

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет соціології

Кафедра методології та методів соціологічних досліджень



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Сучасні методи аналізу текстів

для студентів

галузь знань 05 Соціальні та поведінкові науки
спеціальність 054 «Соціологія»
освітній рівень магістр
освітня програма «Соціологія»
вид дисципліни вибіркова

Форма навчання	денна
Навчальний рік	20__/20__
Семестр	4
Кількість кредитів ECTS	5
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	іспит

Викладачі: Людмила ЮЗВА

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» ____ 20__ р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) «__» ____ 20__ р.
(підпис, ПІБ, дата)

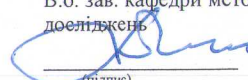
КИЇВ – 2021

Розробниця:

Людмила ЮЗВА, к.соц.н., доцентка кафедри методології та методів соціологічних досліджень

ЗАТВЕРДЖЕНО

В.о. зав. кафедри методології та методів соціологічних досліджень


(підпис) Андрій ГОРБАЧИК
(прізвище та ініціали)

Протокол № 13 від «31» серпня 2021 р.

Схвалено науково - методичною комісією факультету соціології

Протокол від «31» вересня 2021 року № 1

Голова науково-методичної комісії  Тетяна ЧЕРВІНСЬКА

1. Мета дисципліни – навчання студентів сучасним методам аналізу текстів.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

1. Знати базові методи аналізу соціологічної інформації

2. Вміти працювати з комп'ютером

3. Володіти елементарними навичками роботи з текстовими масивами

3. Анотація навчальної дисципліни:

Дисципліна належить до дисциплін за вибором з переліку (№3) в межах освітньої програми «Соціологія» на здобуття освітнього ступеню магістр за спеціальністю 054 «Соціологія», галузі знань 05 «Соціальні та поведінкові науки». Сучасне суспільство характеризується чи не надмірною переповненістю документами різного типу. Візуальні, текстові, відео- та аудіо-документи оточують сучасну особу посекундно. Такий стан справ вимагає певного рівня компетентності у використанні відповідних методів. До того ж із щоденним стрімким продукуванням інформації в онлайн-середовищі та поза ним й паралельним розвитком ІТ-технологій (програмування) видозмінюються методи аналізу текстів. Курс спрямований на поглиблення та доповнення знань з аналізу документів, вироблення практичних вмінь застосування сучасних методів та технік аналізу текстів.

4. Завдання (навчальні цілі):

- Ознайомити з основними сучасними методами аналізу текстів
- Навчити використовувати сучасні методи аналізу текстів
- Навчити збирати дані для подальшого аналізу
- Навчити аналізувати отримані емпіричні дані та робити аналітичні висновки на їх основі
- Дати навички у візуальній подачі та інтерпретації отриманих даних

Це спрямовано на формування компетентностей:

- ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.
- ЗК02. Здатність працювати автономно.
- СК03. Здатність проектувати й виконувати соціологічні дослідження, розробляти й обґрунтовувати їхню методологію.
- СК04. Здатність збирати та аналізувати емпіричні дані з використанням сучасних методів соціологічних досліджень.

5. Результати навчання за дисципліною:

Результат навчання (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність)		Форми (та/або методи і технології) викладання і навчання	Методи оцінювання та пороговий критерій оцінювання	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання			
1.1	Знати математичні методи й моделі в лінгвістиці	Лекція, практична робота, самостійна робота	практична робота, тест	10
1.2	Вміти застосовувати інструменти для кількісного аналізу текстів на основі автоматизованих вибірок	Лекція, практична робота, самостійна робота	практична робота, тест	15
2.1	Знати методи та інструментарій аналітики медіа текстів	Лекція, практична робота, самостійна робота	практична робота, тест	15
2.2	Вміти застосовувати методи аналітики та системи моніторингу медіа-контенту	Лекція, практична робота, самостійна робота	практична робота, тест	20
3.1	Знати методологію комп'ютерного контент-аналізу та принципи аналізу текстів на його основі	Лекція, практична робота, самостійна робота	практична робота, тест	40

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Програмні результати навчання (назва)	Результати навчання дисципліни (код)				
	1.1	1.2	2.1	2.2	2.3
ПР09. Планувати і виконувати наукові дослідження у сфері соціології, аналізувати результати, обґрунтовувати висновки.	+	+	+	+	+
ПР12. Обґрунтовано використовувати новітні методи збору та аналізу соціологічної інформації для вирішення практичних завдань.	+	+	+	+	+

7. Схема формування оцінки.

