

テキストマイニングの実習1

— 形態素解析を利用した集計と分析 —

2015/7/5

ビジネス科学研究科
経営システム科学専攻

テキストマイニング

- テキストデータから有益な情報を抽出する技術の総称
- テキストデータの例
 - 従来
 - 営業日報
 - 自由記述のアンケート
 - 最近 → ビッグデータ活用
 - レビューサイトの口コミ
 - ブログ
 - マイクロブログ (Twitter, Facebook)

ビッグデータ活用とは

- 捉え方は様々…
 - データの利用者やそれを支援するサービスの提供者などの観点によって定義も異なる
- ここでは…
 - 多種多量なデータを生成・収集・蓄積をリアルタイムで行い
 - このデータを利用して,未来の予測や異変の察知等のデータ分析を行うことで
 - 利用者の個々のニーズに即したサービスの提供, 業務の効率化, 新サービスの創出などに活かす取り組み

ビッグデータ活用の拡大

- ・コンピュータの処理能力の向上やストレージ容量の拡大 → ビッグデータの収集・分析可能
- ・社会問題の解決やマーケティング戦略立案などのビジネスに活かす取り組みが活発化

ビーコン/プローブ等の情報を使った交通渋滞予測

複雑な条件を満たした最適な人材配置

大量の購買履歴の分析による商品レコメンド

電子カルテ分析による生活習慣病の重症化プロセス抽出

複数のセンサ情報の分析による急変患者の兆候把握

口コミやクレーム分析による企業リスク回避

- ・情報源のひとつとしてSNS(ソーシャルメディア)上のテキストデータも注目されている

口コミの浸透 (1/2)

- スマフォやSNSの普及により,SNSの口コミやレビューサイトの重要性が増加
 - 口コミの影響力調査
 - 商品購入時に参考とする情報・広告:
 - 購入サイト・レビューサイトの口コミが 47.9%
 - SNSでの口コミ: 17.2%
(スマートフォン保有者 n=535)
 - 影響を受けやすいSNSがある: 全体の20.8%
 - 特に,若年層を中心にSNSの口コミが浸透
 - 10代女性は49.4%, 10代男性は33.7% で高い傾向
(各 n=83)

(上) 総務省「ICTの進化がもたらす社会へのインパクトに関する調査研究」(平成26年)

(下) 日本通信販売協会「ネット通販に関する消費者実態調査2013」

口コミの浸透 (2/2)

- ネット上に膨大に蓄積されている製品やサービスに関する消費者の評価情報

消費者: 有用な情報取得・共有ツール

企業: 消費者の評判に関する情報源



- 企業などでは、自社のビジネスに役立ちそうな情報の収集と分析が重要

口コミサイトの例



- ・ ホテルの口コミ数: 780万件

The screenshot shows the Rakuten Travel website interface. At the top, there's a navigation bar with links like '国内', '海外', '懸賞広場', '割引クーポン', 'グループ予約', '観光案内', 'ヒツブヤク', and 'ログ等で紹介する'. Below the navigation is a green banner with text about entering the Rakuten Card for 2,000 points and links to 'お客様の声', '個人ページ', '予約の確認・変更・取消', 'ヘルプ', '楽天トラベルの使い方', and 'サイトマップ'. A large green header says 'お客様の声' (Customer Reviews) with the number '7,810,155件' (7,810,155 reviews). Below this are several user quotes with small profile pictures. A search bar allows users to search for reviews by location ('国内宿泊' or '海外ホテル') and keyword. A section titled '新着!最新のクチコミ' (Newest! Latest reviews) lists reviews from June 19, 2015, such as '川崎グリーンプラザホテルのクチコミ (689件)' with a rating of 3.58 stars. Another section shows '投稿されたクチコミの画像' (Images of posted reviews) with thumbnails of various travel experiences. On the right side, there's a sidebar with a QR code for mobile submission, a button to 'さっそく投稿する' (Post now), and sections for recommended reviews like '楽天トラベル: 伊東温泉、伊東ホタルニュー岡部 クチコミ・感想・情報' and '楽天トラベル: ベンソン 風の季 クチコミ・感想・情報'.

口コミサイトの例



R 鶴川シーウールドホテル クラ X

HARADA Tomohiko

travel.rakuten.co.jp/HOTEL/2910/review.html

鶴川シーウールドホテル

★お部屋★

●鶴川シーウールドホテル

★レストラン★

●鶴川シーウールドホテル

■温泉大浴場★

●鶴川シーウールドホテル

■室内施設★

●鶴川シーウールドホテル

★よくある質問★

●鶴川シーウールドホテル

★アクセス★

●鶴川シーウールドホテル

設備・アメニティ・基本情報

●鶴川シーウールドホテル

写真・画像

●鶴川シーウールドホテル

地図・アクセス

●鶴川シーウールドホテル

クチコミ

●鶴川シーウールドホテル

温泉

外国語サイト

●Book Kamogawa Sea

World Hotel

●Hotels in

Sotobo(Kamogawa,
Katsuura, Onjuku, Mabora)

●KAMOGAWA SEA WORLD
HOTEL 京葉

●外房(鶴川)・勝浦・御宿・茂原)酒店一覧

総合★★★★★ 2

投稿者さんの 鶴川シーウールドホテル のクチコミ (感想・情報)

投稿者さん

2015年06月11日 17:03:57

良かったところ

・部屋からの景色（朝日最高でした）

・食事（品数が多く、朝クとも良かったです）

・フロントの方の対応（お姉さんがとても頑張っていました）以上。

掃除が行き届いているとの口コミを多く見ました

が、そこは思いました。

気かるかと思ははありましたが、フロントのお姉さんが一生懸命で、その笑顔に救われた思いです。

レビューを評価して 選ばれ理由 レビューを報告するこのレビューは参考になりましたか？

[よい] [悪い]

旅行の目的 … レジャー

同伴者 … 家族

宿泊年月 … 2015年06月

ご利用の宿泊

プラン

【洋室 禁煙・特別室】 お部屋からシャツライルカも見える シーウールドと海一宿泊プラン

ご利用のお部屋

【iワズシーウールドが見える特別室禁煙】

総合★★★★★ 4

投稿者さんの 鶴川シーウールドホテル のクチコミ (感想・情報)

投稿者さん

2015年06月11日 07:33:49

夫、2歳娘と5ヶ月の子どもの4人で宿泊しました。

【立地】当たり前ですが鶴川シーウールドにとても近く、ゆっくり窓内を見できました。

【部屋】至って普通です。（古いから、壁の声は少し聞こえます。）トイレ掃除などはしっかりされていました。清淨機などもTEL一本ですぐに届けて下さりました。

【食事】夜朝共にバイキング。イヌですが子ども用バス、エプロン、ベビーベッドを用意して下さいまして。キッズスペースも食事時間中に専門のスタッフの方がおりゆっくり食事ができました。

【風呂】小さな子ども(赤ちゃん)用のグッズペリーベッド、コーナー、バス、おもちゃ、泡ソーフ、支えのいろいろが揃っていました。お父さん達にも多く気兼ねなく楽しめました。しかしお風呂入りひとつないので、温泉を楽しむいう雰囲気ではなく、銭湯のお湯が温泉という感じです。

また、23時頃にお風呂に行くと、アメニティやシャンプーが空だったのは少し残念でした。

【サービス】受付スタッフの皆さんとともに親切、丁寧です。チェックアウト後に子どものものを冷蔵庫にいておいて欲しいとダメ元で頼むと快く入

鶴川シーウールドホテル

2015年06月11日 19:25:48

この度は、ご利用頂きまして誠にありがとうございました。

客室内清掃の件、大変申し訳ございませんでした。

重要な改善として、早急に対応いたします。

今後は、この様な事の無いよう、清掃・点検を強化いたします。

フロントスタッフへのお言葉、誠にありがとうございました。

モチベーションアップに繋がりますので、お客様からの声として、スタッフと共有させて頂きます。

機会がございましたら、またご利用をお待ちしております。

いい値！バリュープラン

【最安料金（目安）】 10,186円～

(消費税込11,100円～)

【当日15:50からアシカと記念写真】笑うアシカ

と一緒にバス付プラン

【最安料金（目安）】 10,278円～

(消費税込11,100円～)

【当日13:40～エコ・アクアローム&コミュニケーションタイム】1日3回開催

【最安料金（目安）】 10,278円～

(消費税込11,100円～)

【夜の水族館接続付】3月～10月の火・木曜日限

定プラン

【最安料金（目安）】 10,278円～

(消費税込11,100円～)

【当日14:50からイルカと一緒にバス付】2室

【最安料金（目安）】 10,463円～

(消費税込11,300円～)

今しかない！★アピア料

付バス付・スイート

【最安料金（目安）】 10,463円～

(消費税込11,300円～)

便利な赤ちゃんグッズ付バス付・お部屋はお母さん

も入浴可★赤ちゃんちゃん

が寝る付プラン

【最安料金（目安）】 10,926円～

(消費税込11,800円～)

便利な赤ちゃんグッズ付バス付・お部屋はお母さん

も入浴可★赤ちゃんちゃん

が寝る付プラン

【最安料金（目安）】 10,926円～

(消費税込11,800円～)

お子様にも大好評！オーシャンピューラン

【最安料金（目安）】 11,112円～

(消費税込12,000円～)

【80cmのジャングルサイズ】海の王者アザラシぬいぐるみ付プラン

【最安料金（目安）】 11,204円～

(消費税込12,100円～)

