

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ»**

Кафедра прикладної математики та інформатики

**МЕТОДИЧНІ ВКАЗІВКИ
до виконання практичних занять
з дисципліни «Методи та засоби обробки
текстової інформації»**

**для студентів ОС «Магістр»
галузі знань F Інформаційні технології
спеціальності F3 Комп'ютерні науки
денної та заочної форм навчання**

Дрогобич, 2025

УДК 004.912(072)

М 54

Методичні вказівки до виконання практичних занять з дисципліни «Методи та засоби обробки текстової інформації» для студентів ОС «Магістр» галузі знань F Інформаційні технології спеціальності F3 Комп'ютерні науки денної та заочної форм навчання / уклад. Т. В. Алтухова. – Дрогобич: ДВНЗ «ДонНТУ», 2025. – 18 с.

Методичні вказівки до виконання практичних занять з дисципліни «Методи та засоби обробки текстової інформації» спрямовані на систематизацію теоретичних та практичних знань з основ застосування сучасних методів і засобів обробки текстової інформації природної мови та містять завдання із зазначенням 20 варіантів, основні вимоги до формування звіту та контрольні питання. Завдання виконуються студентами у відповідності до заданого варіанту.

Укладач: Тетяна АЛТУХОВА, к.т.н, доцент кафедри прикладної математики та інформатики ДВНЗ «ДонНТУ»

Рецензент: Людмила СЕРГІЄНКО, канд. пед. наук, доц., завідувачка кафедри вищої математики та фізики ДВНЗ «ДонНТУ»

Відповідальний за випуск: Ярослав ДОРОГИЙ, д-р. техн. наук, доц., в.о. завідувача кафедри прикладної математики та інформатики ДВНЗ «ДонНТУ»

Затверджено навчально-методичним відділом ДонНТУ,
протокол № 7 від 27.05.2025 р.

Розглянуто на засіданні кафедри прикладної математики та інформатики ДВНЗ «ДонНТУ» протокол №5 від 30.04.2025 р.

© ДВНЗ «ДонНТУ», 2025

ЗМІСТ

ВСТУП	4
1. Практичне заняття. СТВОРЕННЯ КОНТЕКСТНО-ЗАЛЕЖНОЇ ГРАМАТИКИ	5
2. Практичне заняття. СТВОРЕННЯ ТРАНСФОРМАЦІЙНОЇ ПОРОДЖУЮЧОЇ ГРАМАТИКИ	8
3. Практичне заняття. СТВОРЕННЯ РОЗШИРЕНОЇ МЕРЕЖІ ПЕРЕХОДІВ	11
4. Практичне заняття. ОБЧИСЛЕННЯ ЙМОВІРНОСТІ ПРОПОЗИЦІЇ НА ПІДСТАВІ БІГРАМ І ТРИГРАМ	13
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ	17
ДОДАТОК А. ТИТУЛЬНИЙ АРКУШ	19

ВСТУП

Зараз поточний стан розвитку застосування сучасних методів та засобів обробки текстової інформації природної мови перебувають на досить високому рівні. Штучний інтелект, машинне навчання та обробка природної мови використовуються для аналізу текстових даних, автоматичної класифікації, отримання ключової інформації, машинного перекладу, тощо, тим самим дозволяючи збільшити швидкість і точність аналізу великих обсягів текстової інформації. Враховуючи це, основна **мета** виконання практичних занять з дисципліни «Методи та засоби обробки текстової інформації» є ***систематизація теоретичних положень та практичних навичок при застосування сучасних методів та засобів обробки текстової інформації з урахуванням особливостей різних видів граматик.***

Методичні вказівки дозволять студентам систематизувати отриманні знання та практичні навички з питань синтаксичного аналізу на базі БС-граматики, ТП-граматики, мережевої граматики та метода N-грам. Варіант обирається студентом за умови узгодження з керівником практичних занять або лектором. Захист практичних робіт відбувається після здачі студентом і перевірки її викладачем та супроводжується виступом з презентацією.