7.1 Форми оцінювання студентів:

- семестрове оцінювання:

1. Практичні заняття з роботами у формі представлення та обговорення відповідного завдання:
 - ПР¹ з представлення однієї з математичних моделей мови (тексту) з теми 1: РН 1.1 – 3/ 5 балів
 - ПР з формування автоматизованої вибірки для аналізу текстів з теми 2: РН 1.2 – 3/ 5 балів
 - ПР з ознайомлення з одним з інструментів для кількісного аналізу текстів з теми 3: РН 1.2 – 3/ 5 балів
 - ПР з ознайомлення з моделями поширення текстової інформації в медіа з теми 4: РН 2.1 – 3/ 5 балів
 - ПР з ознайомлення з методами аналітики кросмедійних ЗМІ на вибір (√з теми 5: РН 2.1, РН 2.2 – 3/ 5 балів:
 - ✓ аналіз відвідуваності мережевого ресурсу (статистика, тенденції, абсолютні і відносні показники);
 - ✓ аналіз поведінки відвідувачів на сторінці ресурсу;
 - ✓ аналіз юзабіліті (щільності кліків, конверсійних шляхів відвідувачів мережевого ресурсу);
 - ✓ бенчмаркінг (порівняння із загальними тенденціями і з конкурентами за допомогою незалежних дослідних ресурсів, як-от Alexa, GemiusAudience, Google Trends тощо)
 - ПР з ознайомлення з системами моніторингу медіа-контенту та їхніми методами з теми 6: РН 2.2 – 3/ 5 балів
 - ПР з ознайомлення з основами методології комп'ютерного контент-аналізу з теми 7: РН 3.1 – 3/ 5 балів
 - ПР з ознайомлення з принципами аналізу текстів на основі комп'ютерного контент-аналізу з теми 8: РН 3.1 – 3/ 5 балів
2. Контрольна робота (тест) з тем 1-8: РН 1.1, РН 1.2, РН 2.1, РН 2.2, РН 2.3, РН 3.1 - 12/ 20 балів

- підсумкове оцінювання – *іспит*: РН 1.1, РН 1.2, РН 2.1, РН 2.2, РН 2.3, РН 3.1

- умови допуску до підсумкового оцінювання:

Для студентів, які набрали сумарно меншу кількість балів за критично-розрахунковий мінімум (36 балів) для одержання допуску до іспиту обов'язковим є повторне виконання вибраних ПР та виконання контрольної роботи.

7.2 Організація оцінювання:

1. ПР з представлення однієї з математичних моделей мови (тексту) з теми 1: РН 1.1.
2. ПР з формування автоматизованої вибірки для аналізу текстів з теми 2: РН 1.2.
3. ПР з ознайомлення з одним з інструментів для кількісного аналізу текстів з теми 3: РН 1.2.

¹ ПР – практична робота.

4. ПР з ознайомлення з моделями поширення текстової інформації в медіа з теми 4: РН 2.1.
5. ПР з ознайомлення з методами аналітики кросмедійних ЗМІ на вибір (□) з теми 5: РН 2.1, РН 2.2.
6. ПР з ознайомлення з системами моніторингу медіа-контенту та їхніми методами з теми 6: РН 2.2.
7. ПР з ознайомлення з основами методології комп'ютерного контент-аналізу з теми 7: РН 3.1.
8. ПР з ознайомлення з принципами аналізу текстів на основі комп'ютерного контент-аналізу з теми 8: РН 3.1.
9. Контрольна робота (тест) з тем 1-8: РН 1.1, РН 1.2, РН 2.1, РН 2.2, РН 2.3, РН 3.1.