宿泊料金+タマーバー料

+マーブラーフードランチ付プラン

【最安料金（目安）】 11,389円～

(消費税込12,300円～)

【当日14:50～ルルカ

2015/7/2, 9

テキストマイニング

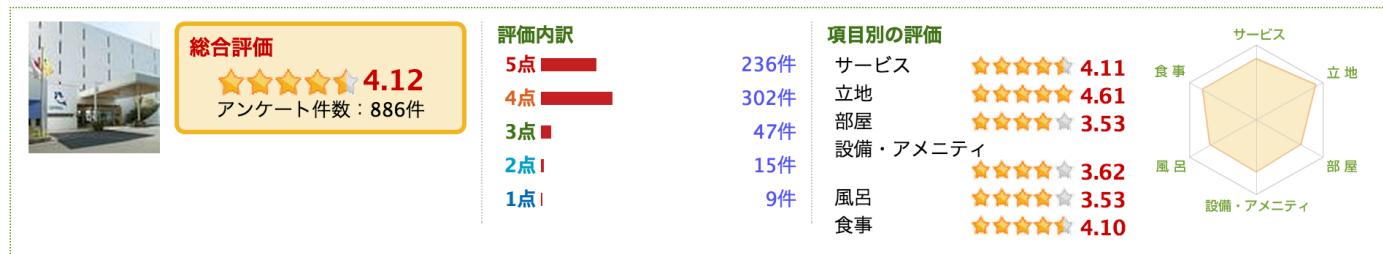
口コミサイトの例



施設紹介 プラン一覧 フォトギャラリー(76) 地図・アクセス お客様の声(886) クーポン一覧 プレゼント

鴨川シーワールドホテルのクチコミ・お客様の声

●ホテル・旅行のクチコミTOPへ



総合 ★★★★★ 2
投稿者さんの 鴨川シーワールドホテル のクチコミ (感想)

投稿者さん 2015年06月11日 17:03:57

良かったところ

- ・部屋からの景色（朝日最高でした）
- ・食事（品数も多く、朝夕とも良かったです）
- ・フロントの方の対応（お姉さんがとても頑張っていました）以上。

掃除が行き届いているとの口コミを多く見ましたが、そ
うは思いませんでした。

気にかかることは多々ありましたが、フロントのお姉さ
んが一生懸命で、その笑顔に救われた思いです。

評価

評価	総合	★★★★★ 2
サービス	2	
立地	4	
部屋	4	
設備・アメニティ	2	
風呂	2	
食事	4	

旅行の目的

- … レジャー

同伴者

- … 家族

宿泊年月

- … 2015年06月

情報)

鴨川シーワールドホテル 2015年06月11日 19:32:50

この度は、ご利用頂きまして誠にありがとうございました。

客室内清掃の件、大変申し訳ございませんでした。
重要改善として、早急に対応いたします。
今後は、この様な事の無いように、清掃・点検を
強化いたします。

フロントスタッフへのお言葉、
誠にありがとうございます。
モチベーションアップに繋がりますので、
お客様からの声として、
スタッフと共有させて頂きます。

機会がございましたら、またご利用をお待ちしております。

テキストマイニングの手順

1. データの把握 (定量的な分析)

- データの特徴を定量的に把握する
→データの総件数,分類(属性)ごとの件数や割合
 - 旅行目的別の人気エリアは?
 - 同伴者別の人気エリアは?
 - レーティングによる人気エリアの差異は?

2. 分析テーマの設定

- 分析の目的を明確にする →仮説を立てる

3. テキストデータの分析 (定性的な分析)

- データ把握と仮説に基づくテキストの分析

演習 — 準備

- データをダウンロードしてください
 - EXCEL 形式
 - <https://github.com/haradatm/gssm-201507/raw/master/data/rakuten-eval.xlsx>
 - 演習は上記の EXCEL ファイルで説明しますが,テキストファイルもダウンロードできます
 - タブ区切り,UTF-8
 - <https://github.com/haradatm/gssm-201507/raw/master/data/rakuten-eval-utf8.txt>
 - タブ区切り,シフトJIS
 - <https://github.com/haradatm/gssm-201507/raw/master/data/rakuten-eval-sjis.txt>

参考 — 講義用データ

<https://github.com/haradatm/gssm-201507>

The screenshot shows a GitHub repository page for 'gssm-201507'. The commit history for the 'master' branch is displayed, starting with a 'first commit' by Tomohiko HARADA 5 minutes ago. The commits are listed in descending order of age, all being 'first commit' for each file. The files listed are 00-slides, 01-data, 02-tools, 03-samples, .gitignore, and README.md, all committed 5 minutes ago.

File	Commit	Time
00-slides	first commit	5 minutes ago
01-data	first commit	5 minutes ago
02-tools	first commit	5 minutes ago
03-samples	first commit	5 minutes ago
.gitignore	first commit	5 minutes ago
README.md	first commit	5 minutes ago

00-slides:

lecture-23.pdf

01-data:

rakuten-eval.xlsx

rakuten-eval-sjis.txt

rakuten-eval-utf8.txt

02-tools:

mecab-0.996.exe

mecab-0.996.tar.gz

mecab-naist-jdic-0.6.3b-20111013.tar.gz

03-samples:

cmdline-mac.txt

cmdline-win.txt

lecture-2_summary.xlsx

lecture-3_text-words.zip

lecture-3_frequency.xlsx

lecture-3_tf-idf.xlsx

misc/lecture-2.r

misc/lecture-2.pdf

演習 — 使用データ

楽天トラベルの口コミデータ

<http://travel.rakuten.co.jp/>

- 外房 (鴨川・勝浦・御宿・茂原) … 1,918件
 - <http://travel.rakuten.co.jp/yado/chiba/sotobo.html>
- 西伊豆 (戸田・土肥・堂ヶ島・松崎) … 3,051件
 - <http://travel.rakuten.co.jp/yado/shizuoka/nishi.html>
- 南房総 (館山・白浜・千倉) … 1,348件
 - <http://travel.rakuten.co.jp/yado/chiba/tateyama.html>

演習 — データの把握

- データ項目 (左から順に)

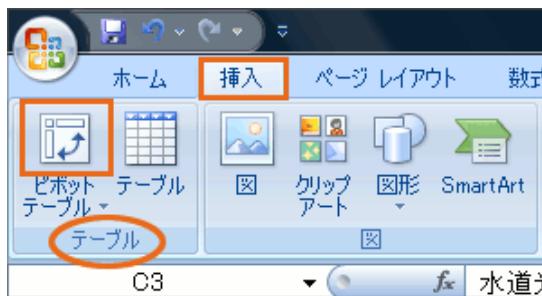
施設情報	3項目	エリア, 施設コード, 施設名
口コミ	1項目	テキスト
ユーザー評価	7項目	総合, サービス, 立地, 部屋, 設備・アメニティ, 風呂, 食事
その他の分類	2項目	旅行の目的, 同伴者
宿泊日	1項目	宿泊年月
ユーザー情報	3項目	ユーザー, 年代, 性別, 投稿回数

- データ集計で分かること
 - 投稿者の属性(年代,性別)は?
 - 旅行目的別の人気エリアは?
 - 同伴者別の人気エリアは?

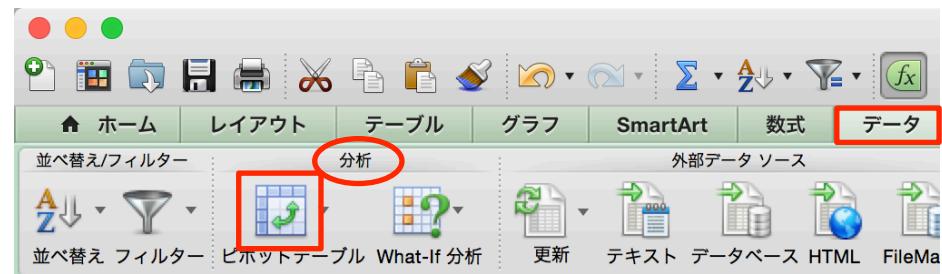
演習 — データの把握

- EXCEL のピボットテーブルを使う
 - ファイル rakuten-eval.xlsx を開く
 - A～R 列を選択し, ピボットテーブルを作成

Windows



Mac



【Windows】 Excel 2007・2010・2013
[挿入] タブ [テーブル] グループの [ピボットテーブル] ボタンをクリックします

【Mac】 Excel 2011
[データ] タブ [分析] グループの [ピボットテーブル] ボタンをクリックします

演習 — データの把握

(エリアごとのデータ件数)

データの個数 : 施設コード	
行ラベル	計
01-外房	1,918
02-西伊豆	3,051
03-南房総	1,348
総計	6,317

(年代別の構成)

データの個数 : 施設コード		列ラベル	01-外房	02-西伊豆	03-南房総	総計
行ラベル						
10才未満			0.10%	0.00%	0.00%	0.03%
10代			0.05%	0.13%	0.00%	0.08%
20代			6.31%	6.46%	5.34%	6.17%
30代			30.92%	24.55%	24.63%	26.50%
40代			34.93%	34.81%	36.65%	35.24%
50代			18.98%	24.09%	23.22%	22.35%
60代			7.19%	8.16%	9.35%	8.12%
70代			1.46%	1.67%	0.59%	1.38%
80代			0.05%	0.07%	0.22%	0.09%
100代			0.00%	0.07%	0.00%	0.03%
総計			100.00%	100.00%	100.00%	100.00%

(旅行目的の構成)

