

だれでもできる！

わかりやすい 何気ない 大きい
しづらい 文章 言葉 方法 可能 いち早い
できる 抽出 解析 しやすい 精度 アンケート
多い 書く 含む 役立つ 意見 データ分析 呼ぶ
拾う 向上 ツール 早い 表す
遠い いく ニーズ 活用 データ分析 顧客 手法 把握
おる 扱う 利用 SNS 単語 分ける
詳しい 大量 テキスト 有益 データマイニング かかる
必要 情報 自然言語処理 消費者 切り捨てる 新しい
難しい 速い 使う 伴う 幅広い
持つ 知る 用いる かける 高い
うまい とてつもない

【1回目】2022/12/14(水) 16:50-17:20 【2回目】2022/12/22(木) 12:40-13:10

担当：理工学図書館ラーニング・サポーター

自己紹介

講習会の流れ

1. 趣旨説明
2. テキストマイニングとは？
3. KH Coder のデモンストレーション
4. 質疑応答

1. 趣旨説明

2. テキストマイニングとは？

3. KH Coder のデモンストレーション

4. 質疑応答

1. テキストマイニングの基礎を知ろう!

2. フリーソフト「KH Coder」の基礎を知ろう!

難しい話は一切しません! 数式やプログラムコードは一切出てきません!

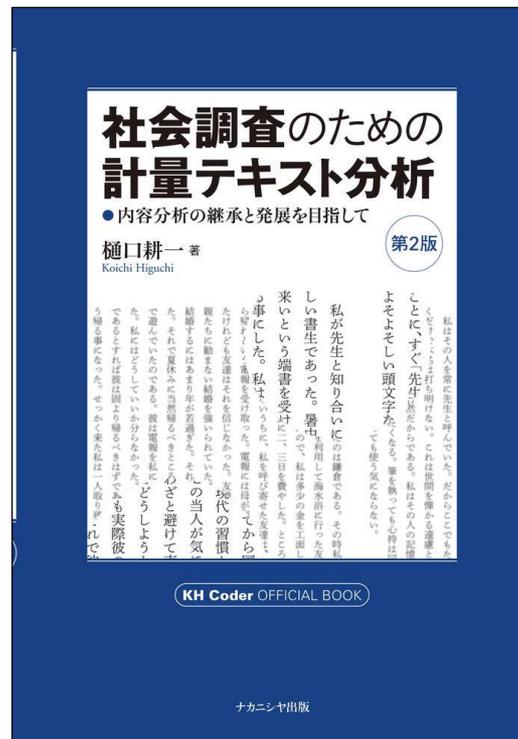
テキストマイニングソフト「KH Coder」のイメージをなんとなく、つかめるようになることが今日の目標です。

※登壇者自身も難しい話はよく理解していません。(例えば、KH Coder 内部のアルゴリズムなどはさっぱりです) あしからず……

趣旨説明

KH Coder のことを細部まで知りたい人は、下記文献を参考にしてください。

KH Coder 開発者自らが、相当詳しく(かつ難しく)解説しています。



書誌ID=2004538143

社会調査のための計量テキスト分析【第2版】

講習会の流れ

1. 趣旨説明

2. テキストマイニングとは？

3. KH Coder のデモンストレーション

4. 質疑応答

テキストマイニングとは？

あらゆる文章を**客観的に**分析する統計手法！

主観的な分析(ただのフィーリング)



こんなに気持ちを入れてラブレターを書いたのに、思いが伝わらなかった…気持ちが足りなかったのか？

客観的な分析(テキストマイニング)



