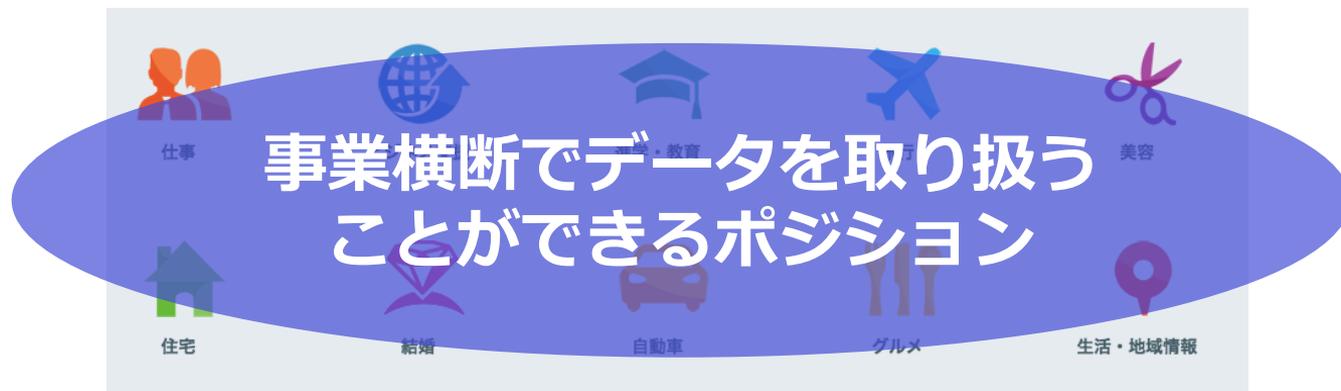
グルメ



ポンパレモール

生活・地域情報

1-2 リクルートテクノロジーズについて



Solutions (IT & Internet marketing)



Development



PC & N/W Support



Marketing Support



Infrastructure



Smart Device



Impact Analysis



Big Data Solution



Service Design



Research & Development

2 公開したデータセットの概要



国内最大級のヘアサロン・リラク&ビューティーサロン検索・予約サイト

📱 サロンをネットで予約する

シュシュシュッと探すと
パパーンって見つかる!

カンタンな
ヘアサロン探しは

HOT PEPPER
Beauty

いまなら
新規会員登録で
1,000円分の
ポイントプレゼント
Check

戻る サロン詳細

美容・美肌・美甲・脱毛・ネイル・エステ

ヘアサロン 表参道店

空席確認・予約する

本日の空席状況: ()空席あり

クーポンメニュー こだわり スタイリスト

2 公開したデータセットの概要



データ概要

国立情報学研究所が株式会社リクルートテクノロジーズから提供を受けて研究者に提供しているデータセットです。

1. ホットペッパービューティーデータ 2014-09-01 新規

ホットペッパービューティーに2012年1月～2014年1月に掲載されたデータです。各種IDはユニーク番号に変換済みで、各データ間はユニーク番号により紐付けが可能です。

- 店舗データ (約8万件)
店舗名, 住所, 店舗メッセージ, ホームページなどのデータです。TSV形式のファイルで約45MB (展開後) です。
- 店舗ブログデータ (約361万件)
タイトル, 本文, 登録日などのデータです。TSV形式のファイルで約2.4GB (展開後) です。
- スタylistデータ (約18万件)
キャリア年数, 得意技術, 自己PR, 性別などのデータです。TSV形式のファイルで約64MB (展開後) です。
- クーポンデータ (約15万件)
クーポン名, 利用条件, 有効期限, 料金などのデータです。TSV形式のファイルで約56MB (展開後) です。
- メニューデータ (約52万件)
メニュー名, 料金, 施術時間などのデータです。TSV形式のファイルで約53MB (展開後) です。
- セットメニューデータ (約7万件)
メニュー名, 料金, 施術時間などのデータです。TSV形式のファイルで約9MB (展開後) です。
- 口コミデータ (約73万件)
ニックネーム, 性別, 世代, レビュー内容, 個別および総合ポイント, 投稿日などのデータです。TSV形式のファイルで約465MB (展開後) です。

2 公開したデータセットの概要（実際の活用状況、抜粋）

主な活用状況

- ・論文「感情極性値を用いたレビューの有用性自動評価」
- ・論文「価値観ベース協調フィルタリングのユーザ特性解析による推薦性能向上に関する研究」（修士論文）
- ・口コミデータを使った時系列分析、計量テキスト分析
- ・自作自演の印象を与える口コミにどのような特徴があるかの分析

ご意見・感想

- ・たくさんの種類のデータがあって、さまざまな角度から分析できる
- ・データの形式がそろっていない部分があり、扱いにくい
- ・肯定的なレビューの数の方が圧倒的に多く、学習時などその偏りをどう克服するかが大変だった
- ・売上データなどもわかると分析の幅が広がるためありがたい

3 公開する意義

① CSRの観点

：研究レベル向上のために民間企業として資することのできる余地？

② 新しい技術やアイデアの獲得

：最先端の研究結果をビジネスに接続できないか？

③ プレゼンス向上

：分析しがいのあるデータを保有する企業であることを訴求

4 データ公開に向けて感じた課題

① リスクテイク

(パーミッション済みとはいえ“最悪のシナリオ”をどう捉えるか)

② 機能組織ならではの悩み

(「データにアクセスできる」 ↔ 「データを取り扱う権限を持つ」)

③ ROI ?

