

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/372774150>

Прикладні аспекти комп'ютерної лінгвістики. Навчально-методичний посібник [Applications of computational linguistics. A handbook]

Book · July 2023

CITATIONS

0

READS

346

1 author:



[Olha Vakhovska](#)

Kyiv National Linguistic University

38 PUBLICATIONS 103 CITATIONS

[SEE PROFILE](#)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЛІНГВІСТИЧНИЙ УНІВЕРСИТЕТ

Ваховська О.В.

ПРИКЛАДНІ АСПЕКТИ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ

Навчально-методичний посібник

Київ
Видавничий центр КНЛУ
2023

УДК 81'322 : 004 (075.8)
В 22

Друкується за рішенням вченої ради
Київського національного лінгвістичного університету
(протокол № 13 від 27 лютого 2023 р.)

Рецензенти: **Андрущенко І.О.** - кандидат філологічних наук, завідувач кафедри фонетики і практики англійської мови Київського національного лінгвістичного університету;
Жихарєва О.О. - доктор філологічних наук, доцент, доцент кафедри іноземних мов Національної академії образотворчого мистецтва і архітектури;
Присяжнюк Л.Ф. - кандидат філологічних наук, доцент, доцент кафедри іноземних мов Інституту міжнародних відносин Київського національного університету імені Тараса Шевченка.

Ваховська, О.В.

В 22 Applications of computational linguistics / Прикладні аспекти комп'ютерної лінгвістики. Навчально-методичний посібник / Ваховська О.В. - К.: Видавничий центр КНЛУ, 2023. - 136 с.

ISBN 978-966-638-362-7

© Ваховська О.В., 2023
© Видавничий центр КНЛУ, 2023

ЗМІСТ

Передмова	4
Розділ 1	
Комп'ютерна лінгвістика й обробка природної мови.....	7
<i>Проектні роботи</i>	30
Розділ 2	
Машинний переклад.....	39
<i>Проектні роботи</i>	41
Розділ 3	
Пошук і кластеризація документів.....	46
<i>Проектні роботи</i>	49
Розділ 4	
Екстракція та узагальнення інформації.....	53
<i>Проектні роботи</i>	56
Розділ 5	
Сентимент-аналіз.....	61
<i>Проектні роботи</i>	63
Розділ 6	
Чат-боти та агенти підтримки діалогу.....	67
<i>Проектні роботи</i>	70
Розділ 7	
Віртуальні світи, ігри та інтерактивна фантастика.....	74
<i>Проектні роботи</i>	76
Розділ 8	
Користувацькі інтерфейси природною мовою.....	80
<i>Проектні роботи</i>	88
Розділ 9	
Системи спільного вирішення проблем та інтелектуальні навчальні системи.....	93
<i>Проектні роботи</i>	96
Розділ 10	
Наділені мовою роботи.....	102
<i>Проектні роботи</i>	104
Завдання для самостійної роботи	110
Післямова	121
Література	123
Нотатки	131

ПЕРЕДМОВА

У навчально-методичному посібнику оглянуто основні поняття комп'ютерної лінгвістики - сучасної галузі знання, метою якої є вивчення природної мови із точки зору комп'ютерного обчислення й розробка артефактів, здатних обробляти і продукувати природну мову в алгоритмічний спосіб, - та розглянуто науково-технічне тло, у яке ці поняття включаються.

Посібник зосереджено, зокрема, на сутності, основних поняттях, цілях і методах комп'ютерної лінгвістики, її витоках і сучасному стані (Розділ 1) й на провідних прикладних здобутках комп'ютерної лінгвістики у створенні систем машинного перекладу (Розділ 2), систем пошуку і кластеризації документів (Розділ 3), систем екстракції та узагальнення інформації (Розділ 4), систем сентимент-аналізу (Розділ 5), чат-ботів та агентів підтримки діалогу (Розділ 6), віртуальних світів, ігор та інтерактивної фантастики (Розділ 7), користувацьких інтерфейсів природною мовою (Розділ 8), систем спільного вирішення проблем та інтелектуальних навчальних систем (Розділ 9), і наділених мовою роботів (Розділ 10).

Розділи посібника передбачають поглиблене опрацювання матеріалу, що його попередньо представлено на відповідних лекціях, із залученням різнотипних завдань і проєктної роботи, задуманих у SMART-форматі як інструменти, за допомогою вибору і поєднання яких можливо сконструювати - згідно з метою і обраним викладачем підходом, а також відповідно до інтересів студентів - навчальний процес як різної форми організації, так і різного рівня складності. Посібник оздоблений завданнями для самостійної роботи студентів над проблемами, що їх окреслено у розділах, і списком літератури й електронних ресурсів.

Перевагою посібника є його вкоріненість в український ґрунт, що її запропоновано студентам у вигляді, по-перше, виконаних нами перекладів фрагментів автентичної англійської фахової літератури на українську мову і, по-друге, дібраної нами інформації насамперед про українські наукові дослідження, інжинірингові розробки і технологічні компанії, які постали криголами у галузі комп'ютерної лінгвістики в Україні, а часом і в світі, і сьогодні продовжують цю галузь розвивати. Студентам пропонується ознайомитися із дотичними роботами працівників провідних українських університетів, зокрема Київського національного лінгвістичного університету, та авторитетних науково-дослідних установ України. Принадою посібника, на нашу думку, є й те, що вміщені у нього завдання і проєктні роботи спрямовані на розвиток у студентів не лише жорстких і цифрових, але й м'яких навичок,

як-от планування, самоорганізація, відповідальність (зокрема, за своє навчання), проєктний і тайм-менеджмент, міжособистісна комунікація, емпатія і командна робота, адаптивність до змін, креативність, мистецтво публічних виступів тощо.

Вихідним для навчання вважаємо принцип *Non scholae, sed vitae discimus*, сформульований Сенекою. У курсі комп'ютерної лінгвістики, що його супроводжує посібник, прагнемо втілити цей принцип через орієнтацію на прикладний характер знань, аби зв'язок між отримуваною студентами інформацією і практичною реальністю, в яку студенти зрештою мають - і воліють - інтегруватися, був максимально унаочнений. Зокрема, у викладанні курсу принагідно посилаємося на реалії сучасного українського і міжнародного ринків праці, знайомимо студентів із релевантними кейсами, скеровуємо студентів на усвідомлення ними свого покликання і на той дієвий пошук можливостей професійної самореалізації, який вони можуть здійснити вже сьогодні. У такий спосіб переносимо увагу з академічних результатів як таких на розвиток насамперед індивідуальних умінь і навичок студентів.

Курс, хоча і задуманий як українськомовний, заохочує студентів знати іноземні мови, зокрема англійську, саме яка наразі домінує на ринку ІТ-технологій та у міжнародному корпоративному ландшафті. Ба більше, курс надихає студентів на вивчення мов програмування, знання яких поступово стає таким, що буде прийматися роботодавцями як належне: студенти мають підготуватися до перспектив і викликів (не такого вже і далекого) майбутнього відносно своєї конкурентоздібності і працевлаштування у ньому. Так, справжнім здобутком курсу вважатимемо розширення кругозору, розвиток загального та емоційного інтелекту студентів, їхнього системного, критичного й аналітичного мислення, здатності до навчання, і, головне, мотивацію студентів до саморозвитку, освіти і самоосвіти протягом життя.

Звертаючись до студентів, дозволимо собі із винятковою вдячністю процитувати нині вже покійного професора Петера Боша, у якого свого часу мали нагоду вчитися комп'ютерній лінгвістиці і чийм неодмінним вітальним словом студентам було таке: “This class is part of an academic study, i.e., we make you an offer to learn a few things, and we will support you in doing so. The decision about how you learn and how much you learn is yours.”¹ Вітаємо студентів Університету на нашому академічному курсі! Академічну культуру визначаємо як насамперед комунікацію: усну і письмову, - й залучення Вас до цієї комунікації вважаємо одним із пріоритетних завдань курсу.

¹ “Цей курс є складовою академічного навчання, що означає, що ми пропонуємо Вам навчитися деяким речам і що ми підтримаємо Вас у цьому навчанні. Як Ви вчитиметеся і чому Ви зрештою навчитеся, вирішувати Вам.” (Тут і далі переклад наш.)

Ця комунікація спирається на відповідний рівень освіченості, що його Вам пропонує підвищити курс комп'ютерної лінгвістики, і передбачає практику академічних свобод і академічної доброчесності. Так, велику увагу у курсі приділено проектним роботам, коли у дискусії висловлюються думки, формулюються та обґрунтовуються точки зору. В якості проектних матеріалів у курс введено наукові статті, робота із якими покликана прилучити Вас до поцінування конвенцій академічного письма і слугувати базою для Ваших власних досліджень. Запрошуємо Вас перейняти погляд на академічну культуру як на спілкування - із підкресленою спільністю дій у досягненні спільної мети. Курс сформувався у своєму нинішньому вигляді саме у спілкуванні і взаємодії із Вашими попередниками - випускниками Університету і студентами старших років навчання. За традицією, курс залишається відкритим для Ваших думок і пропозицій, саме як вітаються цікаві Вам теми й питання і цінується зворотний зв'язок.

Озираючись на той шлях, що цим посібником пройдено від задуму до друку, щиро дякуємо за натхнення, конструктив і підтримку шанованим рецензентам й, особно, кафедрі англійської філології і філософії мови, науково-методичній раді і вченій раді Київського національного лінгвістичного університету, в якому здійлася наша праця.

О.В. Ваховська

РОЗДІЛ 1

Комп'ютерна лінгвістика й обробка природної мови

- ▷ Комп'ютерна лінгвістика як наукова й інжинірингова дисципліна
 - Об'єкт і предмет комп'ютерної лінгвістики
 - Витоки, цілі і методи комп'ютерної лінгвістики
 - Основні поняття комп'ютерної лінгвістики
- ▷ Обробка природної мови
 - Формалізація природної мови
 - Автоматичний синтаксичний аналіз
 - Автоматичний семантичний аналіз
 - Моделювання природної мови у системах штучного інтелекту
- ▷ Комп'ютерна лінгвістика в Україні
 - Провідні наукові установи й основні здобутки
 - Витоки, сучасність і перспективи комп'ютерної лінгвістики в Україні
 - Комп'ютерний фонд української мови

*Людське знання виражено у мові.
Отже, комп'ютерна лінгвістика дуже важлива.*

Марк Стідман

Комп'ютерна лінгвістика² визначається як наукова й інжинірингова дисципліна, завданням якої постає розуміння письмової та усної мови із точки зору її (мови) обчислення, а також розробка артефактів, що здатні ефективно обробляти і продукувати мову, у контексті як великих даних, так і окремого діалогу. Оскільки мова є дзеркалом мислення, розуміння того, як мова обчислюється, дає і розуміння того, як влаштовані людська думка і інтелект. А також, оскільки мова є природним і універсальним засобом спілкування людей, компетентні у лінгвістичному відношенні комп'ютери сприятимуть кращій взаємодії людини й техніки і різноманітного програмного забезпечення, та нададуть людині надійний доступ до безмежних текстових, й не лише, ресурсів Інтернету (Schubert, 2020).

² англ. *Computational Linguistics*, а також *Natural Language Processing*, *Human Language Technology*, *Engineering Linguistics*, *Language Engineering*.

Сутність комп'ютерної лінгвістики розкривається у тих характеристиках цієї дисципліни, що їх протягом курсу розглядається на лекціях і обговорюється під час виконання різних типів завдань на семінарах.

▷ **Комп'ютерна лінгвістика як наукова й інжинірингова дисципліна**



Завдання 1. Уважно прочитайте розділ “Комп'ютерна лінгвістика - новий етап розвитку теоретичної та прикладної лінгвістики” у підручнику (Карпіловська, 2006, с. 7-20).³



Створіть ментальну мапу⁴ інформації, що її подано у розділі. Сформулюйте для себе *a take-home message*⁵ із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 2. Уважно прочитайте визначення комп'ютерної лінгвістики.⁶

³ Завдання базового рівня складності.

⁴ Студентам пропонується ознайомитися із принципами і підходами до створення ментальних мап, а також із відповідними (онлайн-)інструментами.