「好き」等の好意を表す単語が文章全体の1%未満だから、もっとこれらの単語を増やすべきだったかも！

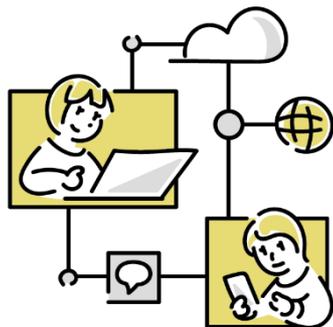
テキストマイニングで分かること

顧客アンケート



- 自由記述欄に共通する意見は？
- 問合せが多いクレームの傾向は？

SNS・Webサイト



- ××関連の投稿の頻度と傾向は？
- △△のWebサイトに多い意見は？

論文・特許



- 最近研究が盛んなキーワードは？
- この分野の特許に共通する技術は？

小説・歌詞

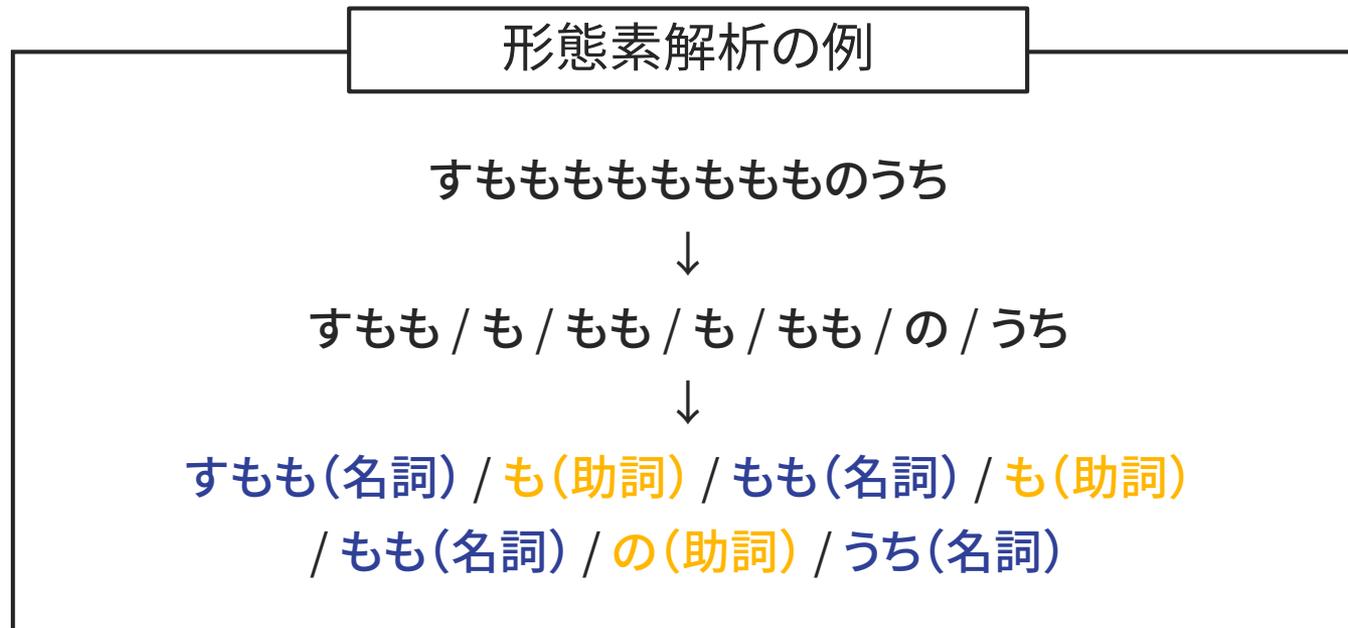


- 作家Bがよく使う表現の特徴は？
- 歌手Aと歌手Bの歌詞の違いは？

テキストマイニングによって、文章中の単語の頻度や傾向、単語間の関連性などが分かる!

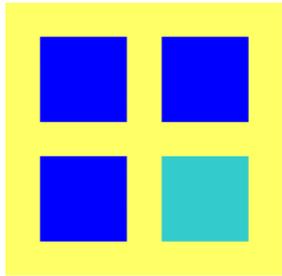
テキストマイニングの流れ

1. 「形態素解析」によって自動的に、文章を単語単位に分割し、それらの品詞を特定する。
2. 分割しすぎた単語を結合したり、ノイズとなる単語を除去する。
3. 様々な分析手法によって、文章中の特定の単語の重要度や、単語同士の関連度などを調べる。



無料テキストマイニングソフトの比較

KH Coder



- プログラミング無しで詳細な解析まで可能。
- ソフトのインストールが必要。Mac版は有料。

AIテキストマイニング by ユーザーローカル



- Webサイト上で手軽に解析できる。
- 語句の取捨選択など詳細な設定ができない。

R



- テキストマイニング以外についても、あらゆる統計解析が可能。
- プログラミングが必要。

無料テキストマイニングソフトの比較

2022年度理工学図書館
LS講習会

だれでもできる！

テキストマイニング

日時

1回目 00/00 (水) 17時 - 18時
2回目 00/00 (水) 17時 - 18時

担当

渡部 (ビジネスエンジニアリング専攻、M2)