(公開準備タスクもろもろ ↔ [CSR/技術獲得/プレゼンス])

4 データ公開に向けて感じた課題

① リスクテイク

(パーミッション済みとはいえ“最悪のシナリオ”をどう捉えるか)

「公開をするとどんなリスクがあるのか、万一発生したら問題ないのか」

※クライアント、カスタマーのデータを活用するため、事業は慎重にならざるをえない



- ・ 想定範囲(学術利用)外に流出する可能性が極めて低いことを説明
- ・ 想定される「最もよくないシナリオ」を列挙、事業に共有をした

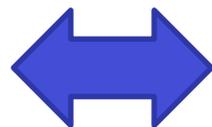
(例) USBメモリの置き忘れ等でデータが流出したら？

(例) 利用者がデータを商用利用したら？

4 データ公開に向けて感じた課題

② 機能組織ならではの悩み

(「データにアクセスできる」⇔「データを取り扱う権限を持つ」)



事業との調整によりスピード感がスポイルされることも

4 データ公開に向けて感じた課題

③ ROI ?

(公開準備タスクもろもろ ↔ [CSR/技術獲得/プレゼンス])

- 1) 協力(賛同)を得られるサービスを探す > 工数割いてまで…リスク…
- 2) 公開対象データの選定 > 個人情報含まず、パーミッション済み…
- 3) データの取得 > 自社内securityレベルも高まり、意外と手間
- 4) データ整形 > マスキング、カラム除外、暗号化…
- 5) 受け渡し

↑ここまで辿り着くのに実は数ヶ月以上も・・・

5 今後にむけて ~とはいえ前向きにデータ公開を推進するために

- ① 公開までのタスクの簡素化、効果の可視化
- ② データ多様化への対応
- ③ コントローラブルに展開 (期間や範囲を限定)

① 公開までのタスクの簡素化、効果の可視化

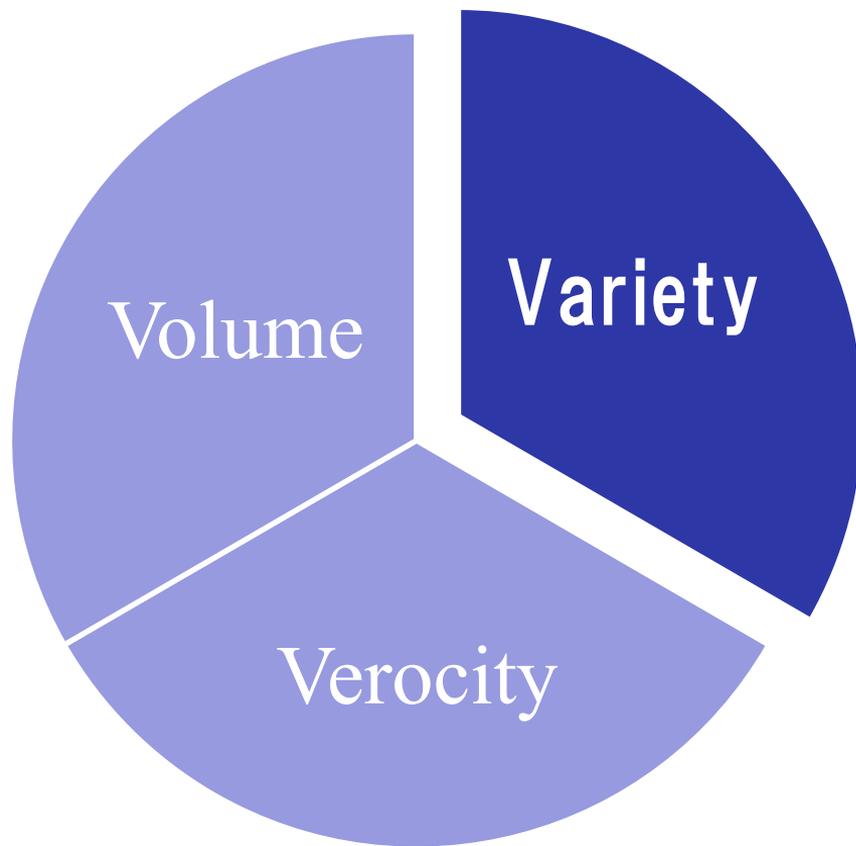
(タスク簡素化)

→ 手間・ステップをシンプルにできればより多くのデータ公開をスピーディに展開できる。ROIも向上する

(効果の可視化)