Ми вдячні професору Світлані Анатоліївні Жаботинській за натхнення залучити практику - і інструменти - створення ментальних мап у викладання. Переконані, що ментальні мапи володіють надзвичайно потужним евристичним потенціалом. Із спілкування з нашими випускниками, знаємо, що багатьом із них саме набуті в Університеті досвід і вміння створювати ментальні мапи відкрили перші двері у світ ІТ-професій, саме як і у світ професійного викладання іноземних мов, де і досі повсюдно ними використовуються. Створення ментальних мап, зрештою, навчає і привчає структурувати інформацію в ефективний спосіб, що в інформаційну епоху набуває особливої ваги.

⁵ *A take-home message* (англ.) розуміємо як “основний посил, основне повідомлення, основна інформація, що їх людина виносить для себе із будь-якого досвіду, зокрема із навчання; висновок” і на семінарах використовуємо як фасилітаційний інструмент.

⁶ Завдання середнього рівня складності.

Комп'ютерна лінгвістика - це ...

... вивчення мови у термінах алгоритмів (Dunlop, & Fetzer, 1993).

... наукове вивчення мови із обчислювальної точки зору (Gallego, & Martin, 2020).

... застосування комп'ютерної науки при аналізі і задля розуміння письмової і усної мови (Gillis, 2022).

... застосування комп'ютерної науки до аналізу, синтезу і розуміння письмової і усної мови (Hunston, 2022).

... застосування лінгвістичних теорій і обчислювальних технік до проблем обробки природної мови (Bolshakov, & Gelbuch, 2004).

... наука про мову, де особлива увага приділяється тим обмеженням, які накладаються на складність обробки природної мови особливостями архітектури людського мислення. Як і більшість інших наук, комп'ютерна лінгвістика має інжинірингові застосування (Hausser, 2014).

... вивчення комп'ютерних систем із метою зрозуміти і генерувати природну мову (Grishman, 1986).

Чому, на Вашу думку, комп'ютерна лінгвістика має ці різні визначення? Чи є вони доцільними?

Знайдіть у цих визначеннях інваріант vs. варіанти. За потреби, доповніть список визначеннями комп'ютерної лінгвістики, що їх подано у вподобаній Вами фаховій літературі.

Сформулюйте власне - на Ваш погляд, необхідне і достатнє - визначення комп'ютерної лінгвістики.



Завдання 3. Уважно прочитайте уривок із першоджерела:⁷

Теоретичні цілі комп'ютерної лінгвістики полягають у створенні таких підходів до граматики і семантики природної мови, які уможливили б обчислення цієї мови із застосуванням **методів синтаксичного і семантичного аналізу**, що є здійсненними практично; виявлення і пояснення механізмів обробки природної мови у мозку і мисленні людини, а також формулювання принципів навчання мови, які охопили б її структурні і дистрибуційні (статистичні) властивості; створення обчислювальних моделей обробки і навчання мови: оскільки ці процеси здійснюються у мозку і мисленні людини, створювані комп'ютерними лінгвістами моделі мають бути нейрофізіологічно і когнітивно правдоподібними.

Практичні цілі комп'ютерної лінгвістики є дуже різноманітними. Деякими із основних є ефективний текстовий пошук за заданою тематикою; ефективний машинний переклад; відповіді на запитання, які включають як прості запитання про певні факти, так і запитання, що вимагають умовиводів і дескриптивних або дискурсивних відповідей (можливо, відповідей із обґрунтуваннями); реферування текстів; аналіз теми, тональності або інших психологічних характеристик текстів, а також усного мовлення; створення діалогових помічників і діалогових систем задля вирішення окремих завдань (здійснення покупок, надання технічного супроводу, організація подорожей, створення розкладу, надання медичної консультації тощо); і, в підсумку, створення комп'ютерних систем, чий механізм засвоєння мови, володіння мовою у діалозі й набуття знань із тексту дорівнювали б тим, які наявні у людей (Schubert, 2020).

Схарактеризуйте своїми словами основні теоретичні і практичні цілі комп'ютерної лінгвістики.

Наведіть приклади реалізації практичних цілей комп'ютерної лінгвістики із свого повсякденного життя. Підсумуйте, наскільки Ви як користувач задоволені досвідом своєї взаємодії із окремими інструментами реалізації цих цілей, що їх пропонує ринок сучасних технологій. У чому Ви як фахівець вбачали б можливість вдосконалення цих інструментів?

⁷ Завдання базового рівня складності.



Завдання 4. На основі авторитетних джерел, підготуйте коротку інформацію про такі системи:⁸

ENIAC - Mauchly, Eckert et al., 1946

МЕЛІМ - Лебедев, 1951-1953

Sad Sam - Lindsay, 1963

ELIZA - Weizenbaum, 1966

Sir - Raphael, 1968

Student - Bobrow, 1968

CYRUS - Kolodner, 1970

Lunar - Woods et al., 1972

PARRY - Colby, 1972

SHRDLU - Winograd, 1972

HACKER - Sussman, 1973

BORIS - Dyer, 1980

Відстежте еволюцію підходів до створення, використовуваних у цих системах.
Визначте спільні й відмінні риси цих систем.



Завдання 5. Вкажіть внесок цих особистостей у становлення і розвиток когнітивної науки, зокрема комп'ютерної науки і комп'ютерної лінгвістики. Якщо вважаєте доцільним, доповніть список:⁹

Л. Баум (Leonard Baum)

Ч. Бебідж (Charles Babbage)

Е. Бут (Andrew Booth)

Н. Вінер (Norbert Wiener)

Т. Віноград (Terry Winograd)

С. Возняк (Stephen Wozniak)

С. Джобс (Steven Jobs)

Р. Квіліан (Ross Quillian)

А. Лавлейс (Ada Lovelace)

⁸ Завдання базового рівня складності.

⁹ Завдання базового рівня складності.

Сергій Олексійович Лебедєв
М. Мінський (Marvin Minsky)
А. Тьюринг (Alan Turing)
У. Уівер (Warren Weaver)
Р. Хартлі (Ralph Hartley)
Н. Чомські (Noam Chomsky)
К. Шеннон (Claude Shannon)

Прізвища у списку розташовані за алфавітом. Намалуйте стрілу часу, яка б унаочнювала послідовність розробок і праць цих особистостей в еволюції науки і техніки.

▷ Основні поняття комп'ютерної лінгвістики

До основних понять комп'ютерної лінгвістики, що із ними пропонується ознайомитися студентам, належать:¹⁰

Алгоритм (Algorithm) - абсолютно надійна послідовність дій або процедур, яку можна виконати кінцевою кількістю кроків, аби вирішити проблему. Алгоритми інколи також називаються ефективними процедурами.

Гомункул (Homunculus) - буквально, “маленька людина,” уявлення про гомункулуса закладено у теоріях мислення, які посилаються на одні когнітивні процеси у своїх поясненнях інших когнітивних процесів. Наприклад, інколи стверджується, що зорове сприйняття передбачає утворення у мисленні образів. Якщо це твердження правильне, то слід було б передбачити, що на ці образи як на внутрішні картини має дивитися певний внутрішній споглядач - гомункулус, але тоді слід було б також передбачити, що і у гомункулуса має бути у мисленні свій гомункулус, який і дивиться на ці картини, і так до безкінечності. Див. також (Кришталь, 2022).

Глибинна структура (Deep structure) - рівень аналізу в *трансформаційній граматиці* (transformational grammar), протиставлений *поверхневій структурі* (surface structure), розрізнення яких було запропоноване Н. Чомскі. Відношення між двома рівнями виявляється у трансформаційній граматиці, за якою поверхнева структура є похідною від глибинної структури. Як вказує термін “глибинна структура,” цей рівень аналізу не відповідає поверхневій структурі конкретного речення безпосередньо, але дозволяє дійти до певних узагальнень стосовно прихованих подібностей у глибинній структурі низки речень, які поверхнево відрізняються одне від одного.

Грамматика структури синтаксичних груп (Phrase structure grammar) - грамматика, яка передбачає групування синтаксичних компонентів за ієрархічним принципом. Так, наприклад, речення може бути розкладене на *іменникову групу* (noun phrase) і *дієслівну групу* (verb phrase), які, за принципами *ієрархії мов за Н. Чомскі* (Chomsky hierarchy), будуть розкладені у свою чергу на менші синтаксичні компоненти.

¹⁰ Якщо не зазначено інакше, терміни, що ними позначаються основні поняття комп'ютерної лінгвістики, у цьому підрозділі визначено услід за (Dunlop, & Fetzer, 1993).

Дерево (Tree) - ієрархічна структура, яка складається із одного **вузла** (a node) (термінологічно позначеного як **кореневий вузол** (the root node)), до якого сходять численні **гілки** (branches), що зв'язують цей вузол із іншими вузлами цього дерева. <...> Використовується для представлення граматичної структури речення; для представлення можливих вирішень однієї проблеми шляхом розбиття цієї проблеми на підпроблеми тощо.

Дієслівна група (Verb phrase, VP) - одиниця граматичного аналізу, яка у найпростішому вигляді складається лише із дієслова або із дієслова і додатка (представленого як **іменникова група** (noun phrase)). Прикладами дієслівних груп є “eat” і “played the banjo.” Один із рівнів аналізу у **граматиці структури синтаксичних груп** (phrase structure grammar).

Експертна система (Expert system) - система штучного інтелекту, яка демонструє рівень володіння специфічним знанням на рівні людини-експерта. Експертна система складається із трьох основних компонентів: **бази знань** (knowledge base), зібраної зі знань людей-експертів певної галузі шляхом процесу, що зветься інженерія знань; набір **правил для прийняття рішень** (decision rules), які визначають, що саме системі слід зробити в певних умовах; і **механізму логічного висновку** (inference engine), який застосовує по відношенню до бази знань відповідні умовам правила.

Ієрархія мов за Н. Чомскі (Chomsky hierarchy) - упорядкування типів граматик, яке відображає спектр формальних мов, що їх може генерувати кожна така граматика. У цьому упорядкуванні, мови, що їх генерують слабші граматики, є підмножинами тих мов, що їх генерують сильніші граматики. В ієрархії виокремлюються чотири рівня, позначені як 0,1,2 і 3, де рівень 0 репрезентує найсильнішу граматику, а рівень 3 - найслабшу. Кожен рівень в ієрархії точно відповідає окремому типу **автомату** (automaton).

Іменникова група (Noun phrase, NP) - одиниця граматичного аналізу, яка складається із займенника або детермінатива (необов'язково) плюс із одного або двох прикметників (необов'язково) плюс іменника. Прикладами іменникових груп є “it,” “she,” “frog,” “a slimy frog” і “the beautiful old guitar.” Один із рівнів аналізу у **граматиці структури синтаксичних груп** (phrase structure grammar).

Інтелект (Intelligence) - інтелект традиційно визначається як здатність істот, наділених мисленням, вчитися. Коли ця здатність оцінюється відносно хронологічного віку людини за використання стандартизованих тестів, то результатом такої оцінки постає певне число як *коефіцієнт інтелекту* (intelligence quotient). Виникнення терміну “штучний інтелект” свідчить про те, що машини, хоча і не наділені мисленням, теж можуть описуватися у термінах інтелекту, тому що машини успішно і передбачувано стабільно виконують певні завдання - особливо такі завдання, вирішення яких зазвичай вимагає залучення людей.

Інтерлінгва (Interlingua) - система символів для представлення значень речень природної мови; використовується у деяких системах обробки природної мови. У межах цього підходу, речення із значно схожими значеннями представляються однаково, навіть якщо їхні поверхневі форми значно відрізняються одна від одної. Наприклад, речення “A threw the ball to B,” “The ball was thrown by A to B” і “B caught the ball that A threw” виражають одну й ту саму головну ідею, що *м'яч був переданий від А до В у просторі*. Використання такого виразу, поданого курсивом, у якості спільної інтерлінгви для цих трьох речень полегшує завдання їхнього схемного представлення. Так, використання єдиної інтерлінгви для представлення значень речень будь-якої природної мови спрощує завдання машинного перекладу, оскільки замість правил, за якими одна мова має бути співвіднесена з іншою, машинний перекладач, базований на інтерлінгві, вимагає лише правил перекладу цих мов на інтерлінгву. Тоді будь-які дві природні мови можуть бути перекладені одна на одну за використання інтерлінгви як їхнього єдиного проміжного представлення. Штучна мова Есперанто була запропонована у якості інтерлінгви для міжнародної комунікації.