1. Практичне заняття. СТВОРЕННЯ КОНТЕКСТНО-ЗАЛЕЖНОЇ ГРАМАТИКИ

Мета: отримати практичні навички створення контекстно-залежної граматики в процесі дослідження представлень.

Завдання до практичного заняття:

1.1 Проаналізувати основні теоретичні положення з контекстно-залежної граматики та визначити особливості її формалізації.

1.2 Виконати генерування (підбір) в індивідуальному порядку речення з вказаною валентністю базової предикатної синтаксеми та вторинною синтаксемою відповідного типу згідно заданого варіанту (табл. 1.1).

Таблиця 1.1 – Вихідні дані

Варіант	Валентність базової предикатної синтаксеми	Тип другої (вторинної) предикатної синтаксеми	Варіант	Валентність базової предикатної синтаксеми	Тип другої (вторинної) предикатної синтаксеми
1	8	атрибутивна	11	4	атрибутивна
2	7	модальна	12	3	модальна
3	6	адвербіальна	13	8	адвербіальна
4	5	атрибутивна	14	7	атрибутивна
5	4	модальна	15	6	модальна
6	3	адвербіальна	16	5	адвербіальна
7	8	атрибутивна	17	4	атрибутивна
8	7	модальна	18	3	модальна
9	6	адвербіальна	19	8	адвербіальна
10	5	атрибутивна	20	7	атрибутивна

1.3 Здійснити ручне трасування відповідної мовної конструкторії на базі контекстно-залежної граматики. Виконати формування опису синтаксичної структури мовної конструкторії шляхом формування дерева складових речення та здійснити відповідне розмічення.

1.4 Виконати обробку отриманих результатів та надати графічне представлення.

Вимоги до змісту звіту з практичного заняття:

1. Титульний аркуш (додаток А)
2. Тема, мета, завдання практичної роботи із зазначенням варіанту.
3. Навести короткий опис теоретичних положень сутності контекстно-залежної граматики з визначеними особливостями її формалізації.
4. Надати опис підбору речення за заданих умов.
5. Надати побудоване дерево складових для сформованого речення в загальному вигляді та розмічене дерево у відповідності до контекстно-залежної граматики.
6. Побудувати дерево синтаксичного підпорядкування для сформованого речення.
7. Навести результати аналізу отриманих результатів роботи.
8. Надати загальні висновки за результатами зіставлення теоретичних положень і отриманих у практичній роботі експериментальних результатів, оцінювання ступеню досягнення мети та виконання поставлених завдань практичного заняття.

Контрольні питання:

1. Дайте визначення синтаксичному аналізу.
2. Що собою представляє синтаксична структура?
3. Поясніть терміни «девіація» та «синтаксична девіація».
4. Охарактеризуйте ядерну конструкцію. Наведіть приклади.
5. Що таке синтаксичний зв'язок? Які види його існують?
6. Які види підпорядкування зв'язку у словосполученнях існують? Надайте коротку характеристику.
7. Що називають субстантивованою частиною мови? Наведіть приклади.
8. Що таке знаменні частини мови? Надайте приклади.