データの個数 : 施設コード		列ラベル	01-外房	02-西伊豆	03-南房総	総計
行ラベル						
レジャー			89.73%	95.51%	93.84%	93.40%
ビジネス			5.89%	0.95%	1.41%	2.55%
その他			1.56%	0.88%	1.56%	1.23%
na			2.82%	2.65%	3.19%	2.82%
総計			100.00%	100.00%	100.00%	100.00%

(性別の構成)

データの個数 : 施設コード	列ラベル	01-外房	02-西伊豆	03-南房総	総計
行ラベル					
男性		60.38%	65.13%	61.80%	62.97%
女性		39.62%	34.87%	38.20%	37.03%
総計		100.00%	100.00%	100.00%	100.00%

(年代別性別の構成)

平均 : 投稿回数	列ラベル	男性	女性	総計
行ラベル				
10才未満		1.0	1.0	1.0
10代		1.0	1.0	1.0
20代		3.0	2.2	2.5
30代		6.9	3.4	5.2
40代		9.7	4.6	7.9
50代		12.9	6.6	11.1
60代		11.1	6.9	10.3
70代		11.4	13.7	11.6
80代		2.8	5.0	3.2
100代		1.0	1.0	1.0
総計		9.8	4.4	7.8

(同伴者の構成)

データの個数 : 施設コード	列ラベル	01-外房	02-西伊豆	03-南房総	総計
行ラベル					
家族		68.98%	69.88%	71.14%	69.87%
一人		14.60%	10.13%	10.91%	11.65%
恋人		6.41%	9.67%	8.90%	8.52%
友達		5.27%	5.77%	4.15%	5.27%
仕事仲間		1.77%	1.11%	1.11%	1.31%
その他		0.16%	0.79%	0.52%	0.54%
na		2.82%	2.65%	3.26%	2.83%
総計		100.00%	100.00%	100.00%	100.00%

演習 — ユーザー評価の傾向

- 評価点による人気エリアの差異は?

(総合評価)

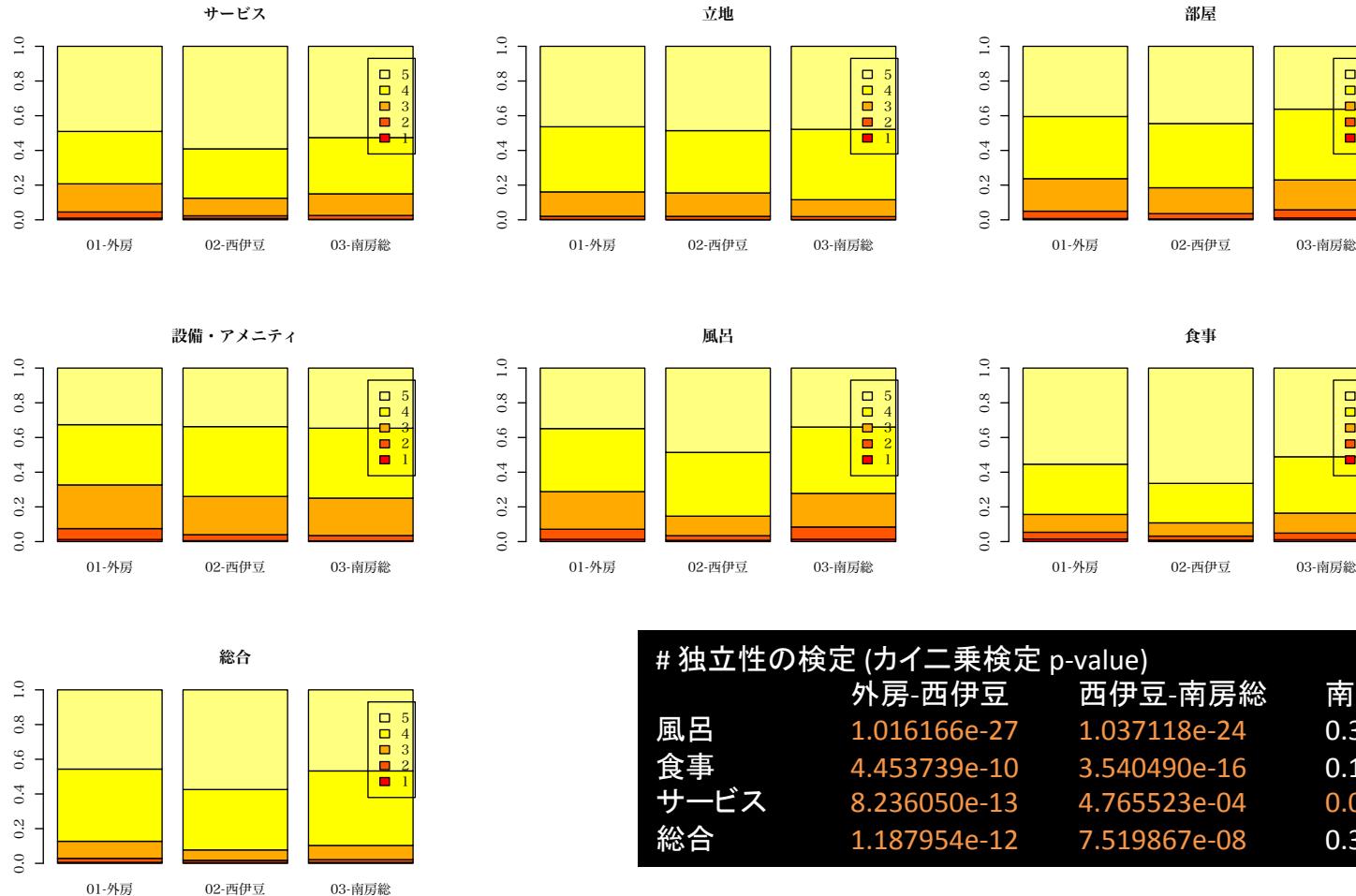
行ラベル	01-外房	02-西伊豆	03-南房総	総計
1	0.52%	0.52%	0.52%	0.52%
2	2.29%	1.21%	1.63%	1.63%
3	9.49%	6.72%	8.16%	7.87%
4	41.45%	34.97%	42.95%	38.64%
5	46.25%	56.57%	46.74%	51.34%
総計	100.00%	100.00%	100.00%	100.00%

(評価項目別)

	01-外房	02-西伊豆	03-南房総	全体
サービス	4.25	4.41	4.35	4.35
立地	4.31	4.29	4.34	4.31
部屋	4.11	4.23	4.06	4.16
設備・アメニティ	3.92	4.03	4.06	4.00
風呂	3.99	4.33	3.96	4.15
食事	4.33	4.48	4.29	4.39
総合	4.31	4.46	4.34	4.39

参考 — ユーザー評価の傾向

- その差は有意か? (独立性を検定する)



独立性の検定 (カイ二乗検定 p-value)

	外房-西伊豆	西伊豆-南房総	南房総-外房
風呂	1.016166e-27	1.037118e-24	0.3187078313
食事	4.453739e-10	3.540490e-16	0.1245526182
サービス	8.236050e-13	4.765523e-04	0.0004802117
総合	1.187954e-12	7.519867e-08	0.3299232019

参考 — 前頁の図と計算例 (R)

```

path1<- "data/rakuten-eval.txt"
data<-read.table(path1,header=T,sep='\t',row.names=NULL)
dim(data); names(data);
head(data[,names(data)!=="テキスト"])
summary(data[,names(data)!=="テキスト"])

# 比率のプロット
t0<-prop.table(xtabs(~data$"総合" + data$"エリア"),margin=2)
t1<-prop.table(xtabs(~data$"サービス" + data$"エリア"),margin=2)
t2<-prop.table(xtabs(~data$"立地" + data$"エリア"),margin=2)
t3<-prop.table(xtabs(~data$"部屋" + data$"エリア"),margin=2)
t4<-prop.table(xtabs(~data$"設備.アメニティ" + data$"エリア")[(2:6)],margin=2)
t5<-prop.table(xtabs(~data$"風呂" + data$"エリア")[(2:6)],margin=2)
t6<-prop.table(xtabs(~data$"食事" + data$"エリア")[(2:6)],margin=2)

quartz(width=12,height=8,type="pdf",file="plot.pdf")
par(family="serif",mfrow=c(3,3))
barplot(t1,col=heat.colors(nrow(t1)),legend.text=rownames(t1),main="サービス")
barplot(t2,col=heat.colors(nrow(t2)),legend.text=rownames(t2),main="立地")
barplot(t3,col=heat.colors(nrow(t3)),legend.text=rownames(t3),main="部屋")
barplot(t4,col=heat.colors(nrow(t4)),legend.text=rownames(t4),main="設備・アメニティ")
barplot(t5,col=heat.colors(nrow(t5)),legend.text=rownames(t5),main="風呂")
barplot(t6,col=heat.colors(nrow(t6)),legend.text=rownames(t6),main="食事")
barplot(t0,col=heat.colors(nrow(t0)),legend.text=rownames(t0),main="総合")
dev.off()

# 検定
t0<-xtabs(~data$"総合" + data$"エリア")
t1<-xtabs(~data$"サービス" + data$"エリア")
t2<-xtabs(~data$"立地" + data$"エリア")
t3<-xtabs(~data$"部屋" + data$"エリア")
t4<-xtabs(~data$"設備.アメニティ" + data$"エリア")[(2:6),]
t5<-xtabs(~data$"風呂" + data$"エリア")[(2:6),]
t6<-xtabs(~data$"食事" + data$"エリア")[(2:6),]