Китайська Кімната (Chinese Room) - приклад, наведений Джоном Серлем як ілюстрація його твердження, що тест Тьюринга є неадекватним критерієм для визначення наявності справжнього мислення. У Китайській Кімнаті знаходиться людина, яка не знає китайської, але у якої наявний повний і витончений набір правил поєднання символів, завдяки чому ці символи, хоча і залишаються без значень, притаманних їм у китайській мові, можна комбінувати у граматично правильні для китайської мови послідовності. Людина отримує зовні символи, які вона у Кімнаті комбінує за правилами й потім повертає назовні. Людина цього не знає, але отримувані символи вона комбінує у граматично правильні відповіді на питання до розповіді,

що її людина також не знає, - цю розповідь знають люди назовні: вони знають китайську і на основі відповідей людини із середини приходять до хибного висновку, що ця людина вільно володіє китайською. Уподібнюючи те, що відбувається у Китайській Кімнаті, до того, що відбувається у комп'ютері із його програмним забезпеченням, Серл стверджує, що оперування символами за правилами, але без знання значень цих символів, не є інтенціональним, а, отже, і не дає підстав для суджень про наявність чи відсутність справжнього мислення.

Комбінаторний вибух (Combinatorial explosion) - різке зростання кількості можливих варіантів аналізу синтаксичної структури речення або його окремих частин при комп'ютерній обробці цього речення.

Лексична субституція (Lexical substitution) - заміна кінцевих вузлів синтаксичного дерева словами мови.

Лематизація (Lemmatization) (від лат. *lemma* "словникова форма") - процедура комп'ютерної і корпусної лінгвістики, що відновлює словникову форму за її словоформою.

Лінгвістичний процесор (Natural language processor) - складна система автоматичної обробки природної мови, здатна до аналізу і синтезу текстів, що забезпечує діалог у системах штучного інтелекту.

Машина Тьюрінга (Turing machine) - абстрактний прилад, який складається із безкінечної "стрічки," розділеної на комірки, кожна із яких може містити лише або 0, або 1. Сканер працює над стрічкою, зчитуючи вміст однієї комірки за один раз і виконуючи набір серійних операцій, кожна з яких сумарно залежить від вмісту цієї конкретної комірки і від стану машини у цей конкретний момент її оперування. Програма машини визначає, що саме машина має зробити у кожен момент її оперування, - наприклад, "Якщо зчитано символ 0 і в цей час машина знаходиться у стані 6, тоді 0 слід замінити на 1 і переключитися у стан 4."

Машинна мова (Machine language) - мова, на яку безпосередньо реагує комп'ютер, на відміну від **мов вищого рівня** (higher-level languages) (Pascal, BASIC, Fortran тощо), які повсюдно використовуються програмістами. Представлені двійковими числами, версії програм вищого

рівня, написані машинною мовою, створюються за допомогою *компілятора* (a compiler) або *інтерпретатора* (an interpreter).

Метамова (Metalanguage) - “засіб спілкування людей, який виник в еволюції людства як засіб передачі почуттів. У мистецтві стільки мистецтва, скільки у ньому метамови. Проте терміни не ідентичні, існує потреба в обох. Метамова - емоційне тло життя, а мистецтво - засіб, щоб тло прикрасити і зробити спільним” (Кришталь, 2020).

Мова думки (Language of thought) - гіпотетична вроджена схема кодування, яка пропонує достатньо можливостей для вираження будь-якої відмінності об'єктів світу, що її взагалі може описати будь-яка природна мова. У сучасній когнітивній науці, ця ідея належить Джеррі Фодору, який стверджує, що така вроджена мова (інколи її називають *менталіз* (mentalese)) є логічно необхідною задля вивчення природної мови, адже дитина, яка опиняється у своєму першому мовному середовищі і починає вчити свою першу природну мову, має почати формулювати припущення про те, до яких об'єктів відносяться мовні вирази, але задля формулювання таких припущень вже потрібна якась попередньо наявна мова. Аби уникнути безкінечної регресії, потрібно постулювати наявність вродженої мови думки.

Мова програмування (Programming language) - набір конструктив *формальної мови* (formal language), які уможливають запис *алгоритмів* (algorithms) у вигляді *програм* (programs), тобто із використанням словника і структури, на які безпосередньо може реагувати комп'ютер. Інколи розрізняються *мови програмування високого і низького рівнів* (high- vs. low-level programming languages). Мови програмування високого рівня, як правило, мають великий спектр конструктив, які можуть у достатній мірі відобразити більшість категорій людської думки і зв'язків між цими категоріями. Такі мови достатньо легко запам'ятовуються і використовуються. Тим не менш, мови програмування високого рівня не повністю відображають ті процеси, що відбуваються у комп'ютері, і тому мають перекладатися *компілятором* (compiler) або *інтерпретатором* (interpreter) в *об'єктний код* (object code), який може використовувати комп'ютер. На відміну від мов високого рівня, мови програмування низького рівня достатньо точно відображають принципи роботи комп'ютера, але їхня невідповідність категоріям і процесам людської думки робить їх значно менш інтуїтивними.

Модель (Model) - спрощений об'єкт (система, процес тощо), що зберігає лише важливі властивості справжнього об'єкта (системи, процесу тощо), який моделюється і саме для вивчення якого створюється модель: дослідження моделі є засобом отримати інформацію про модельований об'єкт.

Комп'ютерні лінгвісти прагнуть створити **обчислювальні моделі** (computational models) різноманітних мовних явищ. Побудова цих моделей залучає правила (такі моделі визначаються як knowledge-based, rule-based або hand-crafted) та/або мовні дані (такі моделі є data-driven, statistical або empirical). Особливості обох типів моделей, услід за (Gallego, & Martin, 2020), можна узагальнити у таблиці:

<i>Моделі, базовані на правилах</i>	<i>Моделі, базовані на даних</i>
Кодування знання про мову, експліцитне	Кодування знання у мові, імпліцитне
Залучення набору створених або відібраних вручну граматичних правил	Залучення статистичних методів або методів машинного навчання
Легкі тестування і відладка	Дещо ускладнені тестування і відладка
Потребують значної участі людини	Потребують не такої значної участі людини
Часто базуються на обмеженому огляді даних, акцент на прототипних прикладах	Повністю керовані даними і вимагають великих даних
Часто не покривають у достатній мірі модельовану царину	Ступінь покриття модельованої царини прямо пропорційна обсягу даних
Часто втрачають ефективність при роботі із зашумленими даними	Легше адаптуються до зашумлених даних

Навчання (Learning) - (а) Будь-яка зміна у поведінці суб'єкта, що її викликає попередній досвід. (б) Самоадаптація суб'єкта до покращеної продуктивності у відповідь на зміну середовища і обставин. Це

покращення зазвичай пояснюється через набуття нових знань і здатність використовувати ці знання.

Неоднозначність (Ambiguity) - якщо слово, фраза або речення мають більше, ніж одну можливу інтерпретацію, то про них говорять, що вони є неоднозначними. Неоднозначність може виникати із різних джерел. **Синтаксична неоднозначність** (Syntactic ambiguity) виникає тоді, коли одна й та сама лінгвістична структура припускає кілька граматичних інтерпретацій - наприклад, різні інтерпретації фрази “old men and women” (Жінки теж були старими?) і речення “I greeted the woman with the flowers” (Квіти були у неї, чи у мене?) <...> У випадках **семантичної неоднозначності** (semantic ambiguity) слово, фраза або речення можуть інтерпретуватися як такі, що мають різні значення, і вимагають контексту для усунення такої неоднозначності - наприклад, іменник “an organ” на позначення частини тіла і музичного інструменту¹¹ або речення “Everybody isn't here,” яке можна інтерпретувати як те, що не всі тут, або як те, що всі не тут.

Об'єктний код (Object code) - програма *машинною мовою* (machine language), створена компілятором із програми, написаної мовою вищого рівня, наприклад, мовою Pascal. Програма об'єктного рівня виконує інструкції, які були визначені в програмі вищого рівня, із якої вона була створена, але інструкції при цьому перекладаються на машинну мову, саме безпосередньо на яку може відреагувати комп'ютер.

Обробка інформації (Information processing) - введення, виведення, зберігання, пошук і оперування символами, здійснювані за допомогою алгоритмів.

Обробка природної мови (Natural language processing) - аналіз генерування і розуміння речення із урахуванням процесів його обчислення. Кінцева мета цього підходу у когнітивній науці - побудувати комп'ютерну модель, яка відображала б усі ті процеси обробки мови, що використовуються людьми, при тому, що наразі повне розуміння цих процесів у науці відсутнє.

Обчислення (Computation) - систематична трансформація вхідних даних у вихідні дані, при цьому зміст цих даних зберігається.

¹¹ Див. також (Vakhovska, 2022), де, зокрема, здійснено аналіз мережі полісемії іменника *an eye*, який у сучасній англійській мові має 44 лексико-семантичних варіанти.

Онтологія (Ontology) - в інжинірингу знань (knowledge engineering), онтологія визначається як артефакт, що у межах певної царини знань специфікує об'єкти (або поняття про ці об'єкти) і зв'язки між ними, а також властивості цих об'єктів (або понять про них) (Peters, & Shrobe, 2003).

Парсер, Лінгвістичний аналізатор (Linguistic parser) - комп'ютерна програма, що аналізує поверхневу синтаксичну структуру речення із метою виявити глибинну синтаксичну структуру цього речення.

Поверхнева структура (Surface structure) - рівень аналізу у *трансформаційній граматиці* (transformational grammar), протиставлений *глибинній структурі* (deep structure) (розрізнення було запроваджене Н. Чомскі). На рівні поверхневої структури, спостерігається пряма відповідність між реченням і його граматичним представленням. Відношення між поверхневою і глибинною структурами речення виявляється трансформаційною граматикою.

Прийменникова група (Prepositional phrase, PP) - мовна структура, яка складається із прийменника і іменникової групи. Одна із одиниць аналізу у *граматиці структури синтаксичних груп* (phrase structure grammar).

Природна мова (Natural language) - будь-яка конвенційна система символів і правил їхнього поєднання, яка виникла природним шляхом, на відміну від *штучних мов* (artificial languages), що виникають внаслідок задуму і свідомої роботи людей, які ці мови створюють. Символи у цій системі можуть бути письмовими, усними або знаковими (у мовах, які використовуються для спілкування з глухими). <...> Природні мови зазвичай слугують широкому спектру комунікативних потреб і тяжіють до багатства і різноманіття вираження, але не до його ефективності та економності.

Семантична мережа (Semantic network) - структура представлення знань (knowledge representation), яка складається із *вузлів* (nodes), які представляють слова або поняття, і *зв'язків* (links), які представляють відношення між словами або поняттями.

Сильний штучний інтелект (Strong AI) - твердження про те, що, у відповідних умовах (наприклад, при запуску вірної програми), комп'ютер і

дійсно отримав би розум; тобто, комп'ютер мав би здатність до мислення у неметафоричному смислі.

Символьний штучний інтелект (Symbolic AI) - підхід до штучного інтелекту, який наголошує важливість правил для оперування символами різного типу. <...> Відображає думку тих когнітологів, які вважають, що людський інтелект виникає внаслідок оперування символами, яке здійснює людський мозок. Протиставляється **конекціонізму** (connectionism).

Синтаксичний аналіз, або **парсинг** (Parsing) - процес аналізу граматичної структури речення із метою її визначення. Вирізняють **синтаксичний аналіз знизу догори** (bottom-up parsing) і **синтаксичний аналіз згори донизу** (top-down parsing). Метою обох підходів є аналіз структури речення відносно набору граматичних правил, але, як свідчать самі терміни, один підхід є протилежним іншому. Наприклад, речення "Pat ate some strawberries" має таку граматику (символ "→" має значення "складається з"):

Sentence → *Noun Phrase* + *Verb Phrase*

Noun Phrase → *Noun*, or *Article* + *Noun*

Verb Phrase → *Verb* + *Noun Phrase*

Відповідний словник включає наступні статті, із відповідними категоріями слів:

Pat: *Noun*

strawberries: *Noun*

ate: *Verb*

some: *Article*

Словник показує, що речення "Pat ate some strawberries" складається із *Noun* + *Verb* + *Article* + *Noun*. Аналіз знизу догори починає із цих одиниць і намагається знайти структури вищого рівня (в кінцевому підсумку *Sentence*), що їх утворюють ці одиниці. Цього можна досягти, якщо зчитувати граматичні правила справа наліво. Так, ця процедура визначає *Article* + *Noun* як складники *Noun Phrase*; отже, початкове *Noun* + *Verb* + *Article* + *Noun* можна переписати у вигляді структури вищого рівня *Noun* + *Verb* + *Noun Phrase*. Послідовне (справа наліво) використання цих саме правил зрештою дає структуру найвищого рівня *Sentence*. Отже, речення "Pat ate some strawberries" є прийнятним згідно із цією граматиною.