9. Які синтаксичні відношення існують?
10. Визначте непередикативні синтаксичні відношення на рівні словосполучень.
11. Визначте, що є результатом синтаксичного аналізу?
12. На що спирається більшість моделей подання синтаксичних структур?
13. Що собою представляє граматику залежностей?
14. Яка основна властивість проєктивності синтаксичного дерева?
15. Визначте основні властивості дерева підпорядкування.
16. Визначте переваги та недоліки граматики залежностей.
17. Охарактеризуйте граматику складових.
18. Надайте класифікацію груп складових.
19. Які фразові категорії ви знаєте?
20. Яким чином здійснюється виявлення складових?
21. Які особливості має дерево складових?
22. Що називають синтаксисом?
23. Дайте визначення синтаксеми.
24. Визначте базовий склад синтаксеми.
25. Що собою представляє базова передикатна синтаксема?
26. Охарактеризуйте валентність базової передикатної синтаксеми.
27. Які різновиди базової передикатних синтаксем ви знаєте?
28. Назвіть різновиди передикатів та надайте їх коротку характеристику.
29. Що собою представляє базова субстанціональна синтаксема (актант)?
30. Охарактеризуйте вторинні субстанціональні синтаксеми та наведіть їх різновиди.
31. Які бувають вторинні передикатні синтаксеми?
32. Коротко охарактеризуйте адвербіальні синтаксеми.
33. Надайте коротку характеристику модальних синтаксем.
34. Що собою представляють атрибутивні синтаксеми?
35. Які методи аналізу мовних конструкцій існують?
36. Охарактеризуйте детерміністський підхід синтаксичного аналізу.

37. Які типи формальних граматик існують? Надайте коротку їх характеристику.

38. Надайте алгоритм контекстно-залежної граматики.

Література [2-5, 7-11, 13, 14]

2. Практичне заняття. СТВОРЕННЯ ТРАНСФОРМАЦІЙНОЇ ПОРОДЖУЮЧОЇ ГРАМАТИКИ

Мета: отримати практичні навички створення трансформаційної породжуючої граматики в процесі дослідження представлень.

Завдання до практичного заняття:

2.1 Проаналізувати теоретичні положення з трансформаційної породжуючої граматики та визначити особливості її формалізації.

2.2 Здійснити ручне трасування відповідної мовної контрукції на базі трансформаційної породжуючої граматики у відповідності до варіанту (табл. 2.1).

Таблиця 2.1 – Вихідні дані

Варіант	Речення	Варіант	Речення
1	Видатний науковець застосував новітню технологію моделювання	11	Молодий аспірант опублікував фахову статтю
2	Висококваліфікований програміст виявив помилку в коді	12	Військовий експерт охарактеризував поточний стан подій
3	Військовий запропонував застосувати нову стратегію	13	Старанний студент виконав завдання вчасно
4	Суддя залучив експертів для проведення експертизи	14	Класний керівник викликав батьків учня
5	Керівник відділу звернув увагу на помилки в документі	15	Провідний програміст розробив початковий прототип

Варіант	Речення	Варіант	Речення
6	Програміст розробив новий мобільний застосунок	16	Технічний інженер використав новітній підхід
7	Викладач здійснив оцінювання знань після лекції	17	Український хореограф поставив найкращий танець
8	Класний керівник був незадоволений поведінкою учнів	18	Закордонний співак презентував ритмічний хіт
9	Український співак виступив на головній сцені міста	19	Відомий актор відіграв чудову роль
10	Талановитий митець створив неперевершений шедевр	20	Куплена статуетка буде гарним експонатом

2.3 Виконати обробку отриманих результатів та надати графічне представлення.

Вимоги до змісту звіту з практичного заняття:

1. Титульний аркуш (додаток А)
2. Тема, мета, завдання практичної роботи із зазначенням варіанту.
3. Навести короткий опис теоретичних положень сутності трансформаційної породжуючої граматики з визначеними особливостями її формалізації.
4. Надати результати опрацювання мовної конструкції.
5. Навести результати аналізу закономірностей.
6. Надати загальні висновки за результатами зіставлення теоретичних положень і отриманих у практичній роботі експериментальних результатів, оцінювання ступеню досягнення мети та виконання поставлених завдань практичного заняття.

Контрольні питання:

1. Що собою представляє трансформаційна породжуюча граматика?

2. З яких компонентів складається трансформаційна породжуюча граматики?

3. Якому компоненту відводиться центральна роль в трансформаційній породжуючій граматиці?

4. Що забезпечує трансформаційний субкомпонент?

5. Що собою представляє глибинна структура простого речення?

6. Охарактеризуйте категоріальні правила.

7. За допомогою яких правил виконується постановка термінальних символів в породжуванні речення?