```

```

chisq_p<-matrix(nrow=4,ncol=3)
rownames(chisq_p)<-c("総合","風呂","食事","サービス")
colnames(chisq_p)<-c("外房-西伊豆","西伊豆-南房総","南房総-外房")
chisq_p["総合","外房-西伊豆"]<-1

# 総合 (t0)
t12<-t0[,c("01-外房","02-西伊豆")]; t12.chisq<-chisq.test(t12); t12.prop<-prop.test(t12)
t23<-t0[,c("02-西伊豆","03-南房総")]; t23.chisq<-chisq.test(t23); t23.prop<-prop.test(t23)
t13<-t0[,c("01-外房","03-南房総")]; t13.chisq<-chisq.test(t13); t13.prop<-prop.test(t13)
t12; t12.chisq; t12.prop; chisq_p["総合","外房-西伊豆"]<-t12.chisq$p.value
t23; t23.chisq; t23.prop; chisq_p["総合","西伊豆-南房総"]<-t23.chisq$p.value
t13; t13.chisq; t13.prop; chisq_p["総合","南房総-外房"]<-t13.chisq$p.value

# 風呂 (t5)
t12<-t5[,c("01-外房","02-西伊豆")]; t12.chisq<-chisq.test(t12); t12.prop<-prop.test(t12)
t23<-t5[,c("02-西伊豆","03-南房総")]; t23.chisq<-chisq.test(t23); t23.prop<-prop.test(t23)
t13<-t5[,c("01-外房","03-南房総")]; t13.chisq<-chisq.test(t13); t13.prop<-prop.test(t13)
t12; t12.chisq; t12.prop; chisq_p["風呂","外房-西伊豆"]<-t12.chisq$p.value
t23; t23.chisq; t23.prop; chisq_p["風呂","西伊豆-南房総"]<-t23.chisq$p.value
t13; t13.chisq; t13.prop; chisq_p["風呂","南房総-外房"]<-t13.chisq$p.value

# 食事 (t6)
t12<-t6[,c("01-外房","02-西伊豆")]; t12.chisq<-chisq.test(t12); t12.prop<-prop.test(t12)
t23<-t6[,c("02-西伊豆","03-南房総")]; t23.chisq<-chisq.test(t23); t23.prop<-prop.test(t23)
t13<-t6[,c("01-外房","03-南房総")]; t13.chisq<-chisq.test(t13); t13.prop<-prop.test(t13)
t12; t12.chisq; t12.prop; chisq_p["食事","外房-西伊豆"]<-t12.chisq$p.value
t23; t23.chisq; t23.prop; chisq_p["食事","西伊豆-南房総"]<-t23.chisq$p.value
t13; t13.chisq; t13.prop; chisq_p["食事","南房総-外房"]<-t13.chisq$p.value

# サービス (t1)
t12<-t1[,c("01-外房","02-西伊豆")]; t12.chisq<-chisq.test(t12); t12.prop<-prop.test(t12)
t23<-t1[,c("02-西伊豆","03-南房総")]; t23.chisq<-chisq.test(t23); t23.prop<-prop.test(t23)
t13<-t1[,c("01-外房","03-南房総")]; t13.chisq<-chisq.test(t13); t13.prop<-prop.test(t13)
t12; t12.chisq; t12.prop; chisq_p["サービス","外房-西伊豆"]<-t12.chisq$p.value
t23; t23.chisq; t23.prop; chisq_p["サービス","西伊豆-南房総"]<-t23.chisq$p.value
t13; t13.chisq; t13.prop; chisq_p["サービス","南房総-外房"]<-t13.chisq$p.value
print(chisq_p)

```

分析テーマの例

- 食事やサービスが良いのは西伊豆,
風呂が悪いのは南房総と外房
 - どんな改善が提案できるか?
- 評価点は 5点～4点 に集中
→そもそも評価点は信頼できるか?

(評価項目別)

	01-外房	02-西伊豆	03-南房総	全体
サービス	4.25	4.41	4.35	4.35
立地	4.31	4.29	4.34	4.31
部屋	4.11	4.23	4.06	4.16
設備・アメニティ	3.92	4.03	4.06	4.00
風呂	3.99	4.33	3.96	4.15
食事	4.33	4.48	4.29	4.39
総合	4.31	4.46	4.34	4.39

テキストデータ分析の手順

1. テキストデータを収集する



2. テキストデータから語を抽出する

- 形態素解析による単語分割
- 品詞情報を使った不要語削除 (動詞は基本形へ変換)



3. 各種分析を行う

- 頻出語を確認する
 - 単語ランキング
- 語と語の結びつきを確認する
 - 共起ランキング, 共起ネットワーク
 - 係り受け解析
- グループを見つける
 - クラスタリング

形態素解析とは

- 文を単語に区切り,品詞を推定する
 - 日本語など分かち書きのない言語で必要
- 3つの処理
 - 単語の分かち書き処理 (tokenization)
 - 活用語処理 (stemming, lemmatization)
 - 品詞推定 (Part Of Speech tagging)
- 代表的なツール
 - MeCab (<http://taku910.github.io/mecab/>)
 - JUMAN (<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>)

形態素解析による単語分割

- テキストデータ

子供の運動会でとても良い映像が撮影できた



- 形態素解析による単語分割

子供/の/運動会/で/とても/良い/映像/が/撮影/でき/た

名詞 助詞 名詞 助詞 副詞 形容詞 名詞 助詞 名詞 動詞 助動詞



- 不要語を除去 (動詞は基本形へ変換)

子供 運動会 良い 映像 撮影 できる

名詞 名詞 形容詞 名詞 名詞 動詞 (内容語のみ残した例)

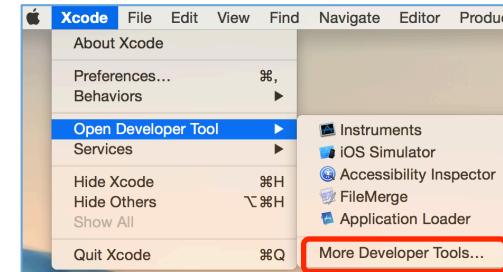
事前準備 — Mac で演習を行うには

- 形態素解析ツール MeCab をインストールするため,以下の手順で Cコンパイラをインストールしてください
 - ① Apple Store から Xcodeをインストール
 - ② Xcode -> Preferences -> Downloads -> Command Line Tools からインストール

① Xcode のインストール



② Command Line Tools のインストール



Description	Release Date
+ Command Line Tools (OS X 10.10) for Xcode 6.3.2	May 18, 2015
+ Command Line Tools (OS X 10.10) for Xcode 6.3.1	Apr 20, 2015
+ Command Line Tools (OS X 10.10) for Xcode 6.3	Apr 8, 2015
+ Command Line Tools (OS X 10.9) for Xcode - Xcode 6.2	Mar 7, 2015
+ Command Line Tools (OS X 10.10) for Xcode - Xcode 6.2	Mar 6, 2015
+ Command Line Tools (OS X 10.10) for Xcode - Xcode 6.1.1	Dec 2, 2014
+ Command Line Tools (OS X 10.9) for Xcode - Xcode 6.1	Dec 2, 2014

注: 使用中の OSX
や Xcode のバージョンが一致する
ものを選択する

演習 — MeCabのインストール

- ダウンロードサイト
 - <http://taku910.github.io/mecab/#download>

Mac の場合

Windows の場合

ダウンロード

- MeCab はフリーソフトウェアです。GPL(the GNU General Public License), LGPL とができます。 詳細は COPYING, GPL, LGPL, BSD各ファイルを参照して下さい。
- MeCab 本体

Source

- [mecab-0.996.tar.gz:ダウンロード](#)
- 辞書は含まれていません。動作には別途辞書が必要です。

Binary package for MS-Windows

- [mecab-0.996.exe:ダウンロード](#)
- Windows 版には コンパイル済みの IPA 辞書が含まれています

演習 — MeCabのインストール

- Mac の場合 (1/3)
 - MeCab のコンパイルとインストール

```
$ cd ~/Downloads  
$ tar zxvf mecab-0.996.tar.gz  
$ cd mecab-0.996  
$ LIBS=-liconv ./configure --with-charset=utf8  
$ make  
$ make check  
$ sudo make install
```

注1: 前ページのダウンロード先を ~/Downloads にしている場合

注2: 「\$」 はコマンドプロンプトです, 入力しないでください

演習 — MeCabのインストール

- Mac の場合 (2/3)
 - 辞書(NAIST-jdic)のインストール

```
$ cd ~/Downloads  
$ wget http://osdn.jp/projects/naist-jdic/downloads/53500/mecab-  
naist-jdic-0.6.3b-20111013.tar.gz  
$ tar zxvf mecab-naist-jdic-0.6.3b-20111013.tar.gz  
$ cd mecab-naist-jdic-0.6.3b-20111013  
$ LIBS=-liconv ./configure --with-charset=utf8  
$ make  
$ sudo make install
```