→ 「どれだけ世の役に立っているか」に加え、具体的な
ビジネス面での成果が生まれる & 可視化されると、
事業のメリットに繋がる

② データ多様化への対応 (ビッグデータの“3V”)



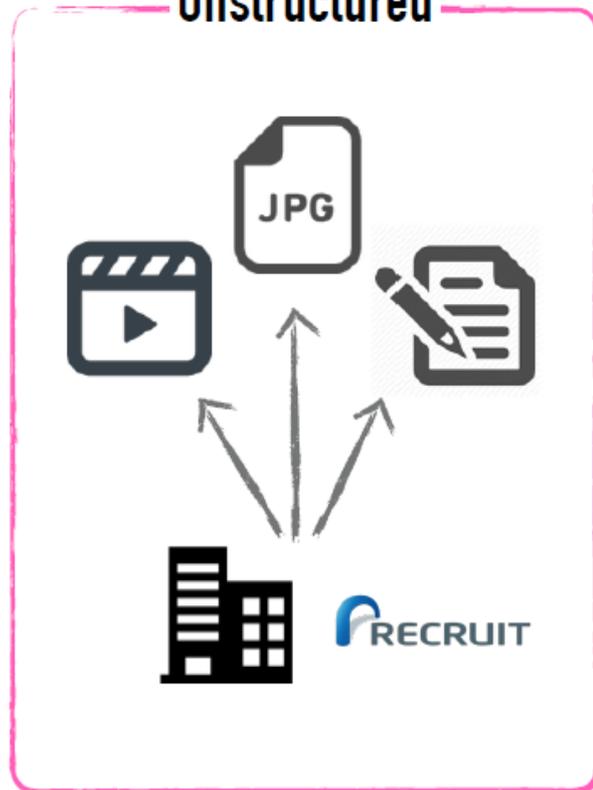
② データ多様化への対応

VARIETY

Structured



Unstructured



② データ多様化への対応

VARIETY



一致率: 98.41%

ホンダ
オデッセイ



トヨタNEW!
プリウス 1.8S ツーリングセレクション



本体	163.0万円	総額	172.3万円
年式	2016年	走行	2.0万km
車種	1.8/10	排気	なし
保証	保証付	整備	正規整備
住所	栃木県宇都宮市		

【購入プラン付】
【ディーラー保証】

電話する (無料)

トヨタNEW!
プリウス 1.8S ツーリングセレクション



本体	163.0万円	総額	172.3万円
年式	2016年	走行	2.0万km
車種	1.8/10	排気	なし
保証	保証付	整備	正規整備
住所	栃木県宇都宮市		

【購入プラン付】
【ディーラー保証】

電話する (無料)

画像や音声等のデータ提供も模索していきたい

③ コントローラブルに展開 (1: 期間限定のプログラム)

データ提供型アカデミック支援プログラム



概要

日本データベース学会（以下、DBSJ）維持会員の企業が保有する実データを、DBSJ学生会員に対し提供することで、実データを使った斬新なアイデア、萌芽的な取り組みを支援するプログラムです。

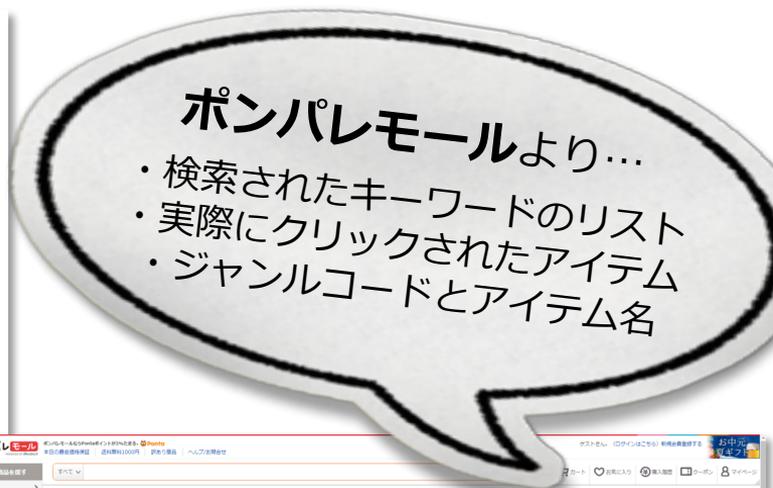
本プログラムに参加する学生会員は、本プログラムにて提供されるデータを対象とした研究テーマを自由に設定し、2017年9月に開催予定のWebDB Forum 2017にて、その成果を報告していただきます。

6月 (SoC)
プログラムスタート

(夏休み)
各自で分析、報告
資料作成

9月 (WebDBF)
成果発表
& 審査/表彰

10月～
企業オンサイトで
引き続き深掘り等



半年間かつ数大学限定のプログラム向けに提供
(事業サイドからも快諾を得られた)

③ コントローラブルに展開 (2 : 自社主催のハッカソン)

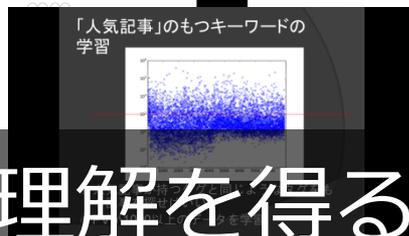


今回のデータは
コチラ!