8. Наведіть основні ознаки іменників.

9. Надайте коротку характеристику власних іменників. Наведіть приклади.

10. Охарактеризуйте одухотворенні іменники та наведіть декілька прикладів.

11. Дайте характеристику ознаці іменника «обчислювальність».

12. Які ознаки дієслова є?

13. Що собою представляють перехідні дієслова?

14. Охарактеризуйте види дієслів.

15. Визначте основні форми дієслів.

16. Визначте, що собою представляють зворотні, незворотні та взаємні дієслова?

17. Визначте основні ознаки прикметників.

18. Охарактеризуйте прикметники відносно розряду за значенням.

19. Надати характеристику прикметників відносно ступенів.

20. Які форми прикметників є?

21. Визначте основні правила лексичного включення. Наведіть приклади лексики трансформаційної породжуючої граматики.

22. Визначте основні цілі методики аналізу складових.

23. Які базові концепти відносяться до безпосередньо складових?

24. На яких вихідних постулатах ґрунтується методика аналізу безпосередньо складових?

25. Визначте основні процедурні правила методики.

26. Визначте переваги та недоліки методики аналізу безпосередньо складових.

Література [2-5, 7-11, 13, 14]

3. Практичне заняття. СТВОРЕННЯ РОЗШИРЕНОЇ МЕРЕЖІ ПЕРЕХОДІВ

Мета: отримати практичні навички зі створення розширеної мережі переходів в процесі дослідження представлень.

Завдання до практичного заняття:

3.1 Проаналізувати теоретичні положення щодо сутності та методів досліджень розширеної мережі переходів.

3.2 Здійснити ручне трасування відповідної мовної контрукції на базі розширеної мережі переходів у відповідності до варіанту (табл. 3.1). Розширена мережа переходів, яка досліджується, представляє трьохступінчасту процедуру (створення множини станів, створення множини дуг з мітками лексичних категорій, створення множини дуг з мітками найменування станів).

3.3 Виконати обробку та аналіз отриманих результатів.

Таблиця 3.1 – Вихідні дані

Варіант	Речення	Варіант	Речення
1	Чи повинен студент публікувати тези?	11	Чи бажає замовник затверджувати проєкт?
2	Чи хоче автор робити доповідь?	12	Чи зможе пацієнт виконувати призначення?
3	Чи планує тренер проводити інструктаж?	13	Чи бажає студент формувати звіт?
4	Чи бажає клієнт замовляти товар?	14	Чи зможе клієнт надати відгук?

Варіант	Речення	Варіант	Речення
5	Чи хоче керівник перевіряти звітність?	15	Чи буде блогер коментувати відео?
6	Чи може архітектор досліджувати макет?	16	Чи зможе каскадер застосовувати екіпіровку?
7	Чи буде вчитель видавати завдання?	17	Чи може спікер розпочати діалог?
8	Чи планує хореограф ставити танок?	18	Чи буде викладач оцінювати завдання?
9	Чи мріє співак виконати сингл?	19	Чи зможе аспірант виступити на конференції?
10	Чи може інженер перевіряти план?	20	Чи зможе програміст розробити програму захисту?

Вимоги до змісту звіту з практичного заняття:

1. Титульний аркуш (додаток А)
2. Тема, мета, завдання практичної роботи із зазначенням варіанту.
3. Навести короткий опис теоретичних положень та сутності розширеної мережі переходів.
4. Надати результати опрацювання мовної конструкції.
5. Навести результати аналізу закономірностей.
6. Надати загальні висновки за результатами зіставлення теоретичних положень і отриманих у практичній роботі експериментальних результатів, оцінювання ступеню досягнення мети та виконання поставлених завдань практичного заняття.

Контрольні питання:

1. Охарактеризуйте мережу переходів.
2. Визначте сутність граматики розширених мереж переходів.
3. Що собою представляє розширена мережа переходів?
4. В якому вигляді відбувається подання мережі переході граматики?
5. Які стани виділяються в мережі переходів?
6. Які мітки зазначаються на дугах мережевого графу переходів?