Для порівняння, аналіз згори донизу починається із структури найвищого рівня *Sentence* і намагається розкласти цю структуру на складники так, аби отримати вихідне речення. Починаючи із *Sentence* і

використовуючи правила зараз вже зліва направо, отримуємо *Noun Phrase* + *Verb Phrase* і так далі, аж поки не буде отримано речення “Pat ate some strawberries,” яке є граматично прийнятним згідно із цими правилами. Утім, при цьому можна отримати і інші речення, наприклад “Strawberries ate some strawberries,” які, хоча і будуть абсурдними, але все ж таки відповідатимуть заданій граматиці. Граматична правильність не дорівнює правильності змістовній.

Обидва підходи - знизу догори і згори донизу - мають використовуватися залежно від мети аналізу і від того, яке саме речення аналізується. Інколи ці два підходи комбінуються між собою.

Слабкий штучний інтелект (Weak AI) - твердження про те, що комп'ютери є корисними і дієвими інструментами для вивчення мислення, наприклад, шляхом формулювання і перевірки гіпотез про механізми обробки мови або роботи пам'яті. Утім, на відміну від **сильного штучного інтелекту (strong AI)**, слабкий штучний інтелект не припускає, що комп'ютери буквально є (або могли б бути) мислячими суб'єктами.

Теорія інформації (Information theory) - заснована Клодом Шенноном, теорія інформації є математичним підходом до вивчення комунікації. Цей підхід наголошує кількість інформації в одиницях (біт - базова одиниця вимірювання інформації), а не її зміст, і уможливорює цифрове представлення цих одиниць. Наприклад, якщо збирається інформація про автомобілі, то категорії старі vs. нові автомобілі, автомобілі вітчизняного vs. іноземного виробництва, великі vs. невеликі автомобілі можуть бути представлені двійковим числом: старі = 0 і нові = 1, вітчизняні = 0 і іноземні = 1, великі = 0 і невеликі = 1, і тоді інформація про будь-який автомобіль може бути представлена трьома бітами, а, наприклад, старий вітчизняний невеликий автомобіль отримав би код 001.

Тест Тьюрінга (Turing test) - мисленнєвий експеримент; операційний тест, запропонований А. Тьюрінгом з метою надати відповідь на питання “Чи може машина думати?” Тест включає людину, яка ставить різноманітні питання на будь-яку тему чоловікові і жінці, які знаходяться в ізолюваному від цієї людини приміщенні; комунікація здійснюється через систему передачі повідомлень. Спочатку, завданням цієї людини є визначити, хто із двох відповідачів є чоловіком, а хто - жінкою. Потім, чоловіка замінює комп'ютер. Якщо людина, яка ставить запитання, продовжує вірити, що обидва відповідачі - люди, то робота комп'ютера

точно відповідає роботі людини, і тоді слід визнати, що комп'ютер здатен думати. Відповідність тесту Тьюринга меті визначити наявність мислення була поставлена під сумнів Джоном Серлем із його відомим антиприкладом під назвою *Китайська Кімната* (Chinese Room).

Трансформаційна граматики (Transformational grammar) - запропонована і розвинена Н. Чомські граматики, яка залучає два рівні репрезентації: глибинну структуру і поверхневу структуру. Сутність цієї гіпотези полягає у тому, що продукування речень природною мовою є результатом деривацій від глибинних структур, які лежать в основі мови і які мовець не усвідомлює.

Універсальна граматики (Universal grammar) - запропонована Н. Чомські, універсальна граматики є набором обмежень, яким має відповідати граматики будь-якої природної мови. На думку Н. Чомські, надзвичайно мала кількість мовних даних, які отримує дитина під час засвоєння мови, має бути пояснена існуванням *вродженої мовної бази* (innate linguistic framework). <...> Із цієї точки зору, усі природні мови (як існуючі, так і можливі) мають у своїй основі певні спільні риси.

Формальна мова (Formal language) - специфікація набору *правильно оформлених формул* (well-formed formulae), шляхом визначення (1) набору символів і (2) набору правил, які застосовуються для допустимого, тобто граматично правильного, комбінування цих символів.

Штучна мова (Artificial language) - будь-яка система запису і правила її використання, яка була спеціально створена із особливою метою, що не є метою звичайної комунікації між членами мовної спільноти. До прикладів належать системи формальних мов у теоретичній математиці (як неінтерпретовані системи), мови програмування у комп'ютерній науці (як інтерпретовані системи) і Есперанто (людська мова загального призначення, покликана полегшити міжнародну комунікацію). На деякі окремі випадки (як, наприклад, азбука Морзе) можна дивитися як на граничні між природними і штучними мовами. Протиставляється *природній мові* (natural language).

Штучний інтелект (Artificial intelligence) - розділ комп'ютерної науки, який зосереджується на тому, наскільки ментальні можливості людських істот можуть бути відтвореними за допомогою машин. Дослідження зі

штучного інтелекту є формуючими для міждисциплінарної царини когнітивної науки у цілому.



Завдання 6. Подані у цьому підрозділі терміни розташовані за алфавітом. Розташуйте їх у вигляді (фрагментів) семантичної мережі, (фрагментів) поняттєвої схеми або (фрагментів) ментальної мапи тощо за поняттєвими зв'язками.¹²

¹² Завдання високого рівня складності. Можливе виконання цього завдання у вигляді підсумкової проєктної роботи для структурації всієї вивченої інформації.

▷ Обробка природної мови

Мова структурується на різних своїх рівнях, починаючи, у випадку із усним мовленням, із певних закономірностей, наявних в акустичному сигналі, що можуть бути визначені як *фони* (відмінні один від одного і, як такі, впізнавані звуки, які змінюють один одного й із яких вибудовується мова). Групи фонів, які є специфічними для конкретної мови (і які, якщо й дещо змінюються, не перешкоджають впізнаванню слів слухачем), визначаються як *фонем* цієї мови. Фонем, у свою чергу, є складовими *морфем* (мінімальних значущих частин слів), які є вже складовими слів. (У випадку письмового мовлення, на відміну від усного, говорять про символи, графем, склади і слова.) Слова об'єднуються у *словосполучення*, як-от іменникові, дієслівні, прикметникові і прийменникові, які виступають структурними компонентами *речень*, здатних виражати завершені думки. На ще вищому рівні, виокремлюють різні типи структури *дискурсу*, яка, втім, є дещо вільнішою, порівняно із структурами нижчих рівнів.

Були створені відповідні техніки автоматичного аналізу мови на всіх її структурних рівнях <...>. Мова слугує передачі значення. Отже, аналіз синтаксичної структури є лише половиною шляху до механізації цієї головної функції мови, й усі можливі переваги і досягнення окремих підходів до синтаксису залежать перш за все від їхньої можливості підтримати семантичний аналіз і генерувати мову із тих значень, що їх потрібно передати.

Це не означає, що синтаксичний аналіз у собі не має жодної цінності - він може бути корисним у таких прикладних аспектах, як перевірка граматичної правильності або статистичний машинний переклад. Але задля амбіційнішої мети виведення і вираження значення мови необхідно є насамперед теорія *семантичної репрезентації*, рівно як і її зв'язок із поверхневою структурою, і її взаємодія із репрезентацією фонових знань і їхнім використанням. <...> [У цьому розділі] ми розумітимемо семантичну інтерпретацію як процес виведення семантичних репрезентацій із потоку слів, і вважатимемо, що цьому процесу передуює, або здійснюється із ним одночасно, процес автоматичного синтаксичного аналізу (Schubert, 2020).



Завдання 7. Уважно прочитайте розділ “Інформаційні моделі лінгвістичних об’єктів” у підручнику (Карпіловська, 2006, с. 20-34).¹³

¹³ Завдання базового рівня складності.



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 8. Уважно прочитайте розділ “Автоматичний синтаксичний аналіз тексту” у підручнику (Карпіловська, 2006, с. 139-144).¹⁴



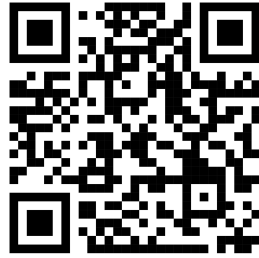
Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 9. Уважно прочитайте розділ “Автоматичний логіко-семантичний аналіз тексту” у підручнику (Карпіловська, 2006, с. 144-153).¹⁵

¹⁴ Завдання базового рівня складності.

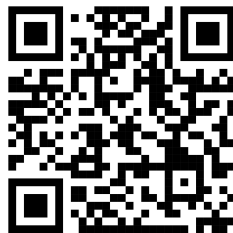
¹⁵ Завдання базового рівня складності.



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 10. Уважно прочитайте розділ “Природний інтелект (=інтелект людини) і штучний інтелект (=інтелект комп’ютера) як його модель” у підручнику (Карпіловська, 2006, с. 106-114).¹⁶



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.

¹⁶ Завдання базового рівня складності.

▷ Комп'ютерна лінгвістика в Україні



Завдання 11. Уважно прочитайте статтю “Комп'ютерна лінгвістика” в Енциклопедії Сучасної України (Карпіловська, 2014).¹⁷



Створіть ментальну мапу інформації, що її подано у статті. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 12. Уважно прочитайте розділ “Комп'ютерний фонд української мови в Інституті мовознавства ім. О.О. Потебні НАН України” у підручнику (Карпіловська, 2006, с. 96-106).¹⁸



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.

¹⁷ Завдання базового рівня складності.

¹⁸ Завдання базового рівня складності.



Завдання 13. Дослідіть призначення і функціонал Корпусу української мови:¹⁹



Дослідіть призначення і функціонал Генерального регіонально анотованого корпусу української мови:²⁰



Знайдіть і дослідіть інші корпуси української мови та слов'янських мов, а також провідні корпуси англійської та німецької мов. Поекспериментуйте із цими корпусами. Як Ви можете використати їх у навчанні, наукових дослідженнях й (майбутній) професійній діяльності?²¹

Структуруйте результати своєї розвідки і представте їх на семінарі.

¹⁹ Завдання високого рівня складності.

²⁰ Див. також (Shvedova, 2020).

Закликаємо студентів відстежувати Генеральний регіональний анотований корпус української мови (ГРАК) у соціальних мережах, зокрема у Фейсбуці, де ГРАК є доволі активним і часто запрошує українських студентів долучитися до своїх проєктів.

²¹ Див. також (Шведова, 2022).

Проектні роботи до Розділу 1



Проектна робота 1. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] У “Короткому англо-українському словнику з комп’ютерної лінгвістики” (Коломієць, 2022), знайдіть приклади, які ілюструють основні положення статті. Підготуйте і проведіть мультимедійну презентацію²² на основі статті і знайдених Вами прикладів.²³

Завадська, В. (2013). Коли “вікно” не є вікном, або ще раз про сучасну українську ІТ-термінологію. *Українське мовознавство*, 43, 20-27.



***Анотація.** Розглядаються особливості перекладу термінів українською мовою, зокрема в ІТ-сфері, в галузі комп’ютерної термінології. Окреслено проблеми, пов’язані з мовою-посередницею та явищем калькування термінів, вказано причини появи термінів, що є нехарактерними для української літературної мови. Подано деякі рекомендації практичного характеру.*

²² Студентам пропонується ознайомитися із принципами і підходами до створення презентацій, а також із різними типами презентацій (мультимедійна, постер, “презентація у ліфті” (англ. *an elevator pitch*) тощо) і відповідними (онлайн-)інструментами. Студенти заохочуються до набуття досвіду й навичок публічних виступів.

Ми вдячні професору Каю-Уве Кюнбергеру за натхнення залучити практику підготовки і проведення різних типів презентацій у викладання. Переконані, що презентації, окрім того, що володіють надзвичайно потужним евристичним потенціалом, тренують водночас цілий набір м’яких навичок.