今期の開催は
2018/1/15-19



リクルート 自然言語処理ハッカソン 🔍



事業サイドも審査員として巻き込み理解を得る

今後に向けて



仕事



ビジネス支援



進学・教育



旅行



美容



住宅



結婚



自動車



グルメ



生活・地域情報

「いかに**高いROI**を担保できるか」を意識しつつも
Variety を豊富にできないか 引き続き検討します！

(最後に…) 大事なことなのでもう一度



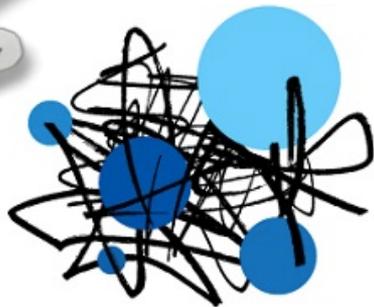
間もなくエントリー締め切りです

(~2017/12/17)

リクルート 自然言語処理ハッカソン



2017/11
新三ツシヨンステートメント
できました!



Technologies for Pleasure

Recruit Technologies Co., Ltd.

リクルートテクノロジーズ



ご清聴ありがとうございました

原島 純

2006年4月 ● 京都大学黒橋・河原研究室配属

情報検索・文書要約

- ・ 同義表現の獲得 (IJCNLP 2008)
- ・ 検索結果の要約 (NLPIX 2010)
- ・ 検索結果のリランキング (IJCNLP 2011)
- ・ 日本語文の圧縮 (COLING 2012)

2013年3月 ● 博士 (情報学) 取得

2013年4月 ● クックパッド入社

サービス開発部門

2016年4月 ● 研究開発部門 (新設) 配属

レシピ情報処理

- ・ レシピコーパスの構築 (LREC 2016)
- ・ レシピの翻訳 (WAT 2016)
- ・ レシピ画像データセットの構築 (SIGIR 2017)
- ・ 手順画像の解析 (CEA 2017)



クックパッド

インターネット上でレシピの投稿・検索等ができる
日本最大の料理レシピサービス



クックパッドデータセット

[クックパッド株式会社](#)が国立情報学研究所と協力して研究者に提供しているデータセットです。

2015/05/29 更新

データ概要

[クックパッド](#)に掲載されたデータで、172万品のレシピやそれらからなる献立に関するデータが含まれています。

1. レシピデータ 2015-02-24 新規

[クックパッド](#)で2014年9月30日までに公開されたレシピに関するデータです。レシピのタイトルや概要、手順、つくれば（「作りましたフォトレポート」の略）、カテゴリなどのデータが含まれています。

2. 献立データ 2015-02-24 新規

[クックパッド](#)で2014年9月30日までに公開された献立に関するデータです。献立のタイトルや献立に含まれるレシピ、各レシピが主菜か副菜かといったデータが含まれています。

7z形式で圧縮したMySQLのバックアップファイルで、サイズは約1.8GB（展開後は約5.5GB）です。

更新情報

- 「クックパッドデータセット」の配布を開始しました。（2015/02/24）**NEW!!**

提供対象者・利用目的

- 本データの利用目的は研究に限ります。
- IDRからは、クックパッド株式会社との契約に基づき、大学および公的研究機関の研究者を対象として提供します。利用の可否をお知りになりたい方は下記「問い合わせ窓口」（IDR事務局）までお問い合わせください。

- Yahoo!データセット
- 楽天データセット
- ニコニコデータセット
- リクルートデータセット
- クックパッドデータセット
- HOME'Sデータセット
- 不満調査データセット
- 国文研データセット
- NTCIRテストコレクション
- 音声コーパス
- 会話コーパス

公開データ (レシピ)

豚のにんにく醤油焼き

レシピを保存



にんにく醤油でご飯がすすむ、スタミナおかずです。

スタッフあこ

材料 (3人分)	
豚薄切り肉	200g
シメジ	1/2パック
★醤油	大匙1.5
★おろしにんにく	大匙1
★みりん	大匙1
★酒	大匙1
★ハチミツ (砂糖でも可)	小匙1
もやし (お好みで)	1パック
炒め用サラダ油	大匙1~
粗挽き黒胡椒	適量
アサツキや小葱など	適量

- 1 もやしはサッと茹でてザルにあけておく。★の調味料は予め合わせておく。
- 2 フライパンにサラダ油を熱し、小房に分けたシメジを炒める。しんなりしたら豚薄切り肉を入れて、肉の色が変わるまで炒める。
- 3 ★の調味料を回し入れ、味を絡めるように少し煮詰める。
- 4 適度に汁気が少なくなったところで出来上がり。お皿に盛り、お好みで粗挽き黒胡椒、アサツキなどを振ってください。