内容

初心者大歓迎！会話や本、論文、Webサイト、歌詞、アンケートなど、文字データは身の回りにあふれています。それらのビッグデータを分析し、有益な情報を見つけ出す「テキストマイニング」の概要と、無料ソフト「KH Coder」の使い方を易しく解説します！

【問合せ先】
理工学図書館利用支援担当
tosyo-rikoh-desk@office.osaka-u.ac.jp

場所

現地：理工学図書館東館1階プレゼンテーションエリア
オンライン：Webex (リンクは申込後メールでお知らせします)

申込方法

イベント前日までに、右のQRコードからお申込みください。

QR
コード

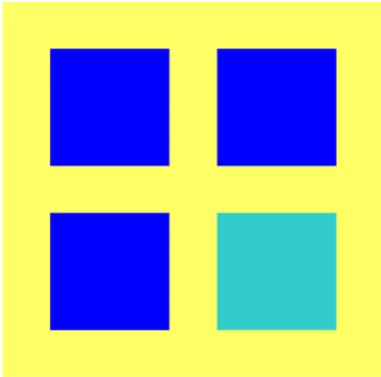
※ 事前にお申込みされていない場合でも、対面の場合は、当日の飛び入り参加を受け付けます。
※ 障がい等で何らかの配慮が必要な方は、お申込みの際にお早めにご相談ください。

あらかじめ準備しておいた「テキストマイニングを説明する文章」を「AIテキストマイニング by ユーザーローカル」の「ワードクラウド」によって分析した結果。

【ワードクラウド】

スコアが高い単語を複数選び出し、その値に応じた大きさに表示する。単語の色は品詞の種類で異なっており、青色が名詞、赤色が動詞、緑色が形容詞、灰色が感動詞を表す。

KH Coder について



- テキストマイニングのためのフリーソフトウェア。
- KH Coder を用いた研究事例は5000件を超える。さらに、官公庁や企業でも使われている。
- Windows 版は無料。
- Mac版は有料だが、無料で使う方法もある。

〈参考〉 Mac で KH Coder を使う方法

Mac OS 上で動かす方法

1. Mac 向け自動設定ソフトウェア (3,980円) を使ってインストール。
2. 無料公開されたソースコードを使ってインストールする。無料だが、必要なソフトウェアを入れるために、UNIX システムについての十分な知識と経験が必要らしい。インストールするだけで相当手間がかかる。

Windows OS 上で動かす方法

1. Boot Camp や Parallels などによって Mac に Windows OS をインストールし、そこで KH Coder を動かす。
2. Windows マシンを用意し、Mac からリモートデスクトップ接続して動かす。

KH Coder ができる分析の例(1/2)

抽出語リスト

分析	サ変名詞	101	
データ	名詞	97	
テキストマイニング	タグ	87	
活用	サ変名詞	35	
情報	名詞	34	
テキスト	名詞	30	
顧客	名詞	24	
解析	サ変名詞	20	
可能	形容動詞	17	
手法	名詞	17	
ツール	名詞	16	
マイニング	サ変名詞	16	
言語	名詞	16	

【目的】多く出現していた単語とその品詞を確認したい!