У нашому курсі, ця практика розтлумачується студентам як завдання підготувати і провести презентацію на тему, що її представлено у відповідній науковій статті: студентам слід виокремити у статті саме той матеріал, який є для них важливим і цікавим; студенти можуть додавати матеріал до презентації, якщо, на їхню думку, цей матеріал сприятиме кращому розумінню й сприйняттю презентації аудиторією; студентам під час презентації слід критично обговорити окремі положення і приклади із статті, додавши своє бачення проблеми та/або власні приклади. Важливо: по завершенню презентації, у студентів є можливість задати питання аудиторії - цей етап покликаний залучити у дискусію усіх учасників семінару (при цьому, не очікується, що як самі доповідачі, так і їхня аудиторія матимуть чіткі відповіді на поставлені питання, адже мета цього етапу - генерувати цивілізовану дискусію як обмін думок). Тривалість презентації - орієнтовно 45 хвилин, плюс час на дискусію.

²³ Уміщені у навчально-методичний посібник проектної роботи намічені як такі, що мають високий рівень складності.

Ключові слова: переклад, термін, ІТ-термінологія, ІТ-переклад, стандартизація термінів.

Коломієць, В.О. (2022). *Короткий англо-український словник з комп'ютерної лінгвістики*. Київ: Видавничий центр КНЛУ.



Анотація. Словник містить переклад близько 1000 англомовних термінів із комп'ютерної лінгвістики, а також слів і словосполучень, які не є термінами, але вживаються в науковій літературі. Укладений на основі корпусу анотацій наукових статей, опублікованих з 2000 по 2017 рік у журналі “*Computational linguistics*”, який є офіційним виданням Міжнародної асоціації комп'ютерної лінгвістики. Для студентів, аспірантів, науковців і спеціалістів з автоматичного опрацювання інформації природною мовою.



Проектна робота 2. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Дослідіть сучасний - на момент вивчення Вами курсу - стан української комп'ютерної лінгвістики. Проаналізуйте еволюцію комп'ютерної лінгвістики в Україні. Підготуйте і проведіть мультимедійну презентацію на основі статті і Вашого аналізу. Висловіть і обґрунтуйте власну думку про стан, потреби і перспективи розвитку комп'ютерної лінгвістики в Україні.

Балог, В. (2005). Сучасний стан української комп'ютерної лінгвістики. *Лексикографічний бюлетень*, 11, 28-35.



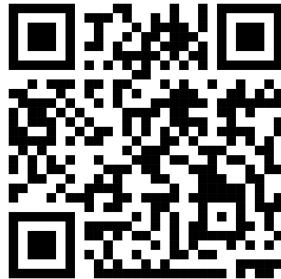
Анотація. У статті розглянуто стан української комп'ютерної лінгвістики на сучасному етапі розвитку мовознавчої науки. Оглядом подано інформацію про наукові установи, де працюють у галузі комп'ютерної лінгвістики, й основні здобутки науковців у цій галузі, а також коротко окреслено перспективи її розвитку.

Ключові слова: мовознавство, минуле і сучасність, перспективи, українська комп'ютерна лінгвістика.



Проектна робота 3. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Гирин, О.В. (2020). Основні проблеми систем обробки природної мови. *Філологічні студії. Збірник наукових праць*, 14, 41-45.



***Анотація.** Стаття присвячена обробці природної мови, а саме автоматичній синтаксичній обробці англійських речень. Висвітлено проблеми, спричинені цим процесом, що пов'язані з графічною, семантичною та синтаксичною неоднозначністю. Визначено шляхи вирішення цих проблем, зумовлених застосуванням автоматичного синтаксичного аналізу, та яким чином такі методи аналізу можуть бути корисні для розробки його нових алгоритмів. Дослідження зосереджене на питаннях, які унеможливають основу обробки природної мови - парсинг. Це процес аналізу речень за їх структурою, змістом і значенням, метою якого є визначення граматичної структури речення, розподіл речень на складові компоненти і визначення зв'язків між ними.*

***Ключові слова:** синтаксичний аналіз, обробка природної мови, статистичне машинне навчання, неоднозначність.*



Проектна робота 4. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Шомко, А.В. (2017). Генеративна граматики як галузь лінгвістичної науки. Проблеми та перспективи її розвитку. *Сучасні філологічні дослідження та навчання іноземної мови в контексті міжкультурної комунікації*, 10, 355-359.



Анотація. У статті проаналізовано етапи становлення та розвитку генеративної граматики Н. Хомські, окреслено її основні проблеми, завдання та здобутки з можливістю практичного застосування у сучасному світі.

Ключові слова: генеративізм, синтаксична структура, генеративна граматики, породжувальний механізм, стандартна теорія, мінімалістська програма.



Проектна робота 5. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Комарницька, О. (2018). Методи автоматизованого семантичного аналізу природномовної інформації. *Філологічний дискурс*, 7. 92-100.



Анотація. У статті розглянуто особливості автоматизованого семантичного аналізу тексту, досліджено проблеми створення автоматизованих лінгвістично-програмних засобів, придатних для застосування в системах екстракції семантики з тексту. Автором систематизовано та здійснено порівняльний аналіз результатів наукових досліджень у галузі розробки моделей і методів семантичного аналізу природномовної інформації. Визначено два основні підходи в напрямку комп'ютерної обробки природномовних текстів: лінгвоаналітичний і статистичний. Автором аргументовано, що найбільш перспективними та ефективними з них є, відповідно, експліцитні методи семантичного аналізу текстової інформації (алгоритми онтологічного семантичного аналізу) та методи латентно-семантичного аналізу. Окреслено можливі шляхи удосконалення існуючих комп'ютерних засобів діагностування релевантності природномовної інформації; обґрунтовано, що найпопулярнішими методами обробки природномовної інформації з метою екстракції та репрезентації семантики мають бути системи, що ґрунтуються на ефективному поєднанні лінгвістичних технологій аналізу (графематичного, морфологічного, синтаксичного, семантичного), зокрема із застосуванням онтологій, та методу латентно-семантичного аналізу. Доведено, що інтеграція технологій експліцитного семантичного

аналізу, латентно-семантичного аналізу, методів теорії нечіткої логіки, штучного інтелекту та ін. є перспективним шляхом розв'язання проблеми автоматизованого семантичного аналізу природномовної інформації.

Ключові слова: *автоматизований семантичний аналіз, природномовна інформація, лінгвістичний аналіз, латентно-семантичний аналіз, метод, онтологія.*



Проектна робота 6. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 1.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 2

Машинний переклад

▷ Machine translation

Машинний переклад визначається як підгалузь комп'ютерної лінгвістики, у центрі уваги якої знаходиться використання програмного забезпечення для перекладу з однієї мови на іншу.

Однією з найстаріших систем машинного перекладу є SYSTRAN, яка була розроблена як система, заснована на правилах (англ. *a rule-based system*), і починаючи з 1960-х років широко використовувалася американськими та європейськими урядовими агентствами, а також Yahoo! Babel Fish та (до 2007 року) у Google Translate. У 2010 році її було гібридизовано зі статистичними методами машинного перекладу. Google Translate нині використовує фразовий (англ. *phrase-based*) машинний переклад, а англійська мова служить мовою-посередником (англ. *an interlingua*) для більшості мовних пар. Перекладач Bing Translator від Microsoft використовує аналіз структури залежностей (англ. *dependency structure analysis*) разом зі статистичним машинним перекладом. Інші розвинені системи машинного перекладу включають Asia Online та WorldLingo. Також існує безліч систем для невеликих мовних груп, наприклад, для перекладу між пенджабі та хінді (система Direct MT) або між кількома європейськими мовами (наприклад, OpenLogos, IdiomaX та GramTrans).

Машинні переклади, як і раніше, мають помилки, але їхня якість зазвичай достатня для того, аби читач зрозумів загальну суть оригінального тексту. У багатьох випадках, наприклад, при перегляді міжнародних веб-сторінок (застосування, яке навряд чи передбачалося протягом десятиліть досліджень у галузі машинного перекладу), знадобиться не більше того. Також, програми машинного перекладу на портативних пристроях, розроблені для допомоги міжнародним мандрівникам, можуть бути достатньо точними для обмежених цілей, наприклад, дізнатися дорогу або звернутися за екстреною допомогою, спілкуватися з транспортним персоналом, здійснювати покупки або бронювання. Коли ж потрібен високоякісний переклад, перевага завжди надається перекладачеві-людині: автоматичні методи можуть бути використані як допомога перекладачеві, але нетипові питання, які вимагають неабияких розумових зусиль, займають - як і раніше - значну частину часу перекладача (Schubert, 2020).



Завдання 14. Наведіть приклади сучасних систем машинного перекладу. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.²⁴

Узагальніть лінгвістичні аспекти створення і використання систем машинного перекладу.



Завдання 15. Оберіть дві-три системи машинного перекладу і здійсніть у них комп'ютерний переклад кількох фрагментів текстів різних жанрів українською та англійською або німецькою мовами на Ваш вибір.²⁵

Зробіть кількісний та якісний аналіз помилок у перекладених фрагментах. Підсумуйте свій досвід користування цими системами.

²⁴ Завдання середнього рівня складності.

²⁵ Завдання середнього рівня складності.

Проектні роботи до Розділу 2



Проектна робота 7. [Усі студенти групи є учасниками цієї проектної роботи]

Оберіть одну із запропонованих тем так, аби між усіма студентами групи та обраними темами встановилося взаємно однозначне відношення. На основі авторитетних джерел, дослідіть обрану Вами тему; дайте темі ємну і привабливу назву. Створіть постерну доповідь для обраної Вами теми. Підготуйтеся до її обговорення; будьте готові висловити і обґрунтувати власну думку.²⁶

1. Переклад: сутність; підходи до визначення; наукові традиції вивчення
2. Види перекладу
3. Засоби комп'ютеризації процесу перекладу
4. Машинний переклад: визначення; цілі і методи; прикладні аспекти
5. Витоки, еволюція і види систем машинного перекладу
6. Машинний переклад vs. автоматизований переклад
7. Машинний переклад vs. інтерактивний переклад
8. Лінгвістична якість машинного перекладу і підходи до її оцінювання
9. Провідні підходи до створення систем машинного перекладу
10. Системи машинного перекладу, базовані на правилах²⁷
11. Статистичні системи машинного перекладу²⁸
12. Системи машинного перекладу, базовані на прикладах²⁹

²⁶ Семінар(и) намічено як постер-конференцію, де доповіді, що із ними попередньо мають ознайомитися усі учасники, обговорюватимуться у форматі круглого столу.

²⁷ англ. *Rule-Based Machine Translation*.

²⁸ англ. *Statistical Machine Translation*.

²⁹ англ. *Example-Based Machine Translation*.

13. Гібридні системи машинного перекладу³⁰
14. Системи машинного перекладу, базовані на штучних нейронних мережах³¹
15. Основні виклики у царині машинного перекладу
16. Створення онтологій³² для систем машинного перекладу
17. Переклад із багатопаралельних джерел³³
18. Використання (систем) машинного перекладу у навчанні
19. Машинний переклад і мова жестів
20. Машинний переклад і авторське право
21. Авторитетні сучасні системи машинного перекладу
22. Майбутнє (систем) машинного перекладу
23. *Запропонуйте власну - доладну і варту - тему*

³⁰ англ. *Hybrid Machine Translation*.

³¹ англ. *Neural Machine Translation*.

³² Див. також (Vakhovska, 2021).

³³ англ. *Translation from Multiparallel Sources*.



Проектна робота 8. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Шипнівська, О. (2013). Визначення типів синтаксичної неоднозначності у знаннеорієнтованій системі машинного перекладу. *Українське мовознавство*, 43, 104-113.



Анотація. Розглянуто теоретико-практичні засади визначення синтаксично неоднозначних конструкцій в автоматичному синтаксичному аналізі. На матеріалі текстів військової тематики проаналізовано основні типи синтаксичної омонімії.

Ключові слова: автоматичний синтаксичний аналіз, знаннеорієнтована система машинного перекладу, синтаксична неоднозначність, типи синтаксичних омонімів, синтаксична структура.



Проектна робота 9. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 2.
Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 3

Пошук і кластеризація документів

▷ Document retrieval and clustering

Інформаційний пошук, що охоплює як пошук структурованих даних, які знаходяться в реляційних базах даних (англ. *relational databases*), так і неструктурованих даних, уміщених у текстових документах, вже давно став однією із центральних тем інформаційної науки (наприклад, див. Salton, 1989).

