

科目名	データサイエンス（ことば）		担当教員	長谷川 信、長尾 純、宮原 淳	
単位	1単位	講義区分		ナンバリング	
期待される学修成果	言語・文学、実務・英語教育				
アクティブ・ラーニングの要素	実習、フィールドワーク				
実務経験					
実務経験を生かした授業内容					
到達目標及びテーマ	現在は、人文科学や社会科学において有用なデータが数多く公開されており、分析用のソフトウェアも整備されている。また、インターネットを利用すれば、さらに膨大な蓄積データに当たることができる。本授業では、言語データを中心として、データの収集・管理・分析のための知識と技術を身につけ、データ構造を理解して分析を進められることが目標となる。				
授業の概要	授業の前半では、文章データの収集・管理・分析の目的や利用方法などを学習する。併せてソフトウェア等を用いた演習を行いデータ分析の知識や技術の定着を図る。授業の後半では、各々で分析テーマを設定して、データの収集～分析を行い、データ取り扱いの方法や、データ理解の視点などを身に付ける。 なお、本授業は「ICT基礎」「ICT活用」「データサイエンス入門」「データサイエンス基礎」の履修を前提として進める。				

授業計画	
第1回	オリエンテーション
第2回	文章データの収集と管理
第3回	コーパスの活用と研究(1)基礎
第4回	コーパスの活用と研究(2)活用
第5回	自然言語処理の技術
第6回	文章データの集計と統計処理
第7回	文章データの可視化
第8回	計量テキスト分析(1)準備
第9回	計量テキスト分析(2)基礎
第10回	計量テキスト分析(3)応用
第11回	機械学習によるテキスト分析・応答(1)準備
第12回	機械学習によるテキスト分析・応答(2)アルゴリズム
第13回	機械学習によるテキスト分析・応答(3)学習
第14回	機械学習によるテキスト分析・応答(4)評価
第15回	機械学習によるテキスト分析・応答(5)応用

事前学修	0.5時間	前提となる授業（ICT基礎、ICT活用、データサイエンス入門、データサイエンス基礎）の内容を復習する。予め提示するテーマに沿って、必要な情報を調査し、授業で利用できるよう準備する。
事後学修	0.5時間	授業の成果を確認・整理してまとめる。授業での成果物を見直して誤りを修正する。処理操作を確認して、復習する。
フィードバックの方法	課題等を回収した際は次の講義にて解説を実施する。	

成績評価方法	割合（％）	評価基準等
レポート	100%	各回およびテーマ毎に確認テスト、レポート等の課題を課す。
定期試験	0%	実施しない
補足事項	受講には管理者権限を有するWindowsパソコンを持参してください(*1)。また、MS Office、テキストエディタは導入を済ませてください。パソコンの利用に関する不明点は情報センターに問い合わせてください。 *1 大学の貸与パソコン、教室パソコンでは受講できません。	

教科書				
書名	著者	出版社	ISBN	備考
ことばのデータサイエンス	小林雄一郎	朝倉書店	9784254510638	2019
参考資料	石川慎一郎「ベーシックコーパス言語学」ひつじ書房(ISBN:978-4894765474) グラム・ニュービッツ、他「自然言語処理の基本と技術」翔泳社(ISBN:978-4798128528) 樋口耕一「社会調査のための計量テキスト分析」ナカニシヤ出版(ISBN:978-4779514746) 赤石雅典「最短コースでわかる Pythonプログラミングとデータ分析」日経BP(ISBN:978-4296201129) その他、必要な資料は授業で配布する。			