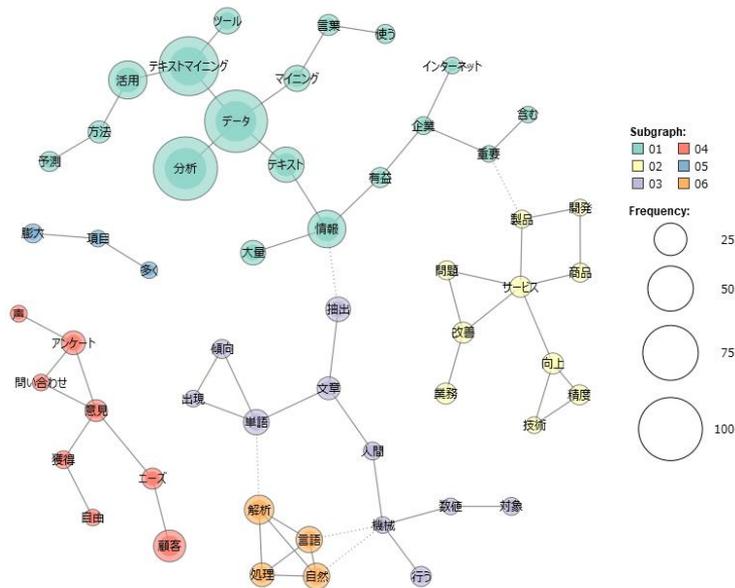
KWICコンコーダンス

分析が可能となった。テキストマイニングの対象としては、顧客が
トデータなどがある。⇒ テキストマイニングの効果⇒テキストマイ
トマイニングの効果⇒ テキストマイニングは、言葉的にはデータマ
で行われるのに対し、テキストマイニングは顧客個人の特性より
取り出すことを総称してテキストマイニングと呼びます。自然言語解
データの活用においても、テキストマイニングは非常に重要な要素と
情報を収集することがテキストマイニングの目的です。⇒ビッグデ
策を打っていくためにはテキストマイニングを有効活用することが当
⇒分析を行う際には、テキストマイニングツールと呼ばれるソフト
ことが一般的です。⇒ テキストマイニングとは? ⇒テキストマイニ
ストマイニングとは? ⇒ テキストマイニングは、大量のテキストを分
。⇒そこで有効なのがテキストマイニングツールの利用です。名前

【目的】単語の本文中での使われ方を確認したい!

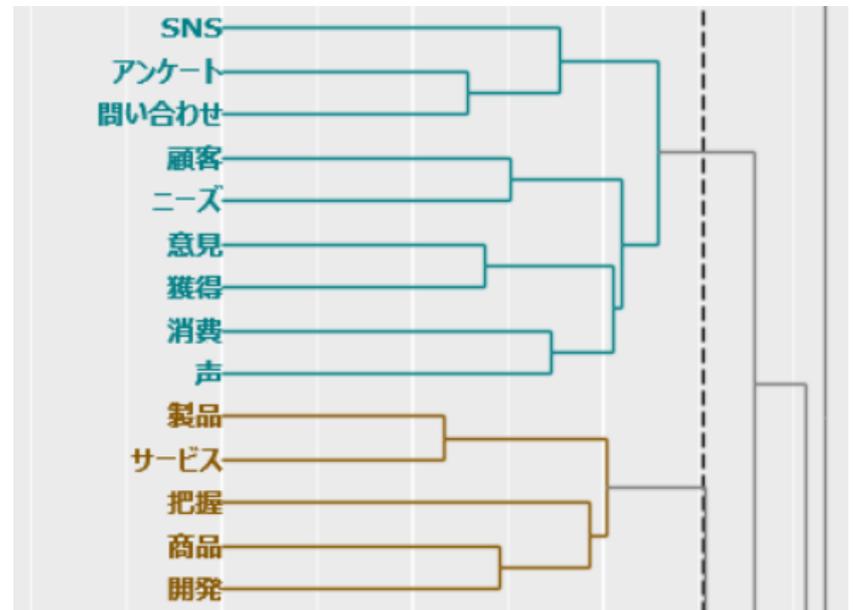
KH Coder できる分析の例(2/2)

共起ネットワーク



【目的】単語同士の結びつき(関連度)をネットワーク図で表したい。

階層的クラスタ分析



【目的】単語同士の結びつき(関連度)を樹形図で表したい。

テキストマイニングの必要性(1/2)

製品Xの顧客満足度調査

株式会社XX

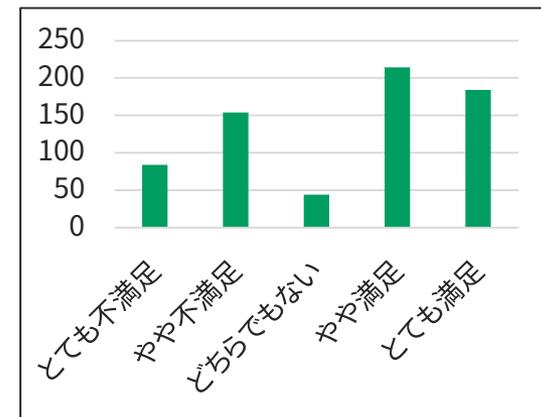
問1 製品Xの満足度を1つ選んでください。

1. とても満足
2. やや満足
3. どちらでもない
4. やや不満足
5. とても不満足

問2 問1の理由を自由にお書きください。

ディスプレイが広くて操作しやすいのが良いです。
また、初心者にも分かりやすいシンプルなユーザーインターフェイスも気に入っています。
ただし、バッテリー持ちが悪いのが玉に瑕です。