Критерії пошуку цих двох типів даних є між собою пов'язаними, оскільки як структуровані, так і неструктуровані дані часто вимагають *пошуку, орієнтованого на зміст* (англ. *content-directed retrieval*). Наприклад, якщо користувачі бази даних співробітників компанії можуть захотіти отримати записи про співробітників за їхнім унікальним іменем або ідентифікатором, то в інший час вони можуть захотіти отримати записи про всіх співробітників у певній категорії зайнятості, можливо, з додатковими обмеженнями, як-от потрапляння до певного діапазону зарплат. Цей пошук здійснюється за допомогою “інвертованих файлів” (англ. “*inverted files*”), які, по суті, індексують об'єкти пошуку за їхніми атрибутами та значеннями (англ. *attributes and values*), а не за їхніми ідентифікаторами (англ. *identifiers*). Так само, текстові документи можуть бути знайдені за деякою унікальною міткою, або ж відповідно до *релевантності* (англ. *relevance*) цих документів певному запиту або заголовку теми. У своєму найпростішому вигляді, поняття релевантності визначається як те, що документи мають містити терміни (слова або короткі словосполучення) запиту. Однак термінам, які є характерними для документа, слід надавати у пошуку більшої важливості. Тому стандартною мірою релевантності для конкретного терміну запиту є *чт-зчд* (*частота терміну - зворотна частота документа*) для цього терміну (англ. *tf-idf* (*term frequency - inverse document frequency*)), яка збільшується (наприклад, логарифмічно) з частотою появи терміну в документі, але зменшується тією мірою, наскільки часто цей термін зустрічається в наборі документів загалом. Підсумовування *чт-зчд* для окремих термінів запиту дозволяє визначити релевантність документа конкретному запиту.

Недоліками цього методу є, по-перше, те, що він недооцінює збіги термінів, якщо кожен термін часто зустрічається в наборі документів (наприклад, для запиту “палички та колбочки ока” збіги для “палички,” “колбочки” та “ока” можуть також характеризувати релевантні документи, хоча всі три терміни зустрічаються досить часто у нефізіологічних контекстах), і,

по-друге, те, що релевантні документи можуть містити невелику кількість збігів термінів запиту, але при цьому містити значну кількість семантично пов'язаних термінів. Деякі з векторних методів (англ. *vector methods*), традиційно згадуваних у зв'язку із кластеризацією документів, можуть бути використані для усунення цих недоліків. Так, можна зменшити розмірність векторного простору на основі термінів за допомогою латентного семантичного аналізу (англ. *latent semantic analysis*), отримавши набагато менший “діапазон понять,” в якому кілька термінів, які, як правило, разом зустрічаються в документах, будуть об'єднані в один вимір (англ. *dimension*). Отже, обмін вимірами, а не конкретними термінами, стає у такому випадку основою визначення релевантності.

Кластеризація документів необхідна, коли велику кількість документів потрібно систематизувати й організувати, часто в ієрархічній формі, для легкого доступу до окремих тематично пов'язаних елементів, наприклад, у зібраннях патентних описів або історій хвороб, в архівах рефератів, юридичних прецедентів або підписаних зображень. Кластеризація також корисна в дослідницькому аналізі даних (наприклад, при визначенні частотності окремих токенів слів у невідомій мові) й опосередковано підтримує різні прикладні напрямки у царині обробки природної мови, адже може використовуватися для оптимізації мовних моделей, наприклад, у створенні кластерів слів, потрібних для конкретизації пошукового простору у випадках наявності лише неповних даних.

Кластеризація широко використовується в інших областях, як-от у біологічних і медичних дослідженнях, епідеміології, маркетингових дослідженнях і таргетованій рекламі, психолого-педагогічних дослідженнях, при аналізі соціальних мереж, геологічному аналізі тощо.

Пошук документів і їхня кластеризація часто служать попередніми кроками у завданнях екстракції інформації (англ. *information extraction*) або текстового майнінгу (англ. *text mining*) - двох суміжних областях, що займаються екстракцією потрібної інформації з документів: основних характеристик іменованих об'єктів (категорія, роль щодо інших об'єктів, місцезнаходження, часові характеристики тощо), певних типів подій, - й виведенням правил, за якими встановлюється семантичний зв'язок між термінами (наприклад, правило, що описує залежність, у якій покупка одного типу продукту корелює з покупкою іншого) (Schubert, 2020).



Завдання 16. Наведіть приклади сучасних систем пошуку і кластеризації документів. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.³⁴

Узагальніть лінгвістичні аспекти створення і використання систем пошуку і кластеризації документів.

³⁴ Завдання середнього рівня складності.

Проектні роботи до Розділу 3



Проектна робота 10. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Вавіленкова, А. (2017). Структура системи порівняльного аналізу електронних текстових документів за змістом. *Вісник Національного університету “Львівська політехніка.” Комп’ютерні науки та інформаційні технології*, 864(1), 277-285.



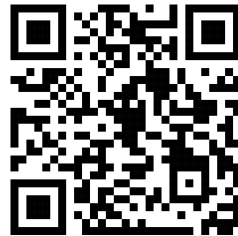
Анотація. Проаналізовано схему функціонування сервісів для визначення унікальності електронних текстових документів, розглянуто їхні основні характеристики під час перевірки на оригінальність наукової статті. Наведено структуру системи порівняльного аналізу електронних текстових документів за змістом, описано принцип функціонування кожної з її основних компонент.

Ключові слова: природна мова, порівняльний аналіз, пошук, логіко-лінгвістична модель, база знань.



Проектна робота 11. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Нога, Р., & Шаховська, Н. (2011). Аналітичний огляд методів та засобів опрацювання текстової інформації. *Вісник Національного університету "Львівська політехніка."* Інформаційні системи та мережі, 715(2), 323-332.



Анотація. У статті проаналізовано проблеми опрацювання розрізнених текстових даних. Виділено задачі опрацювання текстових даних. Показано, що наявне математичне та програмне забезпечення недостатнє для одночасного розв'язання задач опрацювання множинних текстових ресурсів.

Ключові слова: ключове слово, індексування, пошук.



Проектна робота 12. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 3.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 4

Екстракція та узагальнення інформації

▷ Information extraction and summarization

Екстракція інформації та узагальнення змісту неструктурованих даних стають усе важливішими серед прикладних аспектів комп'ютерної лінгвістики, враховуючи велику кількість текстів, продукованих засобами масової інформації, організаціями усіх видів та окремими людьми.

Безперервність і постійне збільшення потоку інформації ускладнюють визначення у цьому потоці і виокремлення із нього тих елементів, що належали б певній темі, задовольняючи певний користувачський запит: даних про людей, інформації про організації та споживчі товари, про діяльність з обслуговування та ремонту продукції, про результати медичних досліджень, або про подробиці аварій, землетрусів, злочинів, корпоративних поглинань тощо.

Один із методів, який широко використовується як при екстракції інформації, так і при певних видах “механічно завченого” її узагальнення (англ. *“rote” summarization*), базується на використанні шаблонів екстракції (англ. *extraction patterns*), розроблених у відповідності до структури мовних виразів, що зазвичай використовуються мовцями для передачі інформації у спілкуванні. Наприклад, текстові корпуси або стрічки новин можуть бути використані для пошуку інформації про компанії шляхом підбору відомих назв компаній і термінів “Corp.,” “.com,” “headquartered at” і “annual revenue of,” а також відповідних частин мови і структурно-семантичних залежностей, і зіставлення шаблонів регулярних виразів із вилученими із корпусу сегментами тексту, які містять ці ключові слова й словосполучення або розташовані близько до них. Іншим прикладом може бути узагальнення повідомлень про землетруси, з яких можна отримати типову інформацію: дані про епіцентр землетрусу, його магнітуду за шкалою Ріхтера, час і тривалість події, постраждалі населені пункти, кількість загиблих, поранених, матеріальні збитки, наслідки у вигляді пожеж і цунамі тощо. Шаблони екстракції зазвичай націлені на певні атрибути об'єктів у заздалегідь визначених атрибутивно-значеннєвих рамках, або фреймах (англ. *predetermined attribute-value frames*) (наприклад, фрейм інформації про компанію або фрейм подробиць землетрусу), а заповнені фрейми можуть розглядатися як узагальнення самі по собі або використовуватися для створення похідних узагальнень. Першими системами такого типу були FRUMP (DeJong, 1982) та JASPER (Andersen et al., 1992). Серед сотень пізніших систем екстракції інформації особливо успішною було

визнано систему Fastus компанії SRI (Hobbs et al., 1997). Розвиток таких систем триває.

Звертаємо увагу, що те, чи буде система, заснована на шаблонах, розглядатися як система екстракції інформації або ж як система узагальнення інформації, залежить суто від того тексту, до якого ця система застосовується. Якщо вся інформація, що цікавить, зібрана в єдиний, розширений текстовий сегмент (як у випадку з повідомленнями про землетруси), то одержані знання можна розглядати як узагальнення сегмента. Якщо ж інформація вибірково екстрагується з різних речень, розкиданих у великих текстових масивах, і при цьому більшість матеріалу ігнорується як невідповідна цілям екстракції, то діяльність такої системи розглядатиметься як екстракція інформації, а не як її узагальнення.

Коли документ, який підлягає узагальненню, не можна віднести до якоїсь передбачуваної категорії, а його зміст структурований і виражений стереотипним чином, то узагальнення зазвичай виконується шляхом вибору та об'єднання “центральної речень” документа. Речення є центральним тою мірою, якою багато інших речень у документі подібні до нього з точки зору спільного змісту слів або із урахуванням складнішої міри подібності, наприклад, заснованої на метриці чт-зчд для термінів або косинусної метрики у векторному просторі зі зменшеною розмірністю (англ. *a cosine metric in a dimensionality-reduced vector space*). Таким чином, окремі речення розглядаються як ніби документи та визначаються кілька інших речень у тексті, “релевантність” яких стосовно цих речень максимальна. Утім, просте повернення послідовності центральної речень, як правило, не дасть адекватного узагальнення відповідного тексту. Наприклад, такі речення можуть містити займенники з невизначеними референтами або кореферентні елементи, референтів яких необхідно шукати в нецентральної реченнях. Крім того, центральні “речення” можуть виявитися частинами, вбудованими у структурно складніші речення, які міститимуть нерелевантну запитові додаткову інформацію. Необхідно застосовувати евристичні методи для виявлення та видалення зайвого матеріалу, а екстраговані фрагменти мають бути попередньо узгоджені й вже після цього об'єднані. В деяких випадках складні описи слід перефразувати, спростивши і скоротивши їх. Наприклад, речення “The tornado carried off the roof of a local farmhouse, and reduced its walls and contents to rubble” можна перефразувати як “The tornado destroyed a local farmhouse.”

Хоча деякі із цих проблем вже частково вирішені в сучасних системах узагальнення інформації, узагальнення, зроблене людиною, завжди характеризуватиметься набагато глибшим розумінням, ніж можна досягти у цей час у будь-яких штучних системах. Ще однією проблемою у цій галузі

(більшою, ніж у машинному перекладі) є оцінка узагальнення. Навіть людські судження дуже різняться, наприклад, залежно від більшої або меншої чутливості оцінювача до граматичних недоліків тексту порівняно із його чутливістю до змістових вад тексту (Schubert, 2020).



Завдання 17. Наведіть приклади сучасних систем екстракції та узагальнення інформації. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.³⁵

Виокреміть лінгвістичні аспекти створення і використання систем екстракції та узагальнення інформації.

³⁵ Завдання середнього рівня складності.

