МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ І СИСТЕМ

**МЕТОДИЧНІ РЕКОМЕНДАЦІЇ**

**до самостійної роботи**

з дисципліни «Технології обробки природомовної інформації»

для здобувачів освітнього ступеня «бакалавр»

зі спеціальності 126 Інформаційні системи та технології

(освітньої програми «Web-технології, Web-дизайн»)

усіх форм навчання

Черкаси

2020

|  |  |
| --- | --- |
| УДК 004.934(07) М 54 | *Затверджено вченою радою ФІТІС,**протокол № 5 від 17.02.2020 р.,**згідно з рішенням кафедри* *інформаційних технологій проектування,**протокол № 8 від 10.01.2020 р.* |

|  |  |
| --- | --- |
| Упорядник: | Тарасенко Я. В., к.т.н. |
| Рецензент: | Федотова-Півень І. М., к.т.н., доцент |
| М 54 |  **Методичні** рекомендації до самостійної роботи з дисципліни «Технології обробки природо мовної інформації» для здобувачів освітнього ступеня «бакалавр» зі спеціальності № 126 «Інформаційні системи та технології» (освітня програма «Web-технології, Web-дизайн») усіх форм навчання [Електронний ресурс] / [Упоряд.: Я.В. Тарасенко]; М-во освіти і науки України, Черкас. держ. технол. ун-т. – Черкаси : ЧДТУ, 2020. – 15 с. – Назва з титульного екрана. |

Методичні рекомендації містять основні теоретичні відомості відповідної теми, які закріплюються та поглиблюються самостійною роботою, перелік запитань та завдань для самостійного опрацювання, перелік літератури, яку здобувачі можуть використовувати при вивченні дисципліни «Технології обробки природо мовної інформації» та методичні вказівки для виконання завдань самостійного опрацювання. Особлива увага приділяється поглибленню навичок застосування алгоритмів пошуку в тексті, проведення автоматизованого реферування текстової інформації, застосування та створення засобів машинного перекладу, проектування та реалізації систем природо мовного діалогу.

Для здобувачів освітнього ступеня бакалавр спеціальності № 126 «Інформаційні системи та технології» (освітньої програми «Web-технології, Web-дизайн») усіх форм навчання.

УДК 004.934(07)

Виробничо-практичне

електронне видання

комбінованого використовування

**МЕТОДИЧНІ РЕКОМЕНДАЦІЇ**

**до самостійної роботи**

з дисципліни «Технології обробки природо мовної інформації»

для здобувачів освітнього ступеня «бакалавр»

зі спеціальності 126 Інформаційні системи та технології

(освітньої програми «Web-технології, Web-дизайн»)

усіх форм навчання

Упорядник **Тарасенко** Ярослав Володимирович

*В авторській редакції.*

ЗМІСТ

[ВСТУП 4](#_Toc33286418)

[Порядок виконання роботи та звітність 5](#_Toc33286419)

[Самостійна робота № 1.](#_Toc33286420) [Технології обробки природо мовної інформації в інформаційній безпеці 6](#_Toc33286421)

[Самостійна робота № 2.](#_Toc33286422) [Лямбда-числення в обробці природо мовної інформації 7](#_Toc33286423)

[Самостійна робота № 3.](#_Toc33286424) [Схожі та відмінні риси формальної граматики мови програмування та природньої мови 8](#_Toc33286425)

[Самостійна робота № 4.](#_Toc33286426) [В Види семантик та способи їх моделювання 9](#_Toc33286427)

[Самостійна робота № 5.](#_Toc33286428) [Основи теорії інтертекстуальності 10](#_Toc33286429)

[Самостійна робота № 6.](#_Toc33286430) [Системи смислового аналізу Text Mining в машинному перекладі 11](#_Toc33286431)

[Самостійна робота № 7.](#_Toc33286432) [Лематизація як етап автоматизованого реферування 12](#_Toc33286433)

[Самостійна робота № 8.](#_Toc33286434) [Словникове забезпечення діалогових систем 13](#_Toc33286435)

[ЛІТЕРАТУРА 14](#_Toc33286436)

# ВСТУП

Навчальна дисципліна «Технології обробки природо мовної інформації» належить до циклу дисциплін професійної підготовки і має міждисциплінарні зв’язки з такими дисциплінами, як «Алгоритмізація та програмування», «Об’єктно-орієнтоване програмування», «Бази даних та знань», «Теорія ймовірності та математична статистика», «Методи та системи штучного інтелекту», «Інтелектуальний аналіз даних», «Людино-машинна взаємодія», Прикладні інтелектуальні системи обробки даних», «Технології та системи підтримки прийняття рішень».