↑
画面の都合上折り返し

注1: ダウンロード先を ~/Downloads にしている場合

注2: 「\$」はコマンドプロンプトです, 入力しないでください

演習 — MeCabのインストール

- Mac の場合 (3/3)
 - 設定ファイルの書き換え

```
$ sudo vi /usr/local/etc/mecabrc
```

以下の行を見つけて編集する

(編集前) ;
dicdir = /usr/local/lib/mecab/dic/ipadic

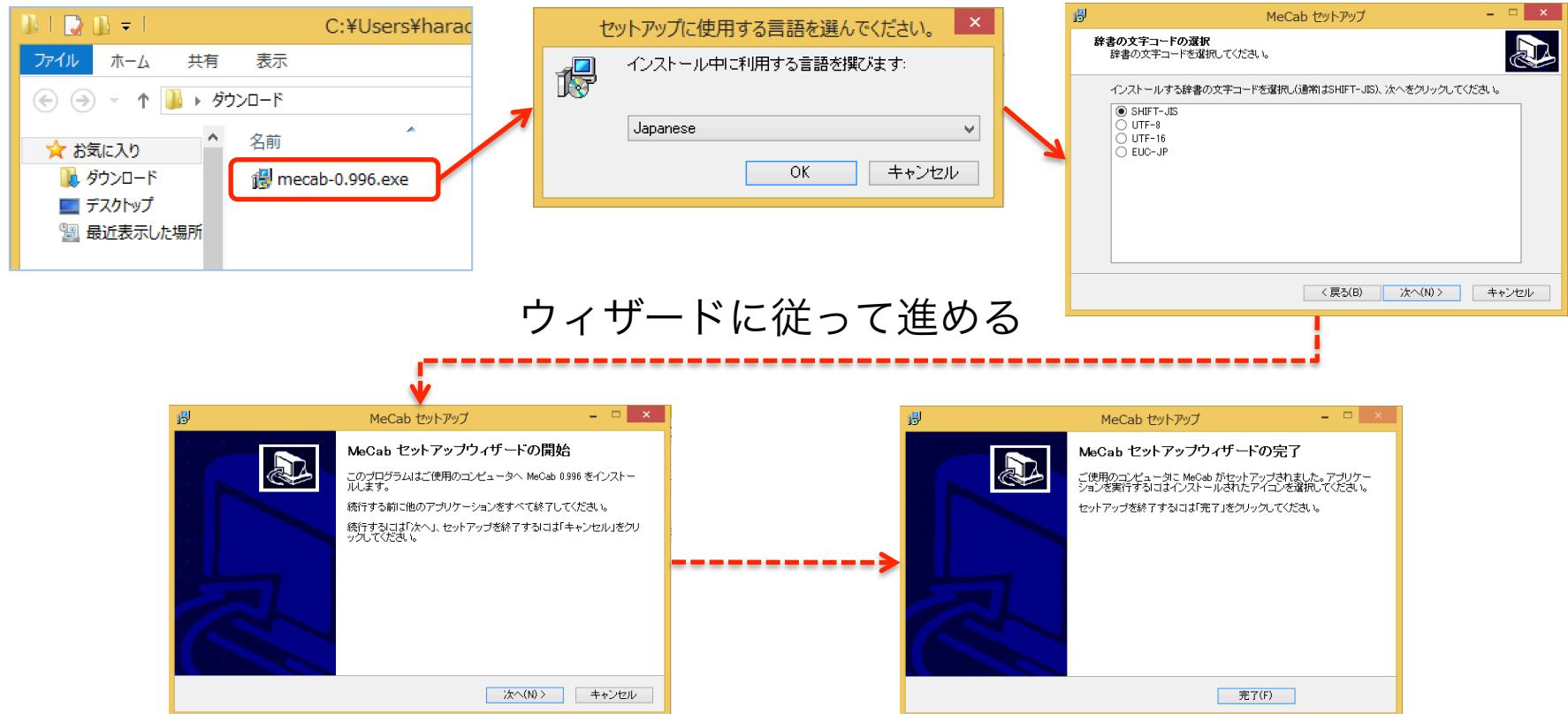


(編集後) ;
#dicdir = /usr/local/lib/mecab/dic/ipadic
dicdir = /usr/local/lib/mecab/dic/naist-jdic

保存して vi コマンドを修了する (:wq コマンド)

演習 — MeCabのインストール

- Windows の場合 (1/5)
 - MeCab のインストール



演習 — MeCabのインストール

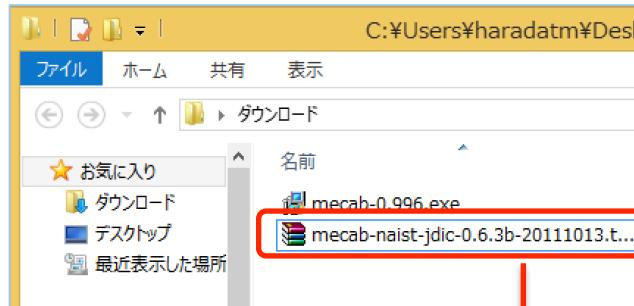
- Windows の場合 (2/5)
 - 辞書(NAIST-jdic)のインストール



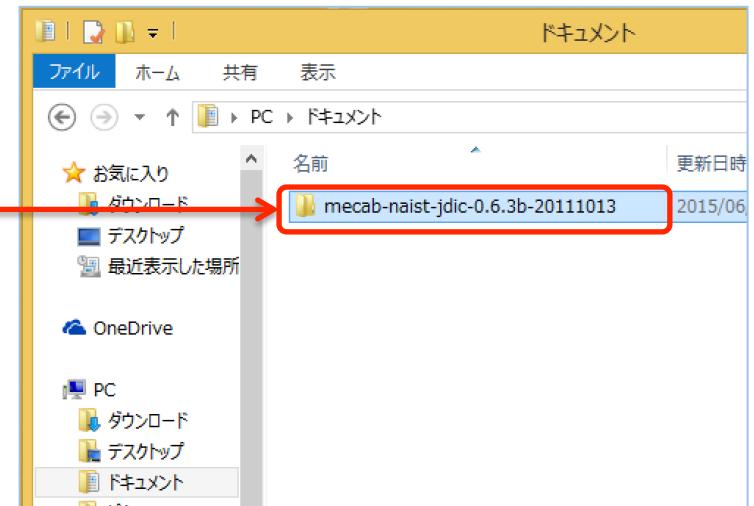
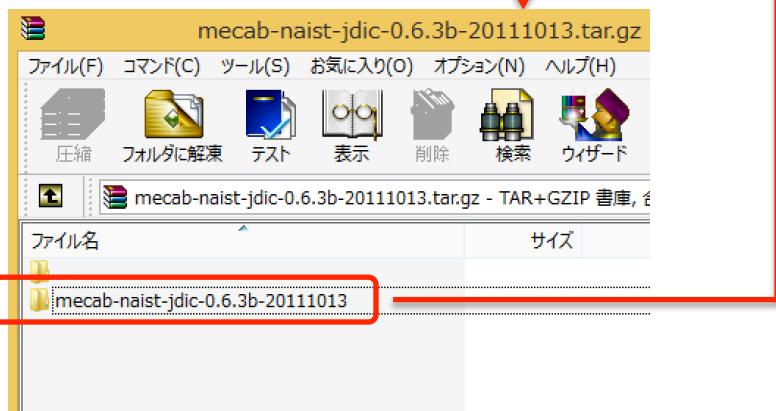
これをダウンロードする

演習 — MeCabのインストール

- Windows の場合 (3/5)
 - 辞書(NAIST-jdic)のインストール



① 任意のツールで解凍する



② ドキュメントフォルダへ移動する

演習 — MeCabのインストール

- Windows の場合 (4/5)
 - 辞書(NAIST-jdic)のインストール

The screenshot shows a Windows Command Prompt window titled "管理者: コマンドプロンプト". The text in the window is as follows:

```
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Windows\system32>cd \Users\ユーザー名\Documents

C:\Users\ユーザー名\Documents> "C:\Program Files (x86)\MeCab\bin\mecab-
-dict-index.exe" -d .\mecab-naist-jdic-0.6.3b-20111013 -o .\mecab-naist-
jdic-0.6.3b-20111013 -f EUC -c SJIS

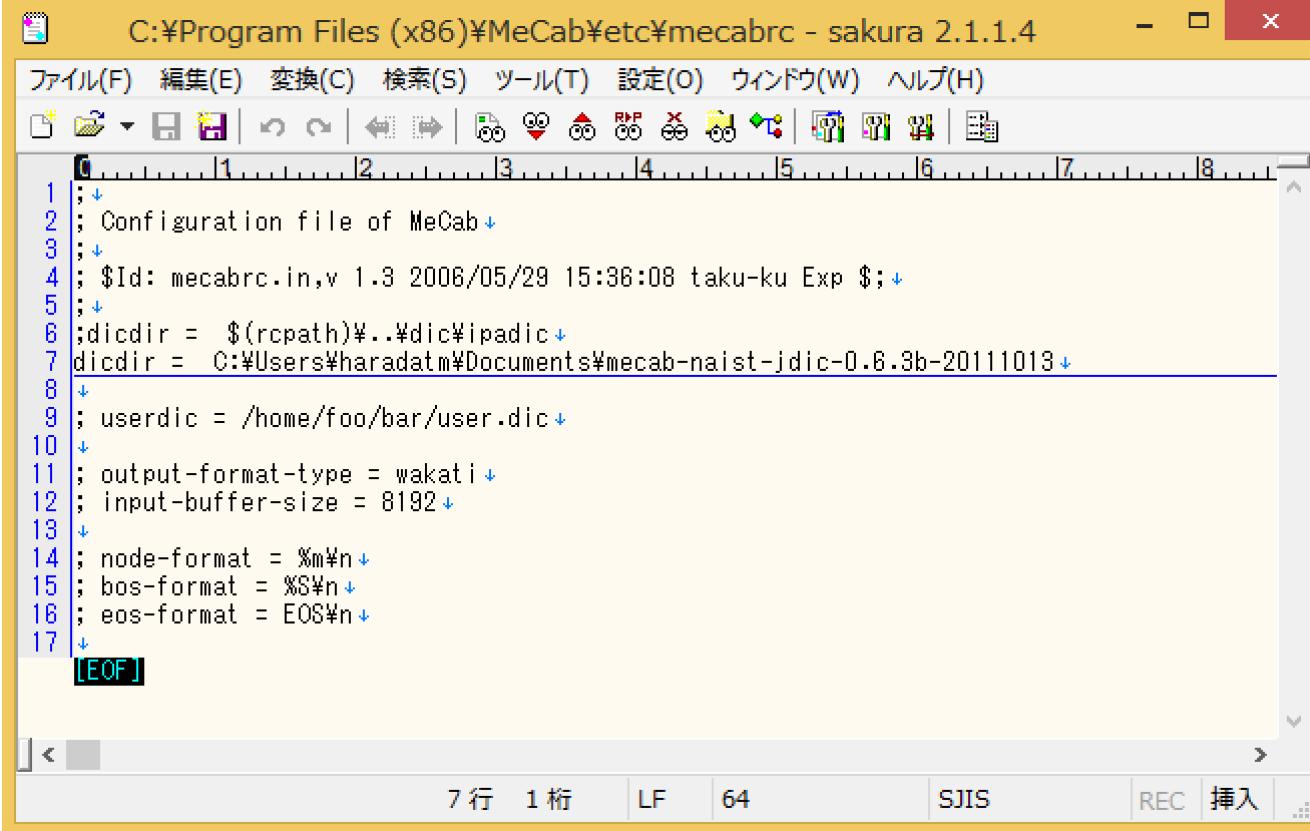
:
done!
```