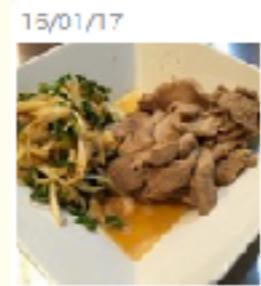
コツ・ポイント
汁気を少し残して、茹でたもやしと一緒にモリモリ食べるのが好きです。菜味はあさつや葱以外にも、茗荷や大葉でも合います。白胡麻もいいね。

このレシピの生い立ち
にんにく系の味つけが大好きな夫のために作ったところ、彼も気に入ったようです。

レシピID: 2232471 公開日: 13/05/23 更新日: 13/05/23

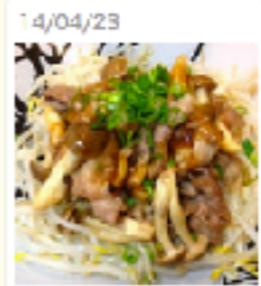
- ・ タイトル
- ・ 概要
- ・ 材料
- ・ 手順
- ・ コツ・ポイント
- ・ このレシピの生い立ち
- ・ つくれぽ
- ・ etc

みんなのつくりましたフォトレポート「つくれぽ」 3件 (3人)



15/01/17
生姜焼きも美味しいけどニンニク焼きも美味しいですね！
Sun Flower

素敵な味と盛りつけですね♪つくれぽありがとうございます！



14/04/23
もやしのシャキ②とニンニク醤油のお肉でご飯食べ過ぎました♪美味
sqqs0725

素敵なつくれぽありがとうございます！美味しく食べてくださいね♪



13/12/19
メインで奥にたっぷり♡食卓そられまくり香りも味も美味でした♡
Marrietty

初れぽありがとうございます♡

公開データ (献立)

ドライカレープレート

 ryoketsuma



この献立のポイント

ドライカレーはこってりした味なので、さっぱりした金柑のサラダで味のバランスを取ります！

調理時間

約 **40** 分

段取りのコツ

カレーを炒めている間にサラダの材料を切っておきましょう。

主



副



- ・ タイトル
- ・ この献立のポイント
- ・ 調理時間
- ・ 段取りのコツ
- ・ 主菜
- ・ 副菜
- ・ 参考になった！
- ・ この献立の主な食材/ジャンル/シーン