Програма навчальної дисципліни складається з двох частин:

1. Основні моделі методи та засоби комп’ютерної обробки природо мовної інформації.

2. Прикладне застосування методики автоматизованої обробки природомовних об’єктів.

Методичні рекомендації налічують опис процесу самостійного опрацювання 8 тем з уточненням мети роботи, основних теоретичних відомостей відповідної теми, які закріплюються та поглиблюються самостійною роботою, переліку запитань та завдань для самостійного опрацювання, переліку літератури, яку здобувачі можуть використовувати при вивченні дисципліни та виконанні завдань самостійного опрацювання.

Метою самостійної роботи є закріплення теоретичних основ автоматизованого аналізу та обробки природньої мови за допомогою інформаційних систем, а також поглиблення практичних навичок обрання і використання математичних методів та технічних засобів обробки природо мовної інформації, застосування алгоритмів пошуку в тексті, проведення автоматизованого реферування текстової інформації, застосування та створення засобів машинного перекладу, систем природо мовного діалогу.

Самостійна робота спроектована таким чином, що по завершенню курсу, здобувачі вищої освіти закріплюють набуті компетенції, які формують здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші), здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень, застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв’язанні задач проектування і використання інформаційних систем та технологій.

Крім того, матеріали методичних вказівок до самостійної роботи можуть бути використані при виконання лабораторних робіт та підготовці до лекційних занять.

При проектуванні самостійної роботи використані такі новітні методи навчання, як модульне та проблемне навчання. При цьому, проблемне інтегроване з модульним.

# Порядок виконання роботи та звітність

Самостійна робота має на меті самостійне поглиблене опрацювання здобувачем вищої освіти тем лекційного курсу використовуючи пропоновану літературу та метод проблемного навчання. Мається на увазі, що в ході роботи здобувач буде зустрічати певні незрозумілі на даному етапі питання. Ці питання формують певну проблематику, обговорення якої під час лекційних занять та лабораторних робіт прибере прогалини у загальних знаннях здобувача із курсу та допоможе у набутті ним відповідних компетенцій, на що розрахований курс.

Самостійна робота проводиться згідно з навантаженням та тематикою, що висвітлені у силабусі та робочій програмі дисципліни. Вона являє собою організовану викладачем діяльність здобувача вищої освіти, направлену на поглиблення набутих умінь та навичок здобувача.

Робота може проводитись у будь-якому приміщенні, обладнаному комп’ютерною технікою та доступом в Інтернет (бібліотека, навчальні аудиторії, домашні умови).

Самостійна робота складається з двох частин:

1) пошук відповіді на запитання, що виносяться на самостійне опрацювання шляхом опрацювання літератури, перегляд пропонованих у силабусі відео-лекцій та електронних ресурсів;

2) вирішення конкретної практичної задачі.

Контроль виконання самостійної роботи здійснюється у формі співбесіди, де розглядаються результати виконання практичного завдання та рівень засвоєння здобувачем теоретичних знань.

Критерії успішного самостійного опрацювання матеріалу:

1) глибокі знання пропонованої літератури;

2) опрацювання додаткових джерел та матеріалів;

3) вирішення сформованої в ході роботи проблематики;

4) набуті практичні навички;

5) системність отриманих знань та інтегрованість у лекційний курс.

За кожною темою самостійної роботи коротко наводиться зв'язок із темою лекційного завдання та наводяться рекомендації з напрямку та розширеної тематики лекційного курсу, на які слід звернути особливу увагу в ході виконання завдання, а також при підготовці до лекційного заняття та виконання лабораторної роботи. Практичне завдання направлене на доповнення навичок, здобутих під час виконання лабораторних робіт, а також розширення навичок, шляхом виконання завдань, аналогічних лабораторним роботам, проте, які відрізняються особливими умовами, оточенням, основною метою чи нестандартним підходом до вирішення поставленої перед здобувачем вищої освіти задачі.

