


Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Философский факультет

УТВЕРЖДАЮ:

Декан Философского факультета

 Е.В. Сухушина

«04» июля 2023 г.

Рабочая программа дисциплины

Анализ естественного языка в гуманитарных исследованиях

по направлению подготовки

39.03.01 Социология

Направленность (профиль) подготовки:

«Социология»

Форма обучения

Очная

Квалификация

Бакалавр

Год приема

2023

Код дисциплины в учебном плане: Б1.О.25

СОГЛАСОВАНО:

Руководитель ОП

 В.В. Кашпур

Председатель УМК

 Т.В. Фаненштиль

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

- ОПК-1 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности. Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.1 Определяет релевантные для решения поставленной задачи источники информации, включая национальные и международные базы данных, электронные библиотечные системы, специализированные пакеты прикладных программ.

ИОПК-1.2. Проводит поиск социологической информации, необходимой для решения поставленной задачи, получает на ее основе социологические данные.

2. Задачи освоения дисциплины

- Освоить методы обработки текста на естественном языке
- Освоить методы обработки слабоструктурированных данных
- Научиться извлекать информацию из открытых источников
- Научиться применять понятийный аппарат Natural Language Processing для решения практических задач профессиональной деятельности
- Научиться использовать современные инструменты анализа естественного языка для решения практических задач в профессиональной деятельности.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к блоку Б.1, является обязательной для изучения.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Семестр 7, зачёт.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: социально-сетевой анализ, методология и методы социологических исследований, сбор и анализ данных, компьютерное обеспечение социологических исследований.

6. Язык реализации

Русский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 2 з.е., 72 часа, из которых:

– практические занятия: 32 ч.;

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Введение в обработку естественного языка

Основные понятия. Возможности анализа естественного языка. Источники данных для анализа. Основные методы текстового анализа в социологии.

Тема 2. Контент-анализ

Возможности, ограничения, процедура проведения, примеры применения в социологии, особенности кодировки текстовой информации.

Тема 3. Дискурс-анализ

Возможности, ограничения, процедура проведения, примеры применения в социологии.

Тема 4. Natural Language Processing

Возможности, ограничения, процедура проведения, примеры применения в социологии, области применения. Сложности NLP: полисемия, омонимия, местоименная анафора, эллипсис.

Тема 5. Нейросети.

Особенности применения, примеры использования, процедура обучения, обучающая выборка, готовые модели NLP.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, подготовке устных сообщений, индивидуального анализа текстов, и фиксируется в форме контрольной точки не менее одного раза в семестр.

Критерии оценивания для текущей аттестации и типовые задания представлены в фонде оценочных материалов. В полном объеме фонд оценочных материалов хранится на кафедре социологии.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет в седьмом семестре проводится в форме выполнения индивидуального проекта, демонстрирующего усвоение основных принципов анализа естественного языка и получения навыков применения специализированного программного обеспечения для текстовой аналитики.

Зачет проводится в устной форме и включает в себя защиту индивидуального проекта, проверяющего ИОПК-1.1., ИОПК-1.2.

Критерии оценивания для промежуточной аттестации, а также типовые задания представлены в Фонде оценочных материалов. В полном объеме фонд оценочных материалов хранится на кафедре социологии.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=3178>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине представлены в Фонде оценочных материалов.

в) План семинарских / практических занятий по дисциплине представлен в фонде оценочных материалов.

г) Методические указания по организации самостоятельной работы студентов размещены в ЭОИС НИ ТГУ.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

1. Aggarwal C. C., Zhai C. X. (ed.). Mining text data. – Springer Science & Business Media, 2012.
2. Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.
3. Jurafsky D., Martin J.H. Speech and Language Processing (3rd ed. draft). – URL: <https://web.stanford.edu/~jurafsky/slp3/>
4. Lembersky G., Ordan N., Wintner S. Language models for machine translation: Original vs. translated texts // Computational Linguistics. – 2012. – Т. 38. – №. 4. – С. 799-825.
5. Sentiment Analysis and Subjectivity // Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.