Annotations in Japanese:

- An arrow points from the text "管理者モードで開く" to the title bar of the window.
- An arrow points from the text "画面の都合上折り返し" to the scroll bar on the right side of the window.
- A callout box at the bottom contains two notes:
 - 注1: ファイルの解凍先を C:\Users\ユーザー名\Documents にしている場合
 - 注2: 「C:***>」はコマンドプロンプトです、入力しないでください

演習 — MeCabのインストール

- Windows の場合 (5/5)
 - 設定ファイルの書き換え



The screenshot shows a Windows Notepad window with the title bar "C:\Program Files (x86)\MeCab\etc\mecabrc - sakura 2.1.1.4". The menu bar includes ファイル(F), 編集(E), 変換(C), 検索(S), ツール(T), 設定(O), ウィンドウ(W), ヘルプ(H). The toolbar contains various icons for file operations. The main text area displays the contents of the 'mecabrc' file:

```
1;+
2; Configuration file of MeCab +
3;+
4; $Id: mecabrc.in,v 1.3 2006/05/29 15:36:08 taku-ku Exp $;+
5;+
6;dicdir = $(rcpath)..\$dic\$ipadic +
7;dicdir = C:\Users\haradatm\Documents\mecab-naist-jdic-0.6.3b-20111013 +
8;+
9; userdic = /home/foo/bar/user.dic +
10;+
11; output-format-type = wakati +
12; input-buffer-size = 8192 +
13;+
14; node-format = %m\n +
15; bos-format = %S\n +
16; eos-format = EOS\n +
17;+
[EOF]
```

The status bar at the bottom shows "7行 1行 | LF | 64 | SJIS | REC | 握入".

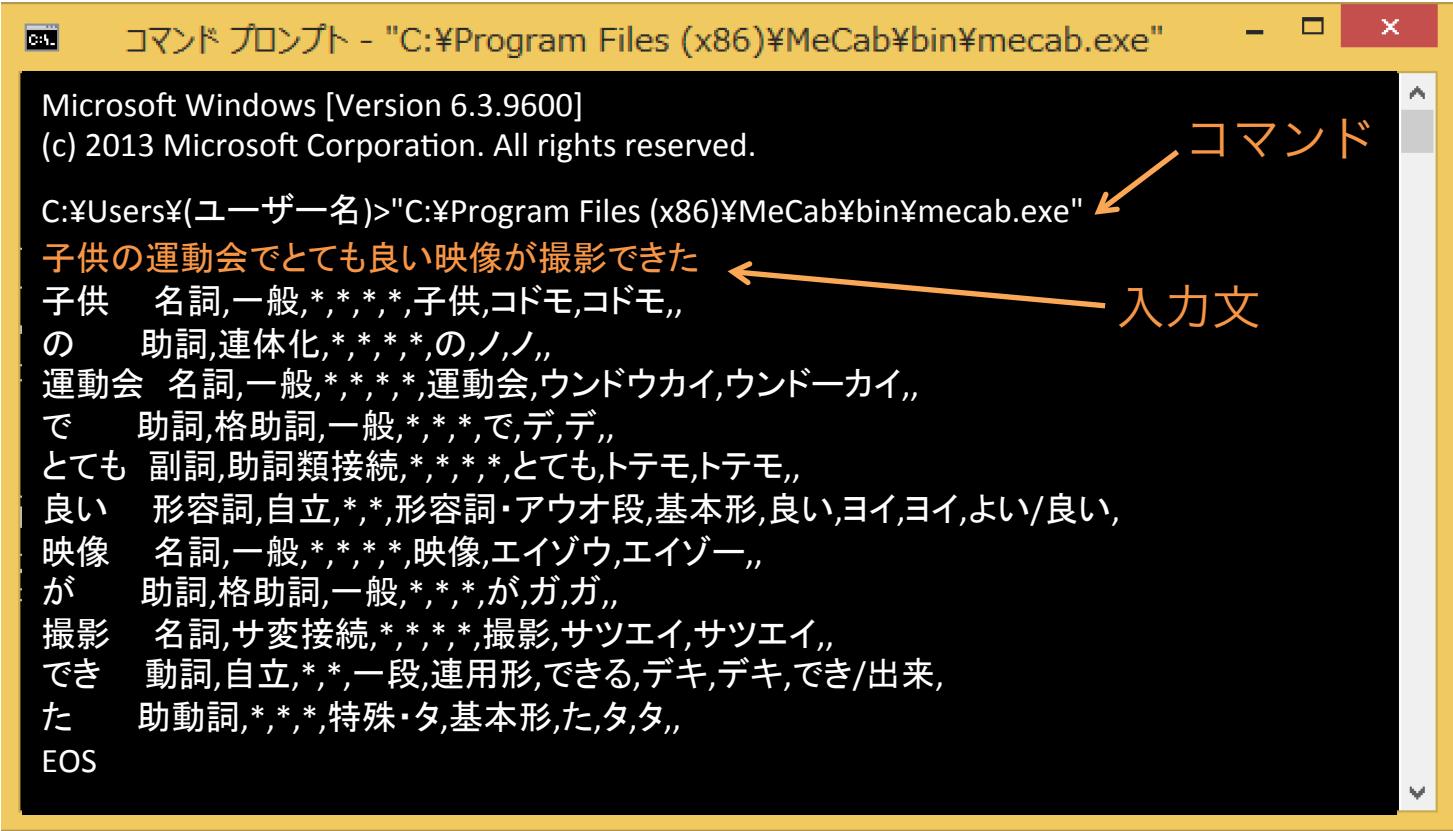
演習 — MeCabの動作確認

- Mac の場合
 - 形態素解析の実行 (以下のような出力がされればOKです)

```
$ cd ~/Documents  
$ mecab ← コマンド  
子供の運動会でとても良い映像が撮影できた ← 入力文  
子供 名詞,一般,*,*,*,*,子供,コドモ,コドモ,,  
の 助詞,連体化,*,*,*,*,の,ノ,ノ,,  
運動会 名詞,一般,*,*,*,*,運動会,ウンドウカイ,ウンドーカイ,,  
で 助詞,格助詞,一般,*,*,*,で,デ,デ,,  
とても 副詞,助詞類接続,*,*,*,*,とても,トテモ,トテモ,,  
良い 形容詞,自立,*,*,形容詞・アウオ段,基本形,良い,ヨイ,ヨイ,よい/良い,  
映像 名詞,一般,*,*,*,映像,エイゾウ,エイゾー,,  
が 助詞,格助詞,一般,*,*,*,が,ガ,ガ,,  
撮影 名詞,サ変接続,*,*,*,撮影,サツエイ,サツエイ,,  
でき 動詞,自立,*,*,一段,連用形,できる,デキ,デキ,でき/出来,  
た 助動詞,*,*,*,特殊・タ,基本形,た,タ,タ,,  
EOS
```

演習 — MeCabの動作確認

- Windows の場合
 - 形態素解析の実行 (以下のような出力がされればOKです)



Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\¥Users¥(ユーザー名)>"C:\¥Program Files (x86)\¥MeCab\¥bin\¥mecab.exe"
子供の運動会でとても良い映像が撮影できた
子供 名詞,一般,*,*,*,*子供,コドモ,コドモ,,
の 助詞,連体化,*,*,*,*の,ノ,ノ,,
運動会 名詞,一般,*,*,*,*運動会,ウンドウカイ,ウンドーカイ,,
で 助詞,格助詞,一般,*,*,*で,デ,デ,,
とても 副詞,助詞類接続,*,*,*,*とても,トテモ,トテモ,,
良い 形容詞,自立,*,*,*形容詞・アウオ段,基本形,良い,ヨイ,ヨイ,よい/良い,
映像 名詞,一般,*,*,*,*映像,エイゾウ,エイゾー,,
が 助詞,格助詞,一般,*,*,*が,ガ,ガ,,
撮影 名詞,サ変接続,*,*,*,*撮影,サツエイ,サツエイ,,
でき 動詞,自立,*,*,*一段,連用形,できる,デキ,デキ,でき/出来,
た 助動詞,*,*,*特殊・タ,基本形,た,タ,タ,,
EOS

コマンド

入力文

The screenshot shows a Windows Command Prompt window with the title "コマンド プロンプト - 'C:\¥Program Files (x86)\¥MeCab\¥bin\¥mecab.exe'". The window displays the output of the MeCab morphological analyzer. An orange arrow points from the word 'コマンド' to the command line at the top. Another orange arrow points from the word '入力文' to the input sentence '子供の運動会でとても良い映像が撮影できた'.