7. Що необхідно додати до кожної дуги, щоб мережа переходів здійснювала трансформаційні перетворення?
8. Визначте структуру опису граматики за допомогою нормальної форми Бекуса.
9. Які типи дуг використовуються в мережевому графі переходів?
10. Які дії присвоює регістру значення форми?
11. Визначте основні типи переходів.
12. Наведіть основні типи форм.
13. Яким чином відбувається опис форм.
14. Скільки види регістрів виділяють в мережі переходів?
15. Визначте переваги та недоліки мережі переходів.

Література [1-9, 12-14]

4. Практичне заняття. ОБЧИСЛЕННЯ ЙМОВІРНОСТІ ПРОПОЗИЦІЇ НА ПІДСТАВІ БІГРАМ І ТРИГРАМ

Мета: отримати практичні навички обчислення ймовірності речень на базі біграм та триграм в процесі дослідження представлень за умови заданої мовної конструкції.

Завдання до практичного заняття:

4.1 Проаналізувати особливості застосування N-грам на базі біграм і триграм в процесі обробки текстової інформації природною мовою.

4.2 Побудувати біграмні та триграмні мовні моделі для зазначеного варіанту мовної конструкції (табл. 4.1). Надати опис та детальне роз'яснення стосовно процесу обчислення ймовірності речення на базі біграм і триграм, який ґрунтується на чотирьохступінчатій процедурі (обчислення ймовірностей

уніграм, обчислення ймовірностей біграм, обчислення ймовірностей триграм, обчислення ймовірностей речення).

4.3 Виконати аналіз отриманих результатів.

Таблиця 4.1 – Вихідні дані

Варіант	Речення	Варіант	Речення
1	Необхідно забрати посилки з пошти	11	Із соку ягоди готують сироп
2	Осінній день схожий на сутінки	12	Незабудка є окрасою весняної флори
3	Могутнім оберегом завжди вважали вінок	13	Щось має статись у світі
4	Нехай ніколи не зраджує пам'ять	14	Музикальність є незбагненною душею мови
5	Пам'ять логічна надає людині індивідуальності	15	Милозвучність є окрасою української мови
6	Зображення кола було магічним знаком	16	Промовлене слово може змінити все
7	Пам'ять окремої людини працює інакше	17	Філологія об'єднує мовознавство та літературознавство
8	Одяг був для людини оберегом	18	Потрібно підкинути дрова в камін
9	Білі плоскі суцвіття прикрашатимуть кущі	19	Характеристики системи можуть бути змінені розробниками
10	В листопаді найменше сонячних днів	20	Зимова пора може бути сонячною та холодною

Вимоги до змісту звіту з практичного заняття:

1. Титульний аркуш (додаток А)
2. Тема, мета, завдання практичної роботи із зазначенням варіанту.
3. Навести опис особливостей застосування N-грамної моделі мови на прикладі біграм та триграм в обробці природомовних текстів.
4. Надати результати застосування біграмної та триграмної мовної моделі із детальним описом головних аспектів виконання процесу обчислення ймовірності речення.

5. Надати загальні висновки за результатами зіставлення теоретичних положень і отриманих у практичній роботі експериментальних результатів, оцінювання ступеню досягнення мети та виконання поставлених завдань практичного заняття.

Контрольні питання:

1. Надайте характеристику N-грам.
2. Що називають N-грамою?
3. Які види N-грам існують?
4. Для чого використовують N-грамні моделі при обробці текстової інформації природної мови?
5. Охарактеризуйте ймовірнісний підхід, що ґрунтується на N-грамах.
6. Визначте мету побудови N-грамних моделей.
7. Що собою представляють біграма?
8. Поясніть, що представляє собою триграма?
9. Визначте переваги та недоліки застосування біграм та триграм.
10. Для чого використовуються N-грамні моделі в сфері обробки природної мови?
11. Наведіть алгоритм побудови біграмної мовної моделі та обчислення складових ймовірностей.
12. Поясніть алгоритм побудови триграмної мовної моделі та обчислення складових ймовірностей.