# Самостійна робота № 1

# Технології обробки природо мовної інформації в інформаційній безпеці

**Мета роботи** – ознайомитися з можливостями обробки природньої мови для захисту інформації. Розширити навички використання інструментальних програмних засобів обробки тексту природньої мови. Поглибити знання сфер застосування дисципліни.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №1 «Базові принципи комп’ютерної обробки природньої мови» має на меті розкриття основних сфер застосування технологій обробки природо мовної інформації. До таких сфер відносяться комп’ютерні засоби автоматизованого перекладу, реферування тексту природньої мови, технології пошуку у тексті та засоби природо мовного діалогу. На самостійне опрацювання виноситься питання використання технологій аналізу та модифікації тексту природньої мови для забезпечення захисту текстової інформації (авторські права, захист від зміни тексту тощо).

Для поглиблення знань в області комп’ютерної лінгвістичної стеганографії (науки приховування повідомлення у текстовій інформації) слід розглянути основні сучасні методи лінгвістичної стеганографії. До них відносяться методи довільного інтервалу, синтаксичні та семантичні методи та шляхи автоматизованого вбудовування цифрового водяного знаку в текст. Крім того, слід розглянути підходи автоматизованого стегоаналізу, зокрема структуру та особливості лінгвістичного процесору.

**Запитання для самостійного опрацювання:**

1. Які існують методи лінгвістичної стеганографії та стегоаналізу?
2. Які математичні моделі лежать в основі методів лінгвістичної стеганографії та стегоаналізу?
3. Наведіть користь комп’ютерної стеганографії для захисту інформації.
4. В чому суть лінгвістичного процесору?
5. Перелічіть сучасні програмні інструменти для приховування цифрового водяного знаку у тексті.

**Практичне завдання для самостійної роботи:**

Освоїти роботу з web-орієнтованим комп’ютерним інструментом приховування та розшифрування прихованого повідомлення у тексті на основі Unicode символів. Доступ до інструменту можливо отримати за посиланням: <http://www.irongeek.com/i.php?page=security/unicode-steganography-homoglyph-encoder>

# Самостійна робота № 2

# Лямбда-числення в обробці природо мовної інформації

**Мета роботи –** ознайомитися з лямбда-числення в обробці природньої мови. Розширити навички аналізу семантичного значення речення. Поглибити знання методики реалізації лямбда числень за допомогою однієї з функціональних мов програмування.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №2 «Математичні методи аналізу природньої мови» має на меті розкриття основних аспектів використання статистичних моделей та дискретного аналізу природніх мов, методики морфологічно-синтаксичного та семантичного аналізу природньої мови, що проводяться засобами PHP та Python. Для цього описуються математичні методи формалізації процесу дослідження природньої мови. На самостійне опрацювання виноситься питання лямбда-обрахунків при обробці тексту природньої мови засобами функціональнх мов програмування.

Для поглиблення знань в області лямбда-числення слід розглянути особливості класичного та розширеного лямбда-числення та формування лямбда-словника для задач формальної семантики. Крім того, необхідно виявити закономірності в трактуванні лямбда-формули як смислу речення. Для комп’ютеризації формалізованих обрахунків за допомогою лямбда-виразів слід звернутися до функціонального підходу програмування, що оснований на лямбда-численні. Рекомендується використовувати мову програмування Haskell, зокрема, звернути увагу на рекурсивне та чисте лямбда числення.

**Запитання для самостійного опрацювання:**

1. Яким чином лямбда-числення застосовують в дослідженні семантики?
2. Що саме являє собою лямбда словник?
3. Які особливості лямбда-формули, яка відповідає певному слову?
4. Яким чином відбувається представлення значення речення з допомогою формули лямбда-обрахунку.
5. Опишіть процес використання четвертого перетворення для визначення екстенсіоналу виразу?

**Практичне завдання для самостійної роботи:**

Використовуючи одну з функціональних мов програмування (рекомендовано Haskell), чи будь-яку іншу, з допомогою якої можливо реалізувати функціональний підхід програмування, на основі лямбда-обрахунків проаналізувати семантичне значення простого речення «Кожен студент навчається».