テキストマイニングの実習2

— 形態素解析を利用した集計と分析 —

2015/7/9

ビジネス科学研究科
経営システム科学専攻

演習 — 頻出語を確認する

- 形態素解析器 MeCab の出力例

```
$ mecab
```

```
子供の運動会でとても良い映像が撮影できた
子供 名詞,一般,*,*,*,*,子供,コドモ,コドモ,,
の 助詞,連体化,*,*,*,*,の,ノ,ノ,,
運動会 名詞,一般,*,*,*,*,運動会,ウンドウカイ,ウンドーカイ,,
で 助詞,格助詞,一般,*,*,*,で,デ,デ,,,
とても副詞,助詞類接続,*,*,*,*,とても,トテモ,トテモ,,
良い 形容詞,自立,*,*,形容詞・アウオ段,基本形,良い,ヨイ,ヨイ,よい/良い,
映像 名詞,一般,*,*,*,*,映像,エイゾウ,エイゾー,,
が 助詞,格助詞,一般,*,*,*,が,ガ,ガ,,,
撮影 名詞,サ変接続,*,*,*,*,撮影,サツエイ,サツエイ,,
でき 動詞,自立,*,*,一段,連用形,できる,デキ,デキ,でき/出来,
た 助動詞,*,*,*,特殊・タ,基本形,た,タ,タ,,,
EOS
```

- 単語 <tab> 品詞1,品詞2,品詞3,品詞4,活用型,活用系,基本形,読み,発音
- EOS: End of Sentence
- 出力フォーマットの変更が可能

演習 — 頻出語を確認する

- MeCab を実行する: Mac の場合

```
$ cd (作業ディレクトリ)
```

```
$ mecab -b 819200 -F"%f[6]\t%f[0]\t%f[1]\n" -U"%m\t未知語\n" -E"EOS\tEOS\n"  
01-外房_テキスト.txt > 01-外房_単語.txt
```

```
$ mecab -b 819200 -F"%f[6]\t%f[0]\t%f[1]\n" -U"%m\t未知語\n" -E"EOS\tEOS\n"  
02-西伊豆_テキスト.txt > 02-西伊豆_単語.txt
```

```
$ mecab -b 819200 -F"%f[6]\t%f[0]\t%f[1]\n" -U"%m\t未知語\n" -E"EOS\tEOS\n"  
03-南房総_テキスト.txt > 03-南房総_単語.txt
```

画面の都合上折り返し

注: バッファ・サイズには “-b 819200” を指定してください

【フォーマットを変更するパラメータ】

-F	… 項目の並び	例) -F“%f[6]\t%f[0]\t%f[1]\n”	→ 子供<tab>名詞<tab>一般
-U	… 未知語の表示法	例) -U“%m\t未知語\n”	→ 未知語<tab><tab><tab>
-E	… EOSの表示法	例) -E“EOS\tEOS\n”	→ EOS<tab>EOS

演習 — 頻出語を確認する

- MeCab を実行する: Windows の場合

```
コマンドプロンプト  
C:\$Users\$(ユーザー名)> "C:\Program Files (x86)\MeCab\bin\mecab.exe"  
-b 819200 -F"%f[6]\t%f[0]\t%f[1]\n" -U"%m\t未知語\n" -E"EOS\tEOS\n"  
01-外房_テキスト.txt > 01-外房_単語.txt  
  
C:\$Users\$(ユーザー名)> "C:\Program Files (x86)\MeCab\bin\mecab.exe"  
-b 819200 -F"%f[6]\t%f[0]\t%f[1]\n" -U"%m\t未知語\n" -E"EOS\tEOS\n"  
02-西伊豆_テキスト.txt > 02-西伊豆_単語.txt  
  
C:\$Users\$(ユーザー名)> "C:\Program Files (x86)\MeCab\bin\mecab.exe"  
-b 819200 -F"%f[6]\t%f[0]\t%f[1]\n" -U"%m\t未知語\n" -E"EOS\tEOS\n"  
03-南房総_テキスト.txt > 03-南房総_単語.txt
```

画面の都合上折り返し

注1: バッファ・サイズには “-b 819200” を指定してください

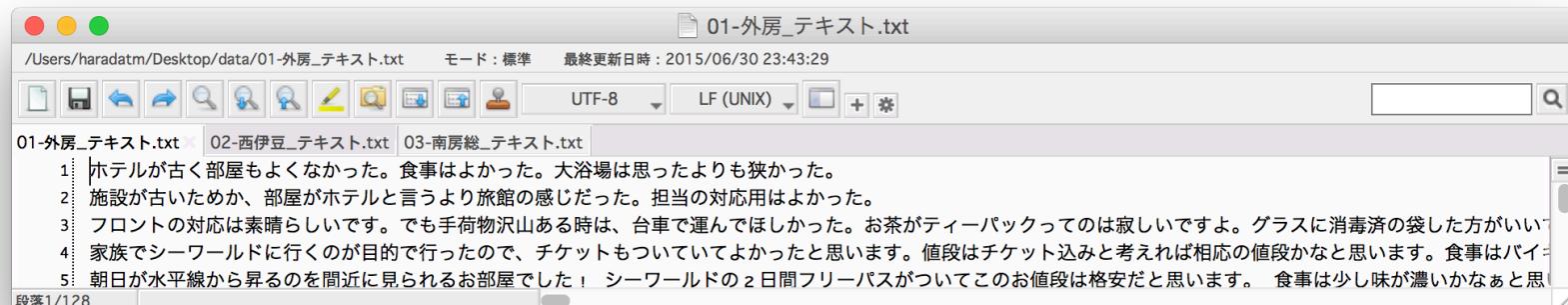
注2: フォーマット変更のパラメータは前頁(Macの場合)と同様です

演習 — 頻出語を確認する

- 風呂の評価が1~2点の口コミを抜き出す

- シート名 [データ (6317)] を開く
- フィルター機能で,A列を [01-外房] のみに絞り込む
- フィルターで機能で,J列を [1,2] のみに絞り込む
- テキストエディタで新しいファイル(ウィンドウ)を開く
- D列の2行目以降を選択し,テキストエディタに貼り付ける
- ファイル名 「01-外房_テキスト.txt」 で保存する

※ 一旦, A列のチェックを外した後, 同様に [02-西伊豆] [03-南房総]についても②~⑥を繰り返す



演習 — 頻出語を確認する

- 単語分割の結果を EXCEL に貼り付ける
 - ① EXCEL の新しいシートの1行目に列名を入力する
→ 列名は左から順に [エリア] [形態素] [品詞1] [品詞2] [単語]
 - ② ファイル「01-外房_単語.txt」をテキストエディタで開く
 - ③ テキストデータの2行目以降を選択し,EXCEL に貼り付ける
→ 貼り付け先: ①のシート上の B列の2行目
 - ④ A列2行目に [01-外房] と入力し,行末までコピーする
 - ⑤ E列2行目に式 [=B2&“(&C2&“)”] を入力し,行末までコピー
→ ④および⑤は, 貼り付けたデータが存在する最終行までコピー
 - ⑥ シート名を [01-外房] に変更する

	A1				
	A	B	C	D	E
1	エリア	形態素	品詞1	品詞2	単語
2	01-外房	ホテル	名詞	一般	ホテル(名詞)
3	01-外房	が	助詞	格助詞	が(助詞)
4	01-外房	古く	副詞	助詞類接続	古く(副詞)

※ 同様に [02-西伊豆] [03-南房総] についても ②～⑨ を繰り返す

演習 — 頻出語を確認する

• 単語の出現頻度を集計する

- ① シート [01-外房] を開き, A~E 列を選択してピボットを作成
→ 新しいシートに作成し, シート名を [集計] に変更

※ 同様に [02-西伊豆] [03-南房総] についても①を行う
→ ピボットは, 比較ができるように同じシート [集計] 上に作成する

A	B	C	D	E	F	G	H
1 エリア	01-外房	エリア	02-西伊豆	エリア	03-南房総		
2 品詞1	(複数の項目)	品詞1	(複数の項目)	品詞1	(複数の項目)	品詞1	(複数の項目)
3 品詞2	(複数の項目)	品詞2	(複数の項目)	品詞2	(複数の項目)	品詞2	(複数の項目)
4							
5 データの個数: 形態素		データの個数: 形態素		データの個数: 形態素		データの個数: 形態素	
6 行ラベル	計	行ラベル	計	行ラベル	計	行ラベル	計
7 部屋 (名詞)	121	部屋 (名詞)	92	部屋 (名詞)	99	部屋 (名詞)	99
8 風呂 (名詞)	99	風呂 (名詞)	84	風呂 (名詞)	77	風呂 (名詞)	77
9 良い (形容詞)	82	良い (形容詞)	75	良い (形容詞)	54	良い (形容詞)	54
10 食事 (名詞)	74	ない (形容詞)	67	食事 (名詞)	3	宿泊	3
11 ない (形容詞)	67	食事 (名詞)	54	料理	8	良い	8
12 ホテル (名詞)	54	残念 (名詞)	50	利用	7	利用	7
13 宿泊 (名詞)	50	温泉 (名詞)	50	料理	5	料理	5
14 人 (名詞)	50	宿泊 (名詞)	42	ホテル	9	ホテル	9
15 残念 (名詞)	42	宿 (名詞)	40	満足	8	満足	8
16 朝食 (名詞)	40	料理 (名詞)	38	ない	6	ない	6
17 利用 (名詞)	38	露天風呂 (名詞)	36	夕食	3	夕食	3
18 海 (名詞)	36	利用 (名詞)		残念	2	残念	2