この献立が「参考になった！」と言っている人

10人

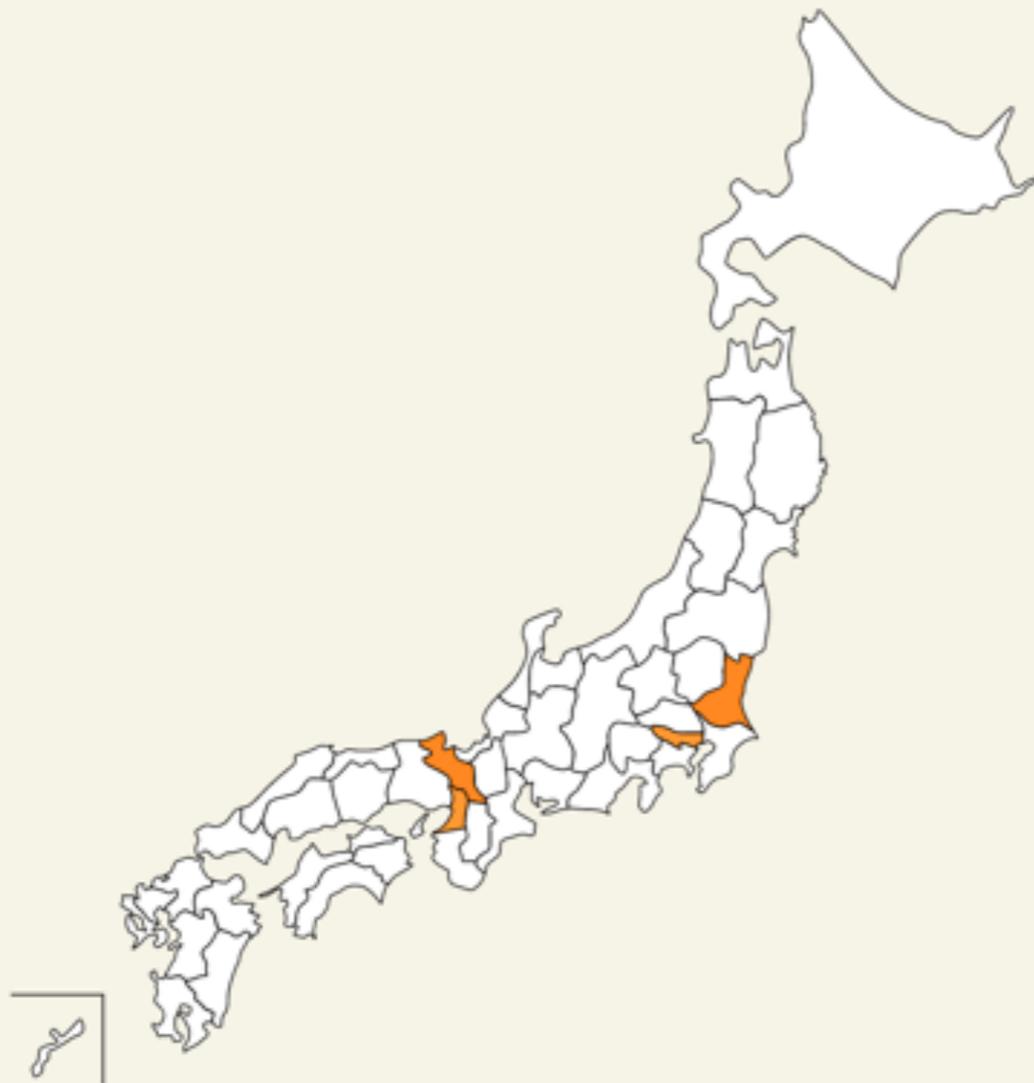
 参考になった！



この献立の 主な食材 / ジャンル / シーン

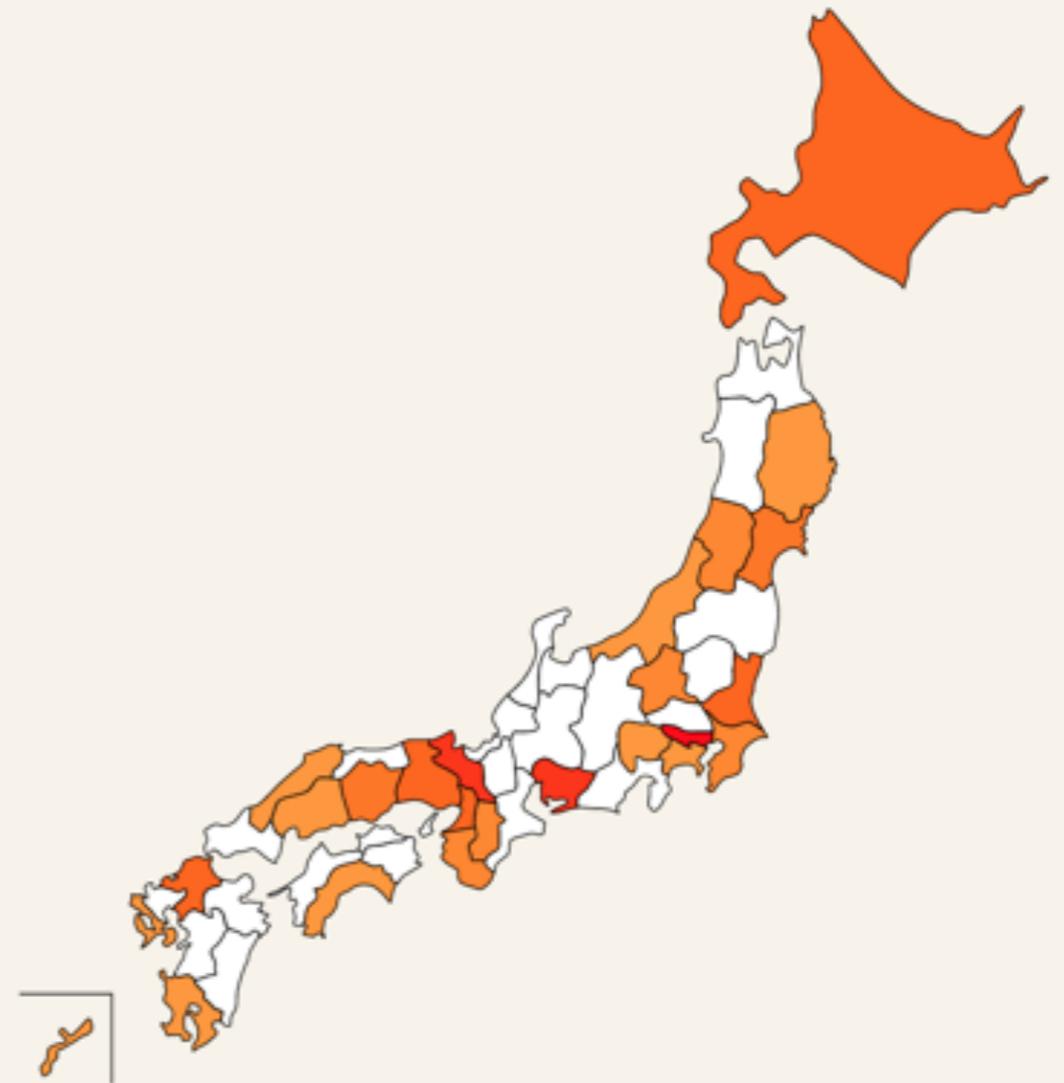
[ごはんもの](#) [洋風](#) [ごはん](#)