Література [1-9, 12-15]

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Інтелектуальна обробка текстів: навч. посіб. / В. Ю. Тарануха. – Київ : КНУ, 2014. – 80 с.
2. Вавіленкова А. І. Теоретичні основи аналізу електронних текстів: монографія / А. І. Вавіленкова. – Київ : ТОВ «СІК ГРУПІ УКРАЇНА», 2016. – 192 с.
3. ChengXiang Zhai Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining / Zhai ChengXiang, Sean Massung. – ACM Books, 2016. – 532 p.
4. Silge Julia Text Mining with R: A Tidy Approach / Julia Silge, David Robinson. – O'Reilly Media, 2017. – 191 p.
5. Ignatow Gabe An Introduction to Text Mining: Research Design, Data Collection, and Analysis / Gabe Ignatow, Rada F. Mihalcea. – SAGE Publications, Inc, 2017. – 346 p.
6. Goldberg Y. Neural Network Methods for Natural Language Processing / Y. Goldberg. – Morgan & Claypool Publishers, 2017. – 309 p.
7. Конспект лекцій з дисципліни «Сучасні методи обробки природно-мовних текстів» для студентів спеціальності 122 Комп'ютерні науки денної та заочної форми навчання [Електронний ресурс] / Є. Є. Федоров, І. В. Ярош, Т. О. Черняк. – Покровськ: ДВНЗ «ДонНТУ», 2018. – 60 с. – Режим доступу: <http://ea.donntu.edu.ua/jspui/handle/123456789/29087>. - Назва з екрана.
8. Прикладна лінгвістика: навч. посіб. / Н. А. Цимбал. – Умань: Візаві, 2019. – 106 с.
9. Zong Chengqing Text Data Mining / Chengqing Zong, Rui Xia, Jiajun Zhang. – Kindle Edition, 2021. – 363 p.
10. Гавриленко С. Ю. Формальні мови, граматики та автомати: навч. посіб. / С. Ю. Гавриленко. – Харків : НТУ «ХПІ», 2021. – 133 с.
11. Lamba Manika Text Mining for Information Professionals: An Uncharted Territory / Manika Lamba, Margam Madhusudhan. – Springer, 2022. – 372 p.

12. George Alexandra Python Text Mining: Perform Text Processing, Word Embedding, Text Classification and Machine Translation / Alexandra George. - BPB Publications, 2022. –320 p.

13. Getting started with NLP [Електронний ресурс]. – Режим доступу: <https://surl.lu/ijsyga>. - Назва з екрана.

14. Технології обробки текстової обробки інформації [Електронний ресурс]. – Режим доступу: <https://surli.cc/ugbsku>. - Назва з екрана.

15. N-грами в НЛЛ: як вони працюють та їхня роль в аналізі тексту [Електронний ресурс]. – Режим доступу: <https://surl.lu/yliwcl>. - Назва з екрана.

ТИТУЛЬНИЙ АРКУШ

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ»
Кафедра прикладної математики та інформатики

ЗВІТ З ПРАКТИЧНОЇ РОБОТИ № ____

з дисципліни «МЕТОДИ ТА ЗАСОБИ ОБРОБКИ ТЕКСТОВОЇ ІНФОРМАЦІЇ»

Варіант ____

Виконав (ла) студент (ка) **1м** курсу групи **КНм-____**

Спеціальності **Ф3 Комп'ютерні науки**

Власне ім'я та ПРІЗВИЩЕ _____

« ____ » _____ 20__ р. *(підпис)*

Перевірив: ступінь, вчене звання, посада

Власне ім'я та ПРІЗВИЩЕ _____

« ____ » _____ 20__ р. *(підпис)*

Дрогобич – 20__ р.