6. The handbook of computational linguistics and natural language processing / Eds Alexander Clark, Chris Fox, Shalom Lappin. – Wiley-Blackwell, 2010. – 801 p.

б) дополнительная литература:

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В.,
2. Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык – Киев: Наук, думка, 1991. – 208 с.
3. Беликов В. П., Крысин Л. П. Социолингвистика. Учебник для бакалавриата и магистратуры. 2-е изд. - М.: Юрайт, 2017.
4. Марчук, Ю.Н. Компьютерная лингвистика: учебное пособие / Ю.Н. Марчук. – М.: АСТ: Восток-Запад, 2007. – 317 с.
5. Николаев И.С., Митренина О.В., Ландо Т.М. Прикладная и компьютерная лингвистика. – URSS, 2017. – 320 с.
6. Новое в зарубежной лингвистике: Вып. XXIV. Компьютерная лингвистика: Пер. с англ./Сост., ред. и вступ. ст. Б. Ю. Городецкого. – М.: Прогресс, 1989. – 432 с.
7. Носков А.А., Пескова О.В., Ягунова Е.В. – М.: МИЭМ, 2011. – 272 с.
8. Шерстяных И.В. Основы социолингвистики [Электронный ресурс]: учебное пособие/ Шерстяных И.В.— Электрон. текстовые данные. — Иркутск: Иркутский государственный лингвистический университет, 2012.— 160 с.
9. Venamara F., Taboada M., Mathieu Y. Evaluative language beyond bags of words: Linguistic insights and computational applications // Computational linguistics. – 2017. – Т. 43. – №. 1. – С. 201-264.
10. Bender E. M. Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax //Synthesis Lectures on Human Language Technologies. – 2013. – Т. 6. – №. 3. – С. 1-184.
11. Cohen K. B., Demner-Fushman D. Biomedical natural language processing. – John Benjamins Publishing Company, 2014.
12. Farzindar A., Inkpen D. Natural language processing for social media //Synthesis Lectures on Human Language Technologies. – 2015. – Т. 8. – №. 2. – С. 1-166.
13. Fitzpatrick E., Bachenko J., Fornaciari T. Automatic detection of verbal deception //Synthesis Lectures on Human Language Technologies. – 2015. – Т. 8. – №. 3. – С. 1-119.
14. Goker A., Davies J. (ed.). Information retrieval: searching in the 21st century. – John Wiley & Sons, 2009.
15. Liu B. Sentiment analysis: Mining opinions, sentiments, and emotions. – Cambridge University Press, 2015.
16. Munzert S., Rubba C., Meißner P., Nyhuis D. Automated data collection with R: A practical guide to web scraping and text mining. – John Wiley & Sons, 2014.
17. Schäfer R., Bildhauer F. Web corpus construction //Synthesis Lectures on Human Language Technologies. – 2013. – Т. 6. – №. 4. – С. 1-145.
18. Taboada M. et al. Lexicon-based methods for sentiment analysis // Computational linguistics. – 2011. – Т. 37. – №. 2. – С. 267-307.

в) ресурсы сети Интернет:

1. Федеральный образовательный портал <http://www.ecsocman.edu.ru/>
2. Сайт журнала «Экономическая социология» <http://ecsoc.hse.ru/>
3. <http://link.springer.com/>
4. <http://www.annualreviews.org/>
5. <http://www.jstor.org/>
6. – Общероссийская Сеть КонсультантПлюс Справочная правовая система. <http://www.consultant.ru>

13. Перечень информационных технологий

- а) лицензионное и свободно распространяемое программное обеспечение:
– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ –
<http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ –
<http://vital.lib.tsu.ru/vital/access/manager/Index>

– ЭБС Лань – <http://e.lanbook.com/>

– ЭБС Консультант студента – <http://www.studentlibrary.ru/>

– Образовательная платформа Юрайт – <https://urait.ru/>

– ЭБС ZNANIUM.com – <https://znanium.com/>

– ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Дунаева Дарья Олеговна, ассистент кафедры социологии, аналитик центра прикладного анализа больших данных