利用状況



公開前 (～2015年2月)

5 大学 6 研究室



公開後 (2017年12月)

82 大学 135 研究室

不満買取センターとは？

いわゆる「ソーシャルリスニングツール」
「世間のあらゆる不満を**買取**してくれる」Webサービス



みんなの不満

ピックアップ

あしたのふーまん
30代 男性
2018/08/07 17:00

暮らし・住まい・
家電器具のデザイン
私費ができる会社がかなり増えたのに、物
に似て別物向けのデザインがあまりない
改善する

2018

日程あるある
30代 女性
会社・業務・不満解消
通勤時間の短縮
通勤時間に電車が多く遅れるので
あまりない。通勤時間のみ、〇分
ほしいのよ。

その不満
買います!

¥

不満買取センターへようこそ
あなたの不満を
買い取ります。

企業の商品やサービス、そして社会をより良くするために
あなたの不満を買い取ります。
あなたの不満の中に眠った「ヒント」を
私たちが企業や社会に届けます。

不満買取センター

検索

Insight Techではよく分析作業を実施



投稿文データを自社で分析するから、データだけほしいな



分析する手間をかけられないから、分析した結果を教えてくださいな

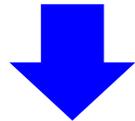


不満インサイトデータ販売

常温で保存可能なヨーグルトがあればいいのに。毎日食べるのだから。

保育園に関する情報でこんなに探しく近隣の待機児童の保育の料金などがたらいいのに。

店で買ったショーツの内側にタグが付いている。取りにくいし、そのまま履くとおしりがかゆい。



不満インサイトサーベイ



分析レポートの一例

一般向けに分析情報の公開も実施しています

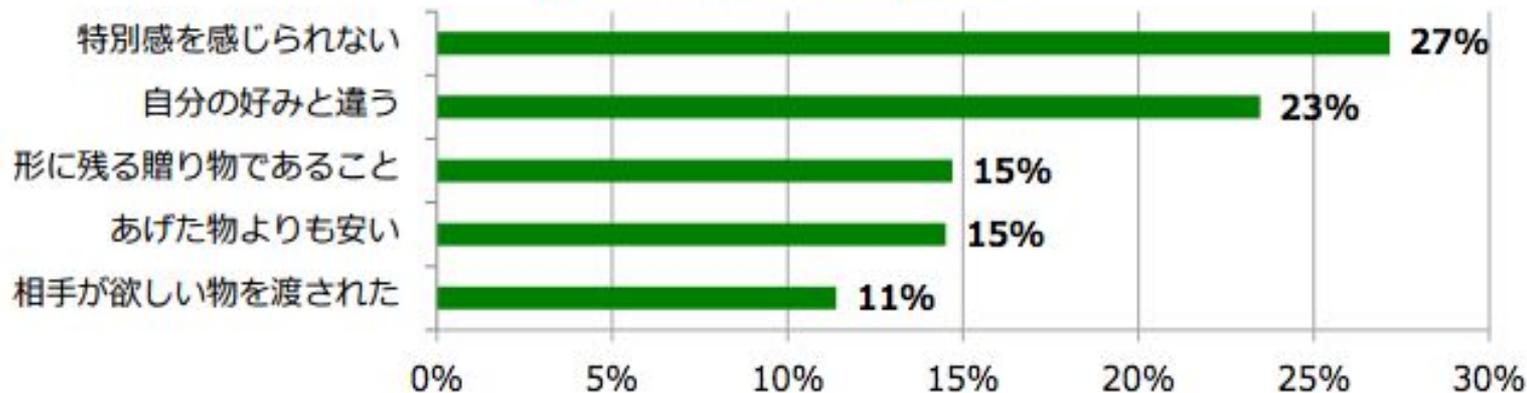
ホワイトデーへの不満調査レポート

不満買取センター ホワイトデー

検索

～選択は良くても特別感がないと喜ばれない!?!～

贈り物の不満理由 トップ5



Insight Techとして嬉しい研究とは？

分析者がデータを俯瞰するときの補助になること

分析者のおおまかな作業フロー

分析課題
の検討

不滿意見投稿