数字で表される**量的(定量)データ**に対しては、グラフや統計検定(平均値の検定など)を使えば客観的な分析が簡単。



数字で表されない**質的(定性)データ**に対しては、客観的な分析は難しい。

- 調査対象680人に多かった意見は？
- 機能と装飾のどちらにより満足してる？
- 機器の操作性の課題は？

テキストマイニングの必要性(2/2)

テキストマイニングがないとき

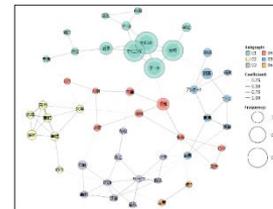


680人の自由記述欄を読むと、インターフェイスに不満を持った顧客が多いように感じました

それってあなたの感想ですよね？ 多いっていうのは、全体の何パーセントですか？ あと、どんな不満が多いんですか？



テキストマイニングがあるとき



自由記述欄の共起ネットワーク分析より、「不満」を表す語句と「遅い」「反応」などの語句が共起しています。そのため、インターフェイスのうち、レスポンスに課題があると考えられます

確かにそれはデータに基づく客観的な意見ですね



講習会の流れ

1. 趣旨説明
2. テキストマイニングとは？
3. KH Coder のデモンストレーション
4. 質疑応答

KH Coder の使い方

1. テキストファイル (txt, word, Excelなど) の準備
2. プロジェクトの新規作成
3. 前処理の実行
4. 抽出語リストの確認
5. 語の取捨選択
6. 再度前処理の実行
7. 各種分析の実行 (抽出語リスト, 共起ネットワーク, 階層的クラスタ分析など)

KH Coder の使い方

1. テキストファイル (txt, word, Excelなど) の準備
2. プロジェクトの新規作成
3. 前処理の実行
4. **とにかくやってみよう!**
抽出語リストの確認
5. 語の取捨選択
6. 再度前処理の実行
7. 各種分析の実行 (抽出語リスト, 共起ネットワーク, 階層的クラスタ分析など)

デモンストレーション開始

- 既にコンピュータに KH Coder がダウンロード&インストールされた状態からスタートします。
- ダウンロードは KH Coder の公式サイトから可能です。

KH Coder の参考サイト

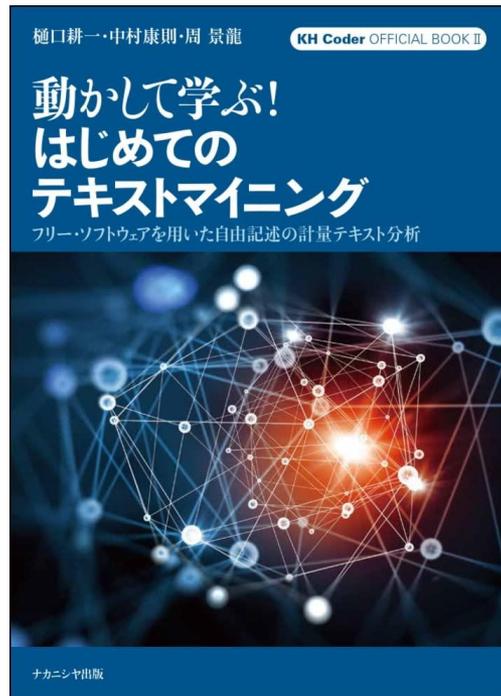
The screenshot displays the KH Coder website's main page, titled "主な機能と分析の手順" (Main Functions and Analysis Procedure). The page includes a navigation menu, a search bar, and a list of social media links. The main content area features a "概要図" (Overview Diagram) section, which contains a flowchart illustrating the analysis process. The flowchart starts with "テキスト型データ" (Text-type data) and "テキストを分割するためのデータ" (Data for text segmentation), both of which are processed as "分析対象ファイル" (Analysis target files) and input into KH Coder. The central step is "データ中から語を自動抽出" (Automatic extraction of words from data), which leads to "外部変数として蓄積" (Accumulation as external variables). This step branches into three main categories: "処理の目的" (Purpose of processing), "KH Coderのコマンド" (KH Coder commands), and "語と語の結びつきを探る" (Exploring word relationships). The "処理の目的" category includes "多く出現していた語の確認" (Confirmation of frequently appearing words) and "内容が似た文書の群を探す" (Searching for clusters of similar documents). The "KH Coderのコマンド" category includes "抽出語リスト" (Extraction word list), "階層的クラスター分析" (Hierarchical cluster analysis), "多次元尺度構成法" (Multiple correspondence analysis), "共起ネットワーク" (Co-occurrence network), "関連語検索" (Related word search), "トピックモデル" (Topic model), "対応分析" (Correspondence analysis), and "クラスター分析" (Cluster analysis). The "語と語の結びつきを探る" category includes "共起ネットワーク" (Co-occurrence network), "関連語検索" (Related word search), "トピックモデル" (Topic model), "対応分析" (Correspondence analysis), and "クラスター分析" (Cluster analysis). The page also features three book covers on the right side, each with a link to a support page: "KH Coderの入門書 (サポートページへ)", "KH Coderの本 (サポートページへ)", and "知を再構築する 異分野融合研究のためのテキストマイニング ※川端亮先生 (KH Coderの)".