- ② フィルター機能を使い, 品詞情報で不要語削除する

- 品詞1を [形容詞, 名詞, 未知語] のみに絞る
- 品詞2から [数, 非自立] のチェックを外す

演習 — 語と語の結びつきを確認する

• 品詞情報をを使った不要語削除する

- ① シート [01-外房] を開く
- ② フィルター機能で,C列を [形容詞,名詞,未知語] のみに絞る
- ③ フィルター機能で,D列から [数,非自立] のチェックを外す
- ④ データのあるセル全体を選択し,新しいシートに貼り付ける
- ⑤ シート名を [01-外房] に変更する

A	B	C	D	E
1 エリア	形態素	品詞1	品詞2	単語
2 01-外房	ホテル	名詞	一般	ホテル (名詞)
5 01-外房	部屋	名詞	一般	部屋 (名詞)
7 01-外房	よい	形容詞	自立	よい (形容詞)
11 01-外房	食事	名詞	サ変接続	食事 (名詞)
13 01-外房	よい	形容詞	自立	よい (形容詞)
17 01-外房	浴場	名詞	一般	浴場 (名詞)
23 01-外房	狭い	形容詞	自立	狭い (形容詞)
26 01-外房	EOS	EOS		EOS (EOS)
27 01-外房	施設	名詞	サ変接続	施設 (名詞)



A	B	C	D	E
1 エリア	形態素	品詞1	品詞2	単語
2 01-外房	ホテル	名詞	一般	ホテル (名詞)
3 01-外房	部屋	名詞	一般	部屋 (名詞)
4 01-外房	よい	形容詞	自立	よい (形容詞)
5 01-外房	食事	名詞	サ変接続	食事 (名詞)
6 01-外房	よい	形容詞	自立	よい (形容詞)
7 01-外房	浴場	名詞	一般	浴場 (名詞)
8 01-外房	狭い	形容詞	自立	狭い (形容詞)
9 01-外房	EOS	EOS		EOS (EOS)
10 01-外房	施設	名詞	サ変接続	施設 (名詞)

演習 — 語と語の結びつきを確認する

- 近くに出現(=共起)する単語の組みを作る
 - ⑥ 新しいシートの1行目にあるC列より右の列名を変更する
[品詞1(前)] [品詞2(前)] [単語(前)] [品詞1(後)] [品詞2(後)] [単語(後)]
[単語(前)-単語(後)]
 - ⑦ 新しいシートの C～E列の3行目以降を行末まで選択し、右側にある同じシートの E列1行目に貼り付ける
 - ⑧ I列2行目に式 [=E2&"-"&H2] を入力し,行末までコピーする

※ 同様に [02-西伊豆] [03-南房総] についても②～⑧を繰り返す

	A	B	C	D	E	F	G	H	I	J
1	エリア	形態素	品詞1(前)	品詞2(前)	単語(前)	品詞1(後)	品詞2(後)	単語(後)	単語(前)-単語(後)	
2	01-外房	ホテル	名詞	一般	ホテル (名詞)	名詞	一般	部屋 (名詞)	ホテル (名詞)-部屋 (名詞)	
3	01-外房	部屋	名詞	一般	部屋 (名詞)	形容詞	自立	よい (形容詞)	部屋 (名詞)-よい (形容詞)	
4	01-外房	よい	形容詞	自立	よい (形容詞)	名詞	サ変接続	食事 (名詞)	よい (形容詞)-食事 (名詞)	
5	01-外房	食事	名詞	サ変接続	食事 (名詞)	形容詞	自立	よい (形容詞)	食事 (名詞)-よい (形容詞)	
6	01-外房	よい	形容詞	自立	よい (形容詞)	名詞	一般	浴場 (名詞)	よい (形容詞)-浴場 (名詞)	
7	01-外房	浴場	名詞	一般	浴場 (名詞)	形容詞	自立	狭い (形容詞)	浴場 (名詞)-狭い (形容詞)	
8	01-外房	狭い	形容詞	自立	狭い (形容詞)	EOS		EOS (EOS)	狭い (形容詞)-EOS (EOS)	
9	01-外房	EOS	EOS		EOS (EOS)	名詞	サ変接続	施設 (名詞)	EOS (EOS)-施設 (名詞)	
10	01-外房	施設	名詞	サ変接続	施設 (名詞)	形容詞	自立	古い (形容詞)	施設 (名詞)-古い (形容詞)	

演習 — 語と語の結びつきを確認する

・ 共起の出現頻度を集計する

- シート [01-外房 (2)] を開き,A~I列を選択し,ピボットを作成
→ 新しいシートに作成し, シート名を [共起] に変更

※ 同様に [02-西伊豆] [03-南房総] についても①を行う
→ ピボットは, 比較ができるように同じシート [共起] 上に作成する

	A	B	C	D	E	F	G	H
1	エリア	01-外房	エリア	02-西伊豆	エリア	03-南房総		
2	品詞1(前)	(複数の項目)	品詞1(前)	(複数の項目)	品詞1(前)	(複数の項目)		
3	品詞2(前)	(すべて)	品詞2(前)	(すべて)	品詞2(前)	(すべて)		
4	品詞1(後)	形容詞	品詞1(後)	形容詞	品詞1(後)	形容詞		
5	品詞2(後)	(すべて)	品詞2(後)	(すべて)	品詞2(後)	(すべて)		
6								
7	データの個数: 形態素		データの個数: 形態素		データの個数: 形態素			
8	行ラベル		行ラベル		行ラベル			
9	風呂 (名詞)-ない (形容詞)	2	風呂 (名詞)-狭い (形容詞)	4	風呂 (名詞)-狭い (形容詞)	4		
10	風呂 (名詞)-狭い (形容詞)	2	風呂 (名詞)-良い (形容詞)	3	露天風呂 (名詞)-ない (形容詞)	4		
11	風呂 (名詞)-広い (形容詞)	2	露天風呂 (名詞)-広い (形容詞)	3	風呂 (名詞)-古い (形容詞)	3		
12	風呂 (名詞)-大きい (形容詞)	2	風呂 (名詞)-広い (形容詞)	2	風呂 (名詞)-良い (形容詞)	2		
13	風呂 (名詞)-良い (形容詞)	2	水風呂 (名詞)-無い (形容詞)	1	風呂 (名詞)-古い (形容詞)	1		
14	露天風呂 (名詞)-ない (形容詞)	2	風呂 (名詞)-大きい (形容詞)	1	風呂 (名詞)-小さい (形容詞)	1		
15	水風呂 (名詞)-ない (形容詞)	1	露天風呂 (名詞)-寒い (形容詞)	1	露天風呂 (名詞)-寒い (形容詞)	1		
16	風呂 (名詞)-すごい (形容詞)	1	露天風呂 (名詞)-寒い (形容詞)	1	風呂 (名詞)-大きい (形容詞)	1		
17	風呂 (名詞)-ぬるい (形容詞)	1	露天風呂 (名詞)-広い (形容詞)	1	風呂 (名詞)-狭い (形容詞)	1		
18	風呂 (名詞)-遠い (形容詞)	1	露天風呂 (名詞)-熱い (形容詞)	1	風呂 (名詞)-大きい (形容詞)	1		
19	風呂 (名詞)-寒い (形容詞)	1	露天風呂 (名詞)-良い (形容詞)	1	風呂 (名詞)-小さい (形容詞)	1		
20	風呂 (名詞)-高い (形容詞)	1	総計	1	露天風呂 (名詞)-寒い (形容詞)	1		
21	風呂 (名詞)-小さい (形容詞)	1						
22	風呂 (名詞)-熱い (形容詞)	1						
23	風呂 (名詞)-怖い (形容詞)	1						
24	風呂場 (名詞)-強い (形容詞)	1						
25	風呂場 (名詞)-狭い (形容詞)	1						
26	露天風呂 (名詞)-ぬるい (形容詞)	1						
27	露天風呂 (名詞)-良い (形容詞)	1						
28	総計	25						

- フィルター機能を使い, 関心のある語(名詞)と評価(形容詞)の共起を確認する

- 行ラベルで [“風呂”を含む] のチェックを外す
- 品詞1(後)を [形容詞] のみに絞る

演習 — チャレンジ課題

- ・ 時間の余った人は,ぜひチャレンジしてください
 - 抽出した単語の重要度(TF・IDF)を求める
 - 先に DF を求め, VLOOKUP で統合します
 - 単語重要度による文書ごベクトルを作る
 - ヒント: ピボットテーブルを利用します