サンプリング

意見
観察

集計軸
決定

ラベリング &
集計

分析
レポート執筆

レポート完成



Insight Techとして嬉しい研究とは？

分析者がデータを俯瞰するときの補助になること

分析者のおおまかな作業フロー

分析上で多くの時間がかかる部分

分析課題
の検討

不滿意見投稿
サンプリング

意見
観察

集計軸
決定

ラベリング &
集計

分析
レポート執筆

レポート完成



Insight Techとして嬉しい研究とは？

分析者がデータを俯瞰するときの補助になること

分析者のおおまかな作業フロー

分析課題
の検討

不滿意見投稿
サンプリング

意見
観察

集計軸
決定

ラベリング
集計

分析
レポート執筆

分析上で多くの時間

この部分の時間が短縮できると
とてもありがたい！

REPORT

Insight Techとして嬉しい研究とは？

分析者がデータを俯瞰するときの補助になること

分析者のおおまかな作業フロー

分析課題
の検討

不滿意見投稿
サンプリング

意見
観察

集計軸
決定

ラベリング
集計

分析
レポート執筆

分析上で多くの時間

Insight Techに限らず
テキスト分析者全般的に
役にたつのではないかな？

REPORT

Insight Techとして嬉しい研究とは？

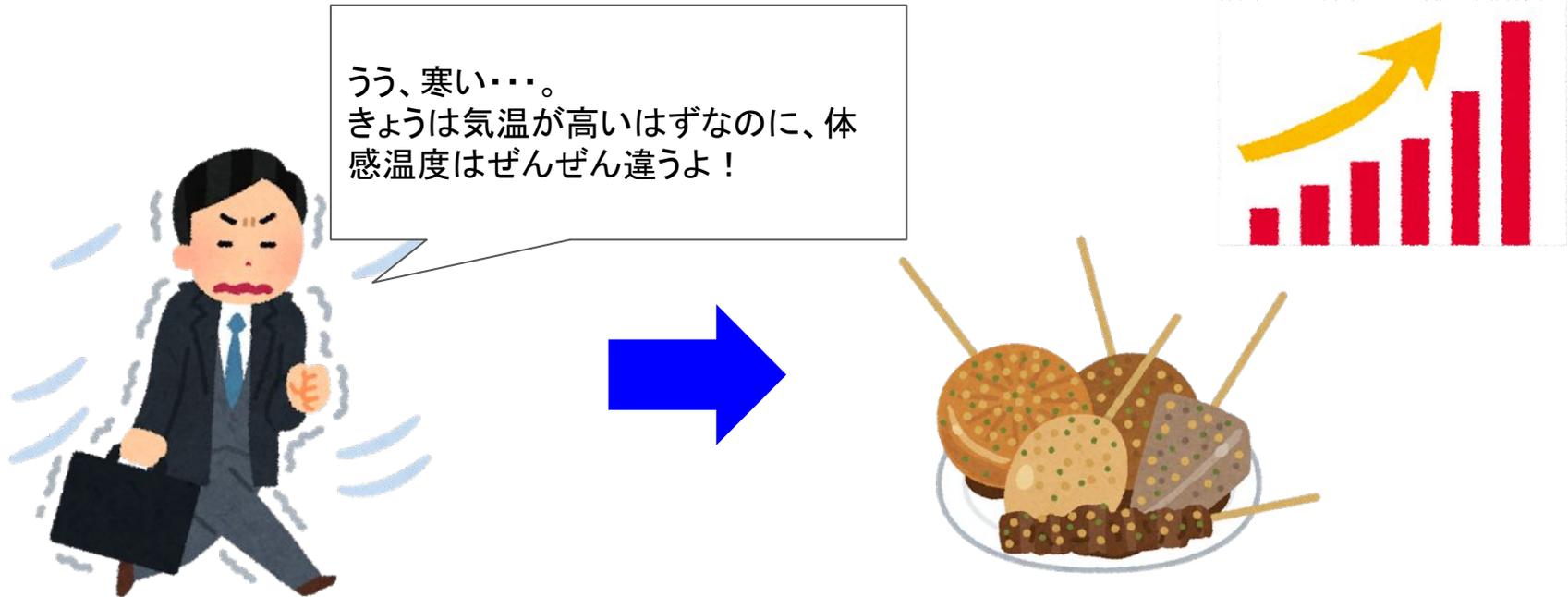
不満投稿と景気指標の関係性

1日に数万件寄せられる不満。
きっとその中には人の消費行動に関する投稿も
多く潜んでいるはず。



Insight Techとして嬉しい研究とは？

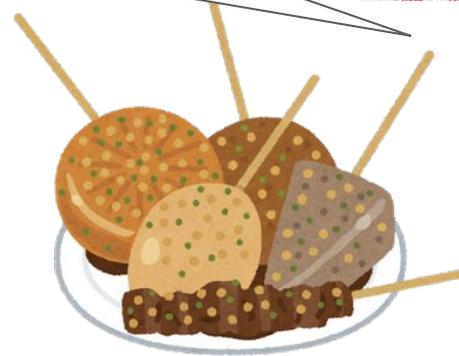
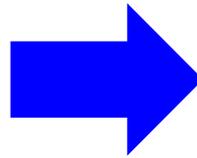
不満投稿と景気指標の関係性で例えば・・・



Insight Techとして嬉しい研究とは？

不満投稿と景気指標の関係性で例えば・・・

気温だけでは予測できなかった人々の「体感温度」で、おでんの売り上げ増加を予測



ディスカッション

- アカデミアが欲しいデータセットは？
- 複数データセットを用いた”ハイブリッド”研究？
- 会場からのトピック