KH Coder の公式サイト。KH Coder のダウンロードはここからできる。基本的な使い方も解説。

<<https://khcoder.net/diagram.html>>

KH Coder のおすすめ参考書（理工学図書館に収蔵）

動かして学ぶ!
はじめてのテキストマイニング



書誌ID=2004546008

開発者が自ら執筆した KH Coder の入門書。
押さえておくべき機能を厳選して紹介。

社会調査のための
計量テキスト分析【第2版】

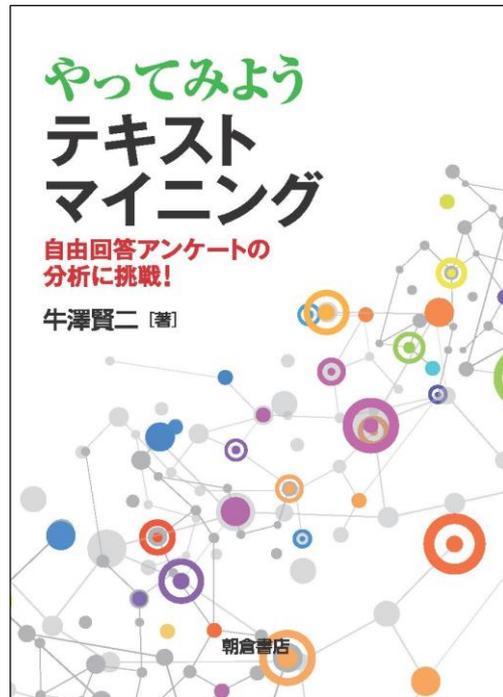


書誌ID=2004538143

開発者が自ら執筆した KH Coder の解説書。
応用まで詳しく述べた上級者向きの一冊。

KH Coder のおすすめ参考書（理工学図書館に収蔵）

やってみようテキストマイニング
: 自由回答アンケートの分析に挑戦!



書誌ID=2004465422

KH Coder の使い方を学ぶ上で、実際のアンケートデータを使った分析が体験できる。

テキストマイニング入門：Excel
とKH Coderでわかるデータ分析



書誌ID=2004489689

KH Coder の基礎と事例について、マンガやイラスト、図解を用いて分かりやすく解説。

1. 趣旨説明
2. テキストマイニングとは？
3. KH Coder のデモンストレーション
- 4. 質疑応答**

質疑応答

- 質問・相談がある方は、挙手をお願いします！
- オンライン参加の方はマイクのミュートを解除して発言いただくか、質問をチャットにお書きください！
- アンケートの回答にもご協力ください。(対面の方は紙に記入、オンラインの方はチャット欄に送信したフォームに記入)

1. 趣旨説明
2. テキストマイニングとは？
3. KH Coder のデモンストレーション
4. 質疑応答

LSデスクのご紹介

大学院のスタッフが駐在し、学生の皆さんの学習をサポートします。
困りごとがあれば、気軽にお越しください！

時間：平日11:00-17:00(金曜日は14:00-17:00)

場所：東館1階LSデスク

相談方法：対面、Zoom、Webフォーム(メール回答)

発表LSの従事時間：毎週水曜日・金曜日 14:00-17:00

ホームページ：<https://www.library.osaka-u.ac.jp/research/ta/>

※LSデスクがお休みになることもあるため、

Twitter(https://twitter.com/LS_Ourikolib)を確認してください！