Проектні роботи до Розділу 4



Проектна робота 13. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Погорілий, С.Д., & Крамов, А.А. (2018). Автоматизована екстракція структурованої інформації з множини веб-сторінок. *Проблеми програмування. Науковий журнал*, 2(3), 149-158. <http://dx.doi.org/10.15407/pp2018.02.149>



Анотація. У статті обґрунтовано доцільність використання методів екстракції структурованих даних з множини HTML-сторінок для здійснення інформаційного пошуку в мережі Internet. Проаналізовано основні методи екстракції структурованих даних з множини веб-сторінок, які сформовані спільним сценарієм, але різними наборами даних. Розглянуто класифікацію методів за ступенем автоматизації (фактору впливу користувача) процесу формування шаблону. Детально описано принципи роботи основних неконтрольованих методів (Roadrunner, FiVaTech, Trinity), розглянуто їхні переваги та недоліки. Обґрунтовано доцільність використання методу Trinity для екстракції даних порівняно з іншими методами. Показано проблему вибору вхідних документів методу серед множини HTML-сторінок для формування узагальненого шаблону. Проведено експериментальну перевірку методу Trinity на множині HTML-сторінок англомовних статей українських наукових журналів. Для формування тестової множини HTML-сторінок виконано автоматизований обхід веб-сайтів журналів за допомогою пошукового роботу. Реалізацію пошукового роботу здійснено за рахунок обробки об'єктної моделі HTML-документів, отриманих з веб-сайтів.

Шаблони (регулярні вирази), сформовані методом Trinity, застосовано до всього набору вхідних HTML-сторінок. Результати екстракції - структуровані дані про статті (назва, автори, анотація, ключові слова) - експортовано до бази даних з можливістю їх подальшого аналізу. Здійснено порівняння отриманих результатів з даними про статті, одержаними за допомогою аналізу об'єктної моделі веб-сторінок власноруч. Обраховано похибку використання методу Trinity на експериментальній множині HTML-сторінок.

Ключові слова: *екстракція даних; методи екстракції; класифікація методів екстракції; метод Trinity; тернарне дерево документу; префіксний обхід дерева; об'єктна модель HTML-сторінки; автоматизоване збирання веб-сторінок; пошуковий робот; шаблон HTML-сторінки; формування регулярного виразу.*



Проектна робота 14. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Погорілий, С.Д., & Крамов, А.А. (2018). Метод розрахунку когерентності українського тексту. *Реєстрація, зберігання і обробка даних. Інформаційно-аналітичні системи обробки даних*, 20(4), 64-75. <http://dx.doi.org/10.35681/1560-9189.2018.20.4.178945>



Анотація. У статті проаналізовано основні методи автоматизованої оцінки когерентності тексту, який написано природною мовою. Запропоновано вдосконалення методу графа семантичної схожості за допомогою попередньої підготовки моделі, а саме здійснення навчання нейронної мережі векторного представлення речень. Проведено експериментальну перевірку роботи методу графа семантичної схожості та його модифікованих версій на множині україномовних статей наукових журналів різної тематики. Ефективність роботи методу та його модифікацій розраховано за допомогою вирішення типових задач оцінки когерентності тексту: задач розрізнення документів і вставки. На основі отриманих результатів визначено найбільш ефективні модифікацію та параметри методу графа семантичної схожості для оцінки когерентності україномовних текстів.

Ключові слова: обробка природної мови, когерентність тексту, граф семантичної схожості, нейронна мережа, модель Doc2Vec, семантична міра схожості речень.



Проектна робота 15. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 4.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 5

Сентимент-аналіз

▷ Sentiment analysis

Сентимент-аналіз визначається як виявлення позитивного чи негативного ставлення (або, більш специфічно, довіри чи зневаги) з боку, наприклад, авторів статей або блогів до комерційних продуктів, фільмів, організацій, осіб, ідеологій тощо.

Сентимент-аналіз розвинувся у дуже активну сферу прикладної комп'ютерної лінгвістики через ту неабияку важливість, що у сучасному суспільстві набули маркетинг продуктів, аналіз соціальних мереж, політичні та розвідувальні дані, визначення психологічних типів особистостей або діагностування психічних розладів на основі написаних особистістю текстів тощо. Використовувані методи зазвичай базуються на лексиконах сентиментів (англ. *sentiment lexicons*), у яких класифіковано афективну полярність словникових одиниць, а також на контрольованому машинному навчанні (англ. *supervised machine learning*), застосовуваному до текстів, із яких були екстраговані ціннісні ознаки слів й словосполучень і які були розмічені вручну як такі, що виражають позитивне або негативне ставлення мовців. Замість ручної розмітки текстів іноді можуть використовуватися вже готові дані для класифікації потрібної інформації. Наприклад, середні оцінки споживчих товарів або фільмів, надані блогерами, можуть бути використані для того, аби навчити алгоритм машинного навчання класифікувати схожі матеріали без оцінок. Афективні лексичні категорії та контрастні ставлення можуть, дійсно, бути вивчені на основі таких даних.

Так, наприклад, часта зустрічність словосполучень “great movie,” “pretty good movie” або “terrible movie” в блогах, присвячених фільмам, з високими, середніми та низькими середніми оцінками, може свідчити про те, що “great,” “pretty good” і “terrible” належать до спектра контрастів від дуже позитивної до дуже негативної полярності оцінки. Знання цих термінів, у свою чергу, може збільшити здатність лексиконів сентиментів охоплювати загальну лексику. Проте сентимент-аналіз, заснований на ціннісних ознаках, умішених у значеннях окремих слів і словосполучень, має очевидні недоліки, як-от ігнорування сарказму і іронії (“This is the most subtle and sensitive movie since The Texas Chainsaw Massacre”), цитування думок, що суперечать авторським (“According to the ads, Siri is the greatest app since iTunes, but in fact ...”), і недостатнє розуміння імплікатур (“You'll be much better off buying a pair of

woolen undies for the winter than purchasing this item”). Тому дослідники намагаються інтегрувати аналіз, заснований на знаннях, і семантичний аналіз (англ. *knowledge-based and semantic analysis*) із поверхневим сентимент-аналізом слів і словосполучень (англ. *superficial word- and phrase-based sentiment analysis*) (Schubert, 2020).



Завдання 18. Наведіть приклади сучасних систем сентимент-аналізу. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.³⁶

Узагальніть лінгвістичні аспекти створення і використання систем сентимент-аналізу.

³⁶ Завдання середнього рівня складності.

Проектні роботи до Розділу 5



Проектна робота 16. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Романюк, А., Романишин, М. (2013). Тональний словник української мови на основі сентимент-анотованого корпусу. *Українське мовознавство*, 43, 63-75.



Анотація. Докладно розглянуто всі етапи створення сентимент-анотованого корпусу українськомовних відгуків і тонального словника на його основі.

Ключові слова: емоційно-смісловий аналіз, тональний словник, сентимент-анотований корпус, засоби для анотування текстів.



Проектна робота 17. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Ялова, К.М., Яшина, К.В., Говорущенко, Т.О., & Тарасюк, О.С. (2021). Сентимент аналіз засобами нейронної мережі. *Математичне моделювання*, 1(44), 30-37.



Анотація. У статті здійснено постановку завдання аналізу тональності вхідної текстової інформації, яке відноситься до розділу прикладної лінгвістики та обробки природної мови. Розроблено двонаправлену нейронну мережу з довгою короткотривалою пам'яттю для розв'язання завдання сентимент аналізу. Обґрунтовано доцільність застосування додаткового шару для нейронної мережі з умовно випадковими полями. Для проведення навчання нейронної мережі застосовано корпус текстових повідомлень з соціальної мережі. Описано результати навчання, валідації та тестування розробленої нейронної мережі. Для оцінювання якості розпізнавання сентиментів застосовано метрики повноти (*precision*), точності (*recall*) та збалансованої міри *F1*. Найкращі значення розпізнавання на тестовому наборі даних були отримані для позитивного сентименту і склали *precision* = 61,92 %, *recall* = 69,21 %, *F1* = 65,36 %.

Ключові слова: сентимент аналіз, двонаправлена нейронна мережа з довгою короткотривалою пам'яттю, умовно випадкові поля.



Проектна робота 18. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 5.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 6

Чат-боти та агенти підтримки діалогу

▷ Chatbots and companionable dialogue agents

Сучасні чат-боти є нащадками ELIZA, створеної Джозефом Вейценбаумом у 1966, і зазвичай використовуються (часто з анімованим персонажем “голова, що говорить”) із розважальною метою або для збудження інтересу у відвідувачів комерційних веб-сайтів із метою підвищення конверсії сайту і утримання клієнтів.

Чат-боти можуть бути оснащені великими, створеними вручну скриптами (схеми питання-відповідь з індексом ключових слів (англ. *keyword-indexed input-response schemas*)), які дозволяють їм відповідати на прості питання про компанію та її продукцію, і надають деяку здатність відповідати на питання довільної тематики й обмінюватись привітаннями і люб'язностями. Менш безпечне застосування - використання чат-ботів, які видають себе за відвідувачів сайтів соціальних мереж або інтерактивних ігор, з метою отримання приватної інформації від введених таким чином в оману користувачів, або з метою рекомендації користувачам певних веб-сайтів або продуктів. Аби запобігти входу ботів, велика кількість сайтів соціальних мереж й інших сайтів, на яких потенційно можуть бути націлені боти, стали використовувати CAPTCHA.

Агенти підтримки діалогу (англ. *companionable dialogue agents*) (або *реляційні агенти* (англ. *relational agents*)) і досі значною мірою спираються на принципи роботи чат-ботів, тобто на вхідні шаблони, що їх очікують від користувачів, та на відповідні вихідні дані, що їх надає такий агент. Утім, мета розробників полягає в тому, щоб вийти за рамки цих методів, створюючи агентів (часто з головами, що говорять, або у вигляді анімованих персонажів), які володіли б особистісними характеристиками та здатністю виявляти емоції та співчуття, мали б семантичну та епізодичну пам'ять, аби, надаючи послуги, тримати інформацію про конкретного користувача у довгостроковій перспективі. Крім спілкування та підтримки, чат-бот може надавати й інші послуги, як-от поради в деяких сферах життя (наприклад, здоров'я та фітнес), ведення розкладу, нагадування, відповіді на запитання, навчання (наприклад, навчання іноземних мов), ігри та інтернет-послуги. Дослідник Йорік Уїлкс припустив, що в ідеалі такі персонажі нагадуватимуть “вікторіанських супутників” (англ. “Victorian companions”) користувача і матимуть такі риси, як

ввічливість, гідність, скромність, життєрадісність та поінформованість (Wilks, 2010).

Утім, такі цілі ще далекі від досягнення, оскільки у штучних системах розпізнавання й розуміння мови здатність до умовиводів та навчання ще розвинена недостатньо. Наприклад, система HWYD (“How Was Your Day”), описана у роботі (Pulman et al., 2010), свого часу отримала приз за кращу демонстрацію на конференції з автономних агентів. Обробка природної мови у цій системі відносно складна і, у порівнянні із іншими подібними системами, витончена. Неглибока синтаксична та семантична обробка (англ. *shallow syntactic and semantic processing*) використовується для пошуку екземплярів близько 30 “шаблонів подій,” як-от “argument at work between *X* and *Y*” або “meeting with *X* about *Y*.” Процес інтерпретації у системі спирається на репрезентацію інформації, підтримувану менеджером діалогу, і включає визначення референтів мовних виразів і відновлення початково опущених фрагментів речень. Цілі, що генеруються менеджером діалогу, уможливають відповіді за допомогою модуля планування, який створює екземпляри і визначає послідовність відповідей. Автори стверджують, що система здатна підтримувати зв’язні діалоги тривалістю понад 20 хвилин.

Розробляються й системи зовсім іншого типу, націлені, наприклад, на клінічно обґрунтоване консультування користувачів із питань здоров'я. Так, системи, описані в роботі (Bickmore et al., 2011), спираються на велику, ретельно розроблену формалізацію клінічно перевірених стратегій консультування та відповідних медичних знань, виражених у мові дескриптивної логіки Web Ontology Language (англ. *a description logic language*; OWL) і мові опису завдань, орієнтованих на досягнення мети (англ. *a goal-directed task description language*). Такі системи довели свою ефективність й їхню роботу можна у певній мірі порівняти з роботою медичних консультантів-людей. Утім, ці діалоги розгортаються за планом (англ. *plan-driven*) і, зрештою, складаються із закриптованих висловлювань системи в парі зі списками відповідей з кількома варіантами вибору, які пропонуються клієнту.

Отже, системи підтримки діалогу залишаються дуже обмеженими в тому, що стосується тем діалогу, які вони можуть опрацьовувати, а також їхнього розуміння природної мови і здатності привнести в розмову фонові знання, не кажучи вже про те, щоб використовувати ці знання інтуїтивно або робити умовиводи на їхній основі (Schubert, 2020).



Завдання 19. Наведіть приклади сучасних чат-ботів та агентів підтримки діалогу. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.³⁷

Узагальніть лінгвістичні аспекти створення і використання чат-ботів та агентів підтримки діалогу.