7.3 Шкала відповідності оцінок

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

8. Структура навчальної дисципліни. Тематичний план лекцій та практичних занять

№ п/п	Номер і назва теми	Кількість годин		
		Лекції	Практичні	Самостійна робота
1. Математичні методи й моделі в лінгвістиці				
1	Математичні методи й моделі в лінгвістиці	2	2	10
2	Автоматизовані вибірки в аналізі текстів	2	2	10
3	Корпус інструментів для кількісного аналізу текстів	4	4	14
2. Методи та інструментарій аналітики медіа текстів				
4	Моделі поширення текстової інформації в медіа	2	2	12
5	Методи аналітики кросмедійних ЗМІ	2	2	12
6	Системи моніторингу медіа-контенту та їхні методи	4	4	14
3. Комп'ютерний контент-аналіз				
7	Методологія комп'ютерного контент-аналізу: основи	4	4	14
8	Принципи аналізу текстів на основі комп'ютерного контент-аналізу	4	2	14
	<i>Контрольна робота (тест)</i>		2	2
	ВСЬОГО	24	24	102

Загальний обсяг 150 год., в тому числі:

Лекцій – 24 год.

Практичні заняття – 24 год.

Самостійна робота – 102 год.

9. Рекомендовані джерела:

Основна: (Базова)

1. Горбачик А. Компьютерная обработка текстов в качественных и количественных социологических исследованиях // Социология: теория, методы, маркетинг. - №1. - 2006. - С. 124-133.
2. Bates M. Models of natural language understanding / Madeleine Bates // Proceedings of the National Academy of Sciences of the United States of America. – 1995. – Vol. 92, № 22. – P. 9977-9982.
3. Chomsky N. Three models for the description of language / Noam Chomsky // IEEE Transactions on Information Theory. – 1956. – Vol. 2, № 3. – P. 113-124.
4. Alan McKee A. Textual Analysis. 2003. - <https://methods.sagepub.com/Book/textual-analysis>
5. Li W. Random texts exhibit Zipf's-law-like word frequency distribution / Wentian Li // IEEE Transactions on Information Theory. – 1992. – Vol. 38, № 6. – P. 1842-1850.

Додаткова:

1. Васильев О., Чалий О., Васильева І. (2019) Математичні методи й моделі в лінгвістиці. Verba et Numeri: Кількісні підходи до аналізу мови й тексту, 27: 9-28.
2. Захарченко А. П. (2017) Поширення контенту найпопулярніших медіа-публікацій у соціальних мережах як соціальна дія. - <http://dx.doi.org/10.16926/j.2017.13.10>
3. Іванов В. Загальні принципи аналізу документів - journal.univ.kiev.ua
4. Іванов О. В. (2013) Кількісний аналіз тексту чи продукування числових артефактів: аудит контент-аналітичних досліджень. Наукові записки, №148: 11–14.
5. Іванов О. В. (2012) Кількісний контент-аналіз: проблема контексту. Вісник Харківського університету імені В.Н. Каразіна, №999: 95-98. - <https://periodicals.karazin.ua/ssms/article/view/13926>

