科目名	テキストマイニング演習
科目属性	専門科目Ａ群
担当教員	斎藤　俊則
単位数	２単位（SC　0.25 単位）

【授業の目的・ねらい】

この科目では自由記述アンケートや SNS の投稿などのテキストデータを計算機を用いて分析する手法であるテキストマイニングの入門的な知識と，プログラミングによるテキストマイニングの実行方法の習得を目指す．プログラミング言語は統計解析などの用途に利用される R 言語を使用する．

【授業概要】

テキストマイニングは自然言語で書かれたテキストデータから処理の事前には明らかでなかった有用な情報を計算機による自動処理を用いて取り出すデータ分析手法である．この手法を用いることで，たとえば質問紙調査における自由記述や調査現場で記述したフィールドノーツ，あるいは官公庁が公開する大量の議事録などから，手法の統計学的な正当性を担保しつつ意味のある情報を抽出することが可能となる．

この授業では，上記の手法を必要とする学生を対象に，自習によるこの手法の習得と各自の研究データへの応用を支援する．自習を効果的に行うために合計 2 回（1 回 90 分）のオンラインによるスクーリングを行う．また，そのほかに Google Classroom 上で学習に役立つリソースを公開し，質問等を受け付ける．

履修を検討する際には以下の点に留意されたい．

- 指定教科書とプログラミングに必要な機材（教科書の指示に基づくソフトウェアおよびそのソフトウェアを問題なく実行できる PC，ただしソフトウェアは基本的に無料のものを選択可能である）とネットワークを自力で準備できること
- プログラミング経験は問わないが，機材のセットアップからすべてを自習によって行う準備があること（詳細は次の項目を参照）．
- 「研究方法論 1」の事前履修を推奨する．
- この授業を通して自習によるプログラミングの学習に必要な基本的なスキルとマインドセットを獲得することに強い意欲を持つこと．必要なスキルは以下のものである．
  - 不明な事柄をネット上から自力で調べ出す検索スキル
  - 検索によって発見した英文を含めた各種資料を読み解き，問題解決に役立ちそうなアイデアは即座に試す（プログラムに反映し実行する）実践的な学習スキル
  - バグ（プログラムの不具合）が生じた際にその原因の可能性を一つ一つ潰しながら突き止める推論および試行錯誤のスキル
  - 自力で解消できない不明な点が生じた際には Google Classroom 上で回答者が極力回答しやすい仕方で自ら聞き出せる質問スキル

  必要なマインドセットは以下のものである．
  - 新しい概念や仕組みの発見・理解・構築・習得・創造を楽しむ知的好奇心
  - 上手くゆかない場合に諦めずに何度でもやり直すレジリエンス
  - コンピュータの自動処理を用いて面倒を解消するためならばプログラミングの面倒を厭わない「ものぐさ精神」

【授業の到達目標】

履修者が各自の研究に関連するテキストデータの性質を，自作したプログラムを用いて分析できること．分析手法は対応分析，ワードクラウドとネットワークグラフの作成，クラスター分析，トピックモデル分析などを想定する．

【授業計画】

1. 学習環境のセットアップ（教科書第 1，2 章）
2. R の基本 1：プロジェクトからデータフレームの操作まで（教科書第 3 章）
3. R の基本 3：リストからグラフィックスまで（教科書第 3 章）
4. 文字列処理と正規表現（教科書第 4 章）

5. RMeCab によるテキスト解析 1: RMeCab パッケージから頻度表の作成と利用まで（教科書第 5 章）
6. RMeCab によるテキスト解析 2: 単語文書行列の生成から関数一覧まで（教科書第 5 章）
7. ウェブスクレイピングによるウェブサイト上のテキストデータの分析（教科書第 6 章）
8. アンケート自由記述文の分析（対応分析）（教科書第 7 章）
9. 文学作品の解析（ワードクラウドとネットワークグラフ）（教科書第 8 章）
10. テキストの分類（クラスター分析，トピックモデル）（教科書第 9 章）
11. 書き手の判別（文体比較）（教科書第 10 章）
12. SNS のテキストデータの分析（API の利用）（教科書第 11 章）
13. テキストマイニングの応用 1: 分析計画，データの準備
14. テキストマイニングの応用 2: 分析の実行
15. テキストマイニングの応用 3: 分析結果の評価

【評価方法】

評価材料及びその配分は以下の通りとします.

1. スクーリング（30%）
2. レポート課題（30%）
3. 科目習得試験（40%）

スクーリングはスクーリング出席状況（質問，議論への参加等）に置き換えて評価材料とします.

レポート課題の出題内容は学修指導書にて詳述します．評価観点は以下の通りとします.

A) 学修指導書にしたがってスクーリング第 2 回までの自習を着実に進められていること

科目習得試験の出題内容は学習指導書にて詳述します．評価観点は以下の通りとします.

A) テキスト分析の目的を適切に定義し，かつ定義された目的に従って最適な分析手法を選択できること
B) 分析のためのプログラムが適切に動作し，かつ目的にかなった分析結果を得られること
C) 自習によるテキストマイニングに積極的に取り組み，成果を得るための試行錯誤を惜しまずに行いきることができること

【教科書】

1. 石田基広， R によるテキストマイニング入門 第 2 版，森北出版 ISBN 9784627848429

【参考図書】

スクーリング時および Google Classroom 上で適宜示します.