# Самостійна робота № 3

# Схожі та відмінні риси формальної граматики мови програмування та природньої мови

**Мета роботи –** ознайомитися з формальною граматикою в світлі мов програмування. Розширити навички застосування формальних граматик природніх мов. Поглибити знання особливостей представлення правил формальної граматики.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №3 «Формальна граматика природньої мови» має на меті розкриття основних аспектів формальної граматики природньої мови. Для цього описуються особливості та пояснюються властивості контекстно-вільної та контекстно-залежної граматики, генеративної граматики та граматики зв’язків. На самостійне опрацювання виноситься питання схожості формальної граматики природньої мови та мови програмування для закріплення розуміння лекційної теми.

Для поглиблення знань в області формальної граматики природніх мов слід особливу увагу звернути на контекстно-вільні граматики, як такі, що використовуються для опису мов програмування. Необхідно також розглянути процес виведення правил контекстно-вільних граматик із правил контекстно-залежних граматик. Крім того, слід також звернути увагу на спеціальні правила співвіднесення (аналог виключення) в мові програмування C++ та порівняти з подібними правилами для природньої мови.

**Запитання для самостійного опрацювання:**

1. Як можна класифікувати контекстно-вільні та залежні граматики за Хомським? Які ще існують класи, не розглянуті на лекції?
2. Яким чином описуються виключення у формальній граматиці мов програмування?
3. Що описують форми Бекуса-Науера?
4. Чи можливо використання скінченного автомату при роботі з формальною граматикою природньої мови та мови програмування?
5. В чому схожість правил синтаксису природньої мови та мови програмування?

**Практичне завдання для самостійної роботи:**

Побудувати орієнтований граф, що представляє формальну граматику висловлювання природньою мовою та блоку команд будь-якої комп’ютерної програми. Порівняти отримані графи між собою, виявити відмінності та спільні риси.

# Самостійна робота № 4

# Види семантик та способи їх моделювання

**Мета роботи –** ознайомитися видами семантик. Розширити навички використання інтенсіональної логіки для дослідження синонімічних семантичних значень у реченні природньої мови. Поглибити знання підходів моделювання семантик.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №4 «Формальна семантика природньої мови» має на меті розкриття основних аспектів формалізації семантики природньої мови та дослідження смислу текстової інформації. Для цього описуються особливості онтологічного підходу та методи формальної семантики, розкривається особливість дискурсного аналізу, семантики Монтегю та інтенсіональної логіки. На самостійне опрацювання виноситься питання семантики можливих світів та видів семантик та інтенсіональних логік, їх особливостей та спільних рис для кращого розуміння процесу формалізації семантики природньої мови.

Для поглиблення знань в області формальної семантики природніх мов слід особливу увагу звернути на інтенсіональні контексти та семантику можливих світів. Необхідно окрім формальної семантики засвоїти лексичну, когнітивну, синтаксичну та комп’ютерну семантику. Слід звернутися, окрім онтологічного підходу моделювання семантики, розглянутого в ході лекційного заняття також до концептоцентричної, синтактикоцентричної та антропоцентричної теорії.

**Запитання для самостійного опрацювання:**

1. Що являє собою семантика можливих світів?
2. Що лежить в основі лексичної, когнітивної, синтаксичної та комп’ютерної семантик?
3. Що таке формально-аксіологічна семантика та яким чином вона моделюється?
4. На чому ґрунтуються концептоцентрична, синтактикоцентрична та антропоцентрична теорії моделювання семантик?
5. Перелічіть види інтенсіональних логік.

**Практичне завдання для самостійної роботи:**

Використовуючи знання про види семантик та способи їх моделювання, базуючись на інтенсіональній логіці дослідити значення речення «Федір розуміє, що фотографія це світлина»

# Самостійна робота № 5

# Основи теорії інтертекстуальності

**Мета** – ознайомитися з теорією інтертекстуальності. Розширити навички використання семантичного пошуку у тексті для здійснення між текстового пошуку. Поглибити знання технологій та інструментальних засобів класичного лінгвістичного аналізу текстової інформації.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №5 «Алгоритми пошуку в тексті» має на меті розкриття основних аспектів організації пошуку по тексту природньої мови. Для цього описується особливості прямого пошуку, алгоритми Кнута, Моріса і Прата, Бойєра-Мура та пояснюються основи семантичного пошуку. На самостійне опрацювання виноситься питання теорії інтертекстуальності для пояснення зв’язків та посилань одного тексту до іншого, що поглиблює знання пошуку у тексті в напрямку між текстового пошуку та розуміння відношення між текстами за принципом відношень в рамках тексту.