6. Іванов О. В. (2013) Класичний контент-аналіз та аналіз тексту: термінологічні та методологічні відмінності. Вісник Харківського університету імені В.Н. Каразіна, №1045: 69-74. - <https://periodicals.karazin.ua/ssms/article/view/13841>
7. Іванов О. В. (2009) Комп'ютерний контент-аналіз: проблеми та перспективи вирішення. Методологія, теорія та практика соціологічного аналізу сучасного суспільства. №15: 335–340.
8. Кікінчук І., Густі М. (2012) Безкоштовні та умовно-безкоштовні комп'ютерні програми для проведення контент-аналізу. Інформація, комунікація, суспільство : матеріали І Міжнародної наукової конференції ІКС-2012. Львів: Видавництво Львівської політехніки: 244–245. - http://ena.lp.edu.ua:8080/bitstream/ntb/14653/1/116_244-245_maket-ena-ntb%28099%29.pdf?fbclid=IwAR35PCoT3p3pDXiUC0X5mxYY00-U_9YImTrDVVKYpM0eMPe9Snaw0zqlsQ
9. Кросмедіа: контент, технології, перспективи: колективна моногр. (2017) (за заг. ред. д.н.соц.ком. В. Е. Шевченко). Київ: 234 с.
10. Лотман Ю. М. (1994) Семантика культури и понятие текста// Логика: Логические основы общения. Хрестоматия / Сост. В. Ф. Берков, Я. С. Яскевич, И. И. Дубинин. - М.: Наука: 308-312.
11. Ситник О. (2017) Методи та інструментарій аналітики кросмедійних ЗМІ. - https://www.academia.edu/35581441/%D0%A8%D0%B5%D0%B2%D1%87%D0%B5%D0%BD%D0%BA%D0%BE_%D0%92%D0%95_%D0%9A%D0%BE%D0%BD%D0%B2%D0%B5%D1%80%D0%B3%D0%B5%D0%BD%D1%82%D0%BD%D1%96%D1%81%D1%82%D1%8C_%D1%8F%D0%BA_%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D0%BD%D0%B0_%D0%BE%D0%B7%D0%BD%D0%B0%D0%BA%D0%B0_%D1%81%D1%83%D1%87%D0%B0%D1%81%D0%BD%D0%B8%D1%85_%D0%BC%D0%B5%D0%B4%D1%96%D0%B0
12. Тарануха В. Ю. (2014) Інтелектуальна обробка текстів / В. Ю. Тарануха. – Київ: 80 с.
13. Шелесток Е. В. Методика выявления количественных показателей истинности, информативности и информационной плотности текстов // Система языка: синхрония и диахрония: Межвузовский сборник научных статей. – Уфа: РИЦ БашГУ, 2009. - С. 151-156.
14. Collins, A., Neville, P., & Bielaczyc, K. (2000). The role of different media in designing learning environments. International Journal of Artificial Intelligence in Education, 11: 144–162 p.
15. Corpus linguistics: An international handbook (Handbooks of linguistics and communication science) (2009), [ed. by A. Lüdeling, M. Kytö]. – Mouton de Gruyter, 2009. – Vol. 1. – P. 1–154.
16. Manning C. D. (1999) Foundations of statistical natural language processing. / Christopher D. Manning, Hinrich Schütze. – MIT Press, P. 22-151, 317-341, 381-407.

4. NoksFishes
5. SoMo
6. Corestone Group
7. Semantrum
8. YouScan
9. ECOSAP
10. Mediateka
11. Медіаєна екосистема (<http://mediaecosystem.org/>)

Підкасти:

<https://www.facebook.com/VoxUkraine.Ukrainian/videos/844263626216532/>

10. Додаткові ресурси (за наявності):

1. <http://voyant-tools.org/>
2. <http://linis-crowd.org/>
3. <http://worditout.com/>
4. <http://www.google.com.ua/trends/>
5. <http://advego.ru/text/seo/>
6. <http://www.analizpisem.ru/>
7. <https://audit.te-st.ru/readability/>
8. Textanz – <http://www.textanz.com/index.php>
9. TextQuest – <http://www.textanz.com/index.php>
10. WordStat – <http://www.provalisresearch.com/wordstat.html>
11. Kwalitan – <http://www.kwalitan.nl/engels/>
12. MAXQDA - <http://www.maxqda.com/lang/ru>

11. Системи моніторингу текстового контенту:

1. InfoStream
2. LOOQME
3. SemanticForce