Для поглиблення знань в області пошуку у тексті, зокрема семантичного пошуку, слід особливу увагу звернути на семіотизацію текстової інформації, знаковий взаємозв’язок із аналогічними знаками в іншому тексті та проблему пошуку цих знаків. Слід засвоїти основні типи прецедентних текстів, особливості формування матриці тексту. Виявити особливості методу концептуального моделювання, методу асоціацій та традиційного лінгвістичного аналізу.

**Запитання для самостійного опрацювання:**

1. Яким чином формувалась теорія інтертексту?
2. Яким чином літературний вплив та запозичення зумовлює між текстовий пошук першоджерела?
3. Які цілі та завдання інтертекстуальності та яку роль відіграє семіотизм?
4. Завдяки чому можливі та як реалізуються інтертекстуальні взаємодії?
5. На чому ґрунтуються методи теорії інтертекстуальності?

**Практичне завдання для самостійної роботи:**

Дослідити уривок тесту (взятий із теоретичних відомостей до поточної самостійної роботи) на зв’язок з іншими текстами за допомогою автоматизованих засобів виявлення плагіату. Провести семантичний пошук та виявити зв’язки з іншими текстами. Порівняти отримані результати.

# Самостійна робота № 6

# Системи смислового аналізу Text Mining в машинному перекладі

**Мета** – ознайомитися з інтелектуальним аналізом текстової інформації. Розширити навички використання спеціалізованих програмних засобів дослідження текстів природньої мови. Поглибити знання технологій організації машинного навчання для автоматизованого перекладу.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №6 «Проектування, розробка та використання засобів реалізації машинного перекладу» має на меті розкриття основних аспектів налаштування та реалізації спеціалізованих інформаційних систем для перекладу текстів. Для цього описуються особливості статистичного та гібридного методів проектування інформаційних систем машинного перекладу. Класифікуються сучасні програмні засоби перекладу та аналізуються підходи машинного перекладу. На самостійне опрацювання виноситься питання використання систем смислового інтелектуального аналізу тексту Text Mining для збору інформації та організації машинного перекладу.

Для поглиблення знань в області машинного перекладу тексту, слід особливу увагу звернути на інструменти та засоби забезпечення процесу інтелектуального аналізу даних. Необхідно виявити різницю в способі зберігання інформації за допомогою баз даних, на відміну від корпусу текстів. Слід навчитися обирати інструментальний засіб (серед таких, як Gate, Knime, Orange, RapidMiner, LPU та інших) для Text Mining та використовувати обраний інструмент.

**Запитання для самостійного опрацювання:**

1. Яким чином організовано зберігання та обробку даних в системах Text Mining?
2. Які обрати інструментальний засіб інтелектуального аналізу тексту?
3. В чому схожість Text Mining та Data Miming?
4. Які елементи інформаційного пошуку застосовуються при виокремленні понять в Text Mining для організації машинного перекладу?
5. Як відбувається тематична індексація у виявленні значень окремих лексичних одиниць?

**Практичне завдання для самостійної роботи:**

Використовуючи програмний засіб Gate провести інтелектуальний аналіз тексту українською мовою для використання результатів при машинному навчанні систем автоматизованого перекладу. Надати результати інтелектуального аналізу тексту.

#

# Самостійна робота № 7

# Лематизація як етап автоматизованого реферування

**Мета** – ознайомитися з процесом лематизації. Розширити навички використання засобів та моделей морфологічного дослідження текст. Поглибити знання технологій та інструментальних засобів автоматизованого реферування текстів природньої мови.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №7 «Автоматизоване реферування текстової інформації» має на меті розкриття основних аспектів забезпечення процесу скорочення текстової інформації за допомогою комп’ютерних засобів. Для цього описуються методи, алгоритми та інструментальні засоби реферування тексту. Описуються етапи реферування та підхід визначення межі семантичного стиснення. На самостійне опрацювання виноситься питання виділення словникової форми слова, як етапу реферування в деяких методиках скорочення тексту.

Для поглиблення знань в області автоматизованого реферування, зокрема в питаннях лематизації, слід особливу увагу звернути на особливості автоматизованого перетворення слів у леми та підрахування повторення цих лем. Слід виявити закономірності у формування бази даних для організації процесу зберігання лем та доступу до них. Необхідно виявити різницю між лематизацією та стемінгом, а також сфери застосування кожної з методик. Увагу слід приділити також правилам приведення форми слова до початкового стану.

**Запитання для самостійного опрацювання:**

1. Які відмінності між лематизацією та стемінгом?
2. Яким чином словоформа приводиться до початкового стану?
3. Яке місце займає лематизація у відомих алгоритмах автоматизованого анотування та скорочення тексту природньої мови?
4. Яким чином визначається частотність повторення лем?
5. Які математичні методи лежать в основі лематизації та які програмні засоби дозволяють їх реалізувати?

**Практичне завдання для самостійної роботи:**

Реалізувати програму лематизації слів української мови, використовуючи мову програмування Python. Для цього слід встановити (якщо відсутня) бібліотеку pymorphy та освоїти її використання.

# Самостійна робота № 8

# Словникове забезпечення діалогових систем

**Мета** – ознайомитися з особливостями формування словникової компоненти діалогових систем. Розширити навички використання систем автоматизованої обробки текстів для автоматизованого формування словнику. Поглибити знання технологій та інструментальних засобів формування словникового забезпечення діалогових систем.

**Теоретичні відомості для поглибленого опрацювання**

Тема лекції №8 «Системи природо мовного діалогу» має на меті розкриття основних аспектів реалізації та використання автоматизованих систем природо мовного діалогу. Для цього описуються особливості створення корпусу текстів та його використання в системах природо мовного діалогу а також процес лінгво-комбінаторного моделювання погано формалізованих систем для формування відповідей в процесі природо мовного діалогу. На самостійне опрацювання виноситься питання словникового забезпечення діалогових систем.

Для поглиблення знань в області систем природо мовного діалогу, зокрема словникового забезпечення діалогових систем, слід особливу увагу звернути на різницю між тезаурусом та словниковою компонентою. Крім того, необхідно засвоїти вміст словникової компоненти. Слід також звернути увагу на основні етапи та особливості процесу наповнення словникової компоненти, в тому числі з використанням комп’ютерних засобів для автоматизації процесу формування словників для забезпечення автоматизованого природо мовного діалогу.

**Запитання для самостійного опрацювання:**

1. В чому полягають особливості словникової компоненти?
2. З яких кроків складається процес наповнення словникової компоненти?
3. Які дані повинен містити словник діалогової системи?
4. Як можливо використовувати лінгвістичний процесор при автоматизованому формуванні словникової компоненти?
5. Які ще процедури інформаційних технологій и спеціальні програмні засоби використовують при формуванні словникової компоненти?

**Практичне завдання для самостійної роботи:**

Обрати будь-який доступний програмний засіб автоматичного синтаксичного аналізу, що надає можливість отримати інформацію про лексичні одиниці тексту. Провести ряд експериментів та сформувати частотний словник на основі отриманих даних для системи природо мовного діалогу.

# ЛІТЕРАТУРА

1. Erkan G., Radev D. LexRank: graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*. 2004. Vol. 22, Issue 1. P. 457-479.
2. Аношин П.И. Автоматический анализ текстов. Синтаксический и семантический анализ. *Евразийский научный журнал*. 2017. № 6. С. 211-213.
3. Прокошенкова Л.П., Гецкина И.Б. Дискурсивный анализ и его роль в современной лингвистике. *Вестник ЧГУ*. 2006. № 4. С. 451-456.
4. Андреев А.В. Формальная логика как инструмент грамматического анализа. *Varietas delectans: Сборник статей к 70-летию Николая Леонидовича Сухачева* : монография. Санкт-Петербург : Нестор-История, 2012. С. 34–47.
5. Jäger G., Roger J. Formal language theory: refining the Chomsky hierarchy. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012. Vol. 367 (1598). P. 1956-1970.
6. Бабина О.И., Дюмин Н.Ю. Корпусный метод автоматического морфологического анализа флективных языков. *Вестник ЮУрГУ. Серия: Лингвистика*. 2012. № 25. С. 38-44.
7. Bird S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural language Toolkit. Sebastopol : O’Reilly Media, 2009. 504 p.
8. Герасимова И.А. Формальная граматика и интенсиональная логика. М.: ИФ РАН, 2000. 156 с.
9. Тарасенко Я.В., Півень О.Б. Особливості обчислення інформаційної ентропії тексту в умовах проведення атаки семантичним стисненням на лінгвістичну стегосистему. *Безпека інформації*. 2018. № 24(2). C. 124-129.
10. Солдатова Г.П., Татаринов А.А., Болдырихин Н.В. Основные алгоритмы поиска подстроки в строке. *Academy*. 2018. №. 5 (32). С. 8-10.
11. Тарасенко Я.В. Метод семантичного стиснення текстової інформації для протидії комп’ютерній лінгвістичній стеганографії : дис. … канд. технічних наук : 05.13.05. Черкаси, 2018. 188 с.
12. Игнатьев М. Лингво-комбинаторное моделирование плохо формализованных систем. *Информационно-управляющие системы*. 2003. № 6. C. 34-37.
13. Станкевич Л.А. Когнитивные системы диалогового общения. *Системный анализ в проектировании и управлении*. 2019. Том 13, № 3. С. 433-443.
14. Воронович В.В. Машинный перевод : Конспект лекций для студентов 5-го курса специальности «Современные иностранные языки». Минск : БГУ, 2013. 39 с.
15. Пентус А.Е., Пентус М.Р. Теория формальных языков: Учебное пособие. М.: Изд-во ЦПИ при механико-математическом ф-те МГУ, 2004. 80 с.
16. Куслий П.С. Формальная семантика и онтология. *Epistemology & Philosophy of Science*. 2012. Том 33, № 3. С. 62-67.
17. Бенгфорт Б., Бильбро Р., Охеда Т. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. СПб.: Питер, 2019. 368 с.
18. Белозёров В.В. Подготовка исходных данных к проведению кластеризации документов. Сборник тезисов докладов конгресса молодых ученых (г. Санкт-Петербург, 9–12 апреля 2013 г.). Выпуск 1. Санкт-Петербург, 2013. С 4-5.
19. Большакова И.Е., Баева Н.В. Автоматический анализ дискурсивной структуры научного текста. *Компьютерная лингвистика и интеллектуальные технологи* : труды междунар. конф. «Диалог’2004» (г. Тверь, 2-7 июня 2004). Москва, 2004. С. 68-73.
20. Gong Y., Liu X. Generic text summarization using relevance measure and latent semantic analysis. *Research and Development in Informational Retrieval* : proceedings of XXIV annual international ACM SIGIR Conference (New Orleans, USA, 09-12 September 2001). New York, 2001. P. 19-25.
21. Conroy J.M. O’leary, D.P. Text summarization via hidden markov models. *Research and Development in Informational Retrieval* : proceedings of XXIV annual international ACM SIGIR Conference (New Orleans, USA, 09-12 September 2001). New York, 2001. P. 406-407.
22. Реформатский А.А. Введение в языковедение. Москва : Аспект Пресс, 1996. 536 с.
23. Крутояров Д.В. Автоматизированная система поиска заимствований в электронных изданиях, опубликованных в сети Интернет : дис. … канд. технічних наук : 05.13.06. Москва, 2006. 191 с.
24. Посевкин Р.В. Метод автоматизированного формирования семантической модели базы данных диалоговой системы. *Программные продукты и системы*. 2018. Том 31, №. 2. С. 291-294.
25. Колесниченко А.В., Жмаєва Н. С. Граматичні труднощі автоматизованого перекладу науково-технічної літератури. *Науковий вісник ПНПУ ім. К. Д. Ушинського*. 2018. № 27. С. 134-141.
26. Малахов Д.А., Сидоренко Ю.А., Атаева О.М., Серебряков В.А. Семантический поиск как средство взаимодействия с электронной библиотекой. *Аналитика и управление данными в областях с интенсивным использованием данных*. 2016. С. 148-154.
27. Шадриков В.Д. Пространство действия и смысловая нагрузка слова. *Мир психологии*. 2014. № 2. С. 80-87.
28. Осипов Г.С. Методы искусственного интеллекта. Москва : ФИЗМАТЛИТ, 2011. 296 с.
29. Андреев А.М., Березкин Д.В., Брик А.В., Кантонистов Ю.А. Вероятностный синтаксический анализатор для информационно-поисковой системы. *Компьютерная хроника*. 1999. № 1. С. 37-85.