³⁷ Завдання середнього рівня складності.

Проектні роботи до Розділу 6



Проектна робота 19. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Бурдаєв, В.П. (2020). Використання чат-бота @es_economy_karkas_bot для онлайн консультації з експертною системою. *Системи обробки інформації*, 1(160), 100-106. <https://doi.org/10.30748/soi.2020.160.13>



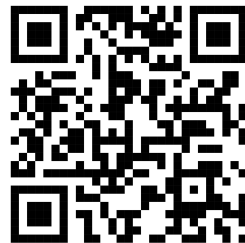
Анотація. *Чат-бот (співрозмовник) - це програма, яка імітує людське спілкування на основі елементів штучного інтелекту. У статті представлені результати інтегрування чат-бота @es_economy_karkas_bot з експертною системою для організації консультування в режимі онлайн. Дано опис архітектури і реалізація імплементації чат-бота месенджера телеграм в експертну систему на базі системи "КАРКАС" - інструментальний засіб для побудови моделей баз знань. Розглянута структура і алгоритм взаємодії чат-бота і агентів експертної системи в онлайн режимі. Виконано аналіз можливостей створення чат-ботів в месенджері ТЕЛЕГРАМ, їх інтеграцію з експертами системи в сфері економіки.*

Ключові слова: *месенджер, чат-бот, Telegram Messenger, агенти повідомлення, база знань, експертна система.*



Проектна робота 20. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Лавренчук, С.В., & Товстенюк, Б.С. (2020). Інформаційний чат-бот для сервісу обміну повідомленнями Telegram в навчальній сфері. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*, 41, 180-185. <https://doi.org/10.36910/6775-2524-0560-2020-41-28>



Анотація. У статті розглянуто чат-бот як засіб сучасної комунікації, окреслено можливості Telegram-ботів. Створено чат-бот для зручного інформування студентів, сформовано графіки активності боту, досліджено актуальність запропонованих функцій.

Ключові слова: Чат-бот, Telegram, Pandas, Matplotlib, сервіс обміну повідомленнями.



Проектна робота 21. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 6.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 7

Віртуальні світи, ігри та інтерактивна фантастика

▷ Virtual worlds, games, and interactive fiction

Текстові пригодницькі (квестові) комп'ютерні ігри, як-от Dungeons and Dragons, Hunt the Wumpus (в оригінальній версії) та Advent, почали розроблятися на початку та в середині 1970-х років і зазвичай містили текстові описи обставин і викликів, з якими стикався гравець, дозволяючи гравцеві вводити прості команди для вибору доступних дій (наприклад, “відкрити ящик,” “взяти меч” або “прочитати записку”). Хоча описи ситуацій (часто супроводжувані картинками) могли бути досить докладними, варіанти введення, доступні гравцеві, були обмеженими простими висловлюваннями, які програмісти гри могли передбачити та зібрати під час попереднього тестування гри і відповіді для яких можна було підготувати вручну.

Безумовно, гнучкіше використання природної мови (“відбивайся від гремліна мечем!” або “Якщо я дам тобі золото, ти відкриєш мені браму?”) пожвавило б взаємодію гравця з ігровим світом і персонажами в ньому. У 1980-х і 90-х роках текстові ігри поступилися місцем іграм, заснованим переважно на графіці та анімації: за ці роки зросла онлайн-спільнота гравців, яка стимулювала розвиток ефективного програмного забезпечення для розробки інтерактивної фантастики. Найпопулярнішою програмою (у 2000 році) була “Galatea” Емілі Шорт, яка дозволяла вести діалог з анімованою скульптурою, але все ж таки була ретельно заскриптованою й допускала з боку користувачів лише ті введення, які відповідали одній із запрограмованих розробниками відповідей. Багато ігор цього жанру також, подібно до чат-боту, використовують шаблони введення-виведення й у такий спосіб набувають певного ступеня стійкості до довільних дій або інструкцій гравців.

Найпопулярнішими комп'ютерними відеоіграми в 1990-х роках і в наступні роки були Myst (пригодницька гра від першої особи, розроблена Робіном і Рендом Міллерами) і гра-симулятор The Sims від Maxis Software. Myst, хоча і залучала повідомлення в книгах і щоденниках, була в основному грою без слів, а головний розробник The Sims Уїлл Райт вирішив проблему діалогу природною мовою у грі, змусивши жителів SimCity розмовляти сімською (англ. *Simplish*), мовою-нісенітницею, яка включає елементи української, французької та тагальської мов.

Сьогодні, комерційні пригодницькі ігри та візуальні романи, як і раніше, покладаються на скриптові дерева діалогів (англ. *scripted dialogue trees*), які

пропонують розгалужені альтернативні напрямки, в яких може розгорнутися діалог, за технологією, подібною до тієї, що лежить в основі ELIZA. Складніші підходи до взаємодії гравців і споживачів фантастики із віртуальними персонажами розробляються у дослідницьких лабораторіях, як-от у Центрі моделювання та симуляції людини при Пенсільванському університеті (англ. *the Center for Human Modeling and Simulation at the University of Pennsylvania*) та в Інституті креативних технологій при Університеті Південної Каліфорнії (англ. *the USC-affiliated Institute for Creative Technologies*). Хоча діалоги в цих сценаріях все ще залежать від ретельно розроблених скриптів, інтерпретація висловлювань користувача залучає цілий ряд авторитетних методів розпізнавання мови (англ. *speech recognition*) і управління діалогом (англ. *dialogue management*) та низку алгоритмів умовиводу (англ. *reasoning*). Сучасні дослідження відбуваються на таких майданчиках, як IVA (*Intelligent Virtual Agents*), AIIDE (*AI and Interactive Digital Entertainment*) та AAMAS (*Autonomous Agents and Multiagent Systems*) (Schubert, 2020).



Завдання 20. Наведіть приклади сучасних віртуальних світів, ігор та інтерактивної фантастики. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.³⁸

Узагальніть лінгвістичні аспекти створення і використання віртуальних світів, ігор та інтерактивної фантастики.

³⁸ Завдання середнього рівня складності.

Проектні роботи до Розділу 7



Проектна робота 22. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Штанько, В.І. (2012). Віртуальний комунікативний простір і проблеми самоідентифікації особистості. *Вісник Харківського національного університету імені В.Н. Каразіна. Серія: Теорія культури і філософія науки*, 1029(1), Вип. 47, 5-12.



Анотація. У статті розглядаються особливості конструювання особистості та її самопрезентації в сучасному комунікативному просторі, створеному інформаційно-комп'ютерними технологіями. Особлива увага приділяється особливостям самопрезентації особистості в чатах, інтернет-спільнотах, блогах, мотивах формування віртуальної особистості та аналізу суперечливого впливу віртуального комунікативного простору на самоідентифікацію особистості.

Ключові слова: віртуальний комунікативний простір, самопрезентація особистості, віртуальна ідентичність, самоідентифікація.



Проектна робота 23. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Трач, Ю. (2017). VR-технології як метод і засіб навчання. *Освітологічний дискурс*, 3(18), 309-322.



Анотація. У статті звернено увагу на один з перспективних освітніх методів, запропонованих сучасними інформаційними технологіями, - віртуальну реальність, яка моделюється комп'ютером. Реальних досліджень віртуальної реальності в педагогіці здійснюється вкрай мало. Автори наявних праць зводять застосування технологій віртуальної освіти до використання електронних підручників і тестових оболонок, рідше мультимедійних матеріалів, в окремих випадках - комп'ютерних віртуальних симуляторів і тренажерів. Відтак, мета даної статті - узагальнення накопиченого досвіду вивчення можливостей використання VR-технологій в освітньому процесі. Наведено переваги і недоліки застосування VR-технологій в освіті, формати VR в освіті, а також кращі проекти застосування VR в навчальному процесі. Окреслено перспективи подальших досліджень.

Ключові слова: VR-технології; віртуальна реальність; засіб навчання; метод навчання.



Проектна робота 24. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 7.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 8

Користувацькі інтерфейси природною мовою

▷ Natural language user interfaces

Користувацькі інтерфейси природною мовою охоплюють велике різноманіття додатків, починаючи від текстових систем, що мінімально залежать від розуміння природної мови, і закінчуючи системами зі значними можливостями розуміння мови у письмовій або усній взаємодії із користувачем. До традиційних і сучасних сфер застосування користувацьких інтерфейсів належать такі:

Текстові відповіді на запитання

▷ Text-based question answering

Текстові відповіді на запитання практичні лише тією мірою, яку можна очікувати, коли на запитання є готові відповіді, сховані десь у текстових корпусах, до яких система за потреби звернеться у пошуку відповіді. Цей пошук у час все зростаючого Інтернет-контенту став набагато більш здійсненним, ніж кілька десятиліть тому, хоча запитання, як і раніше, мають бути прямими, фактичними (наприклад, “Who killed President Lincoln?”), а не такими, що вимагають умовиводу (наприклад, “In what century did President Lincoln live?”, не кажучи вже про “Approximately how many 8-foot 2-by-4s do I need to build a 4-foot high, 15-foot long picket fence?”).

Текстові відповіді на запитання починаються з класифікації запитань (наприклад, yes-no questions, who-questions, what-questions, when-questions тощо), потім слідує пошук інформації з визначеного типу запитання, після чого пошук звужується до абзаців та, нарешті, речень, які можуть містити відповідь. При послідовному звуженні пошуку зазвичай використовується зіставлення слів та їхніх семантичних ознак, а зрештою - зіставлення синтаксичних залежностей і ролей, й, можливо, обмежений умовивід на основі тексту для перевірки варіантів відповідей. Такий умовивід може, наприклад, використовувати знання про гіпероніми із WordNet, аби спробувати встановити, що певний варіант відповіді підтверджує істинність запитання у його декларативній формі. Оскільки обране речення може містити нерелевантний матеріал та анафори, залишається екстрагувати релевантний матеріал (який також може включати допоміжний контекст) і згенерувати правильно оформлену (англ. *well-formed*)

підходящу відповідь. Багато раних, до 1976 року, текстових систем відповідей на запитання обговорюються у роботі (Bourne, & Nahn, 2003). Пізніші огляди (наприклад, Maurya, 2004), як правило, включають вже весь спектр методів відповіді на запитання, що їх використовують здебільшого і сьогодні.

У відповідях на запитання відкритої тематики багато запитань зосереджуються на властивостях названих об'єктів, таких як дата народження, місце народження, рід занять та інші особисті дані відомих сучасних та історичних особистостей, місцезнаходження, власність та продукція різних компаній, факти про споживчі товари, географічні факти тощо. Для відповіді на такі запитання має сенс попередньо зібрати відповідні фактоїди (англ. *factoids*) у велику базу знань, використовуючи методи набуття знань (англ. *knowledge acquisition methods*), традиційно використовувані комп'ютерними лінгвістами. Прикладами систем, які містять велику кількість фактоїдів, є кілька систем, розроблених в Університеті Вашингтона, що зберігають фактоїди у вигляді текстових фрагментів, і різні системи, які об'єднують зібрані фактоїди в трійки RDF (Resource Description Framework), як-от ті, що містяться у DBpedia, True Knowledge і YAGO. Деякі з цих систем отримують свої знання не тільки шляхом екстракції відкритої інформації (англ. *open information extraction*) і таргетованої екстракції семантичних зв'язків (англ. *targeted relation extraction*), але і з таких джерел, як "інфобокси" Вікіпедії та (контрольований) краудсорсинг. Тут поняття відповіді на запитання розширюється, оскільки деякі зі згаданих систем вимагають використання ключових слів або шаблонів запиту для пошуку фактоїдів.

Із точки зору звичайного користувача, залишається незрозумілим, які переваги мають побудовані у такий спосіб бази знань, враховуючи надзвичайну здатність Google та інших пошукових систем швидко давати відповіді навіть на такі запитання, як "Which European countries are landlocked?" (введено без лапок - з лапками, Google знаходить найкращу відповідь, використовуючи True Knowledge) або "How many Supreme Court justices did Kennedy appoint?". Утім, і Google, і Microsoft у свій час запустили великі "графи знань" (англ. "*knowledge graphs*"), що містять тисячі зв'язків між сотнями мільйонів об'єктів. Їхня мета - надати прямі відповіді (а не просто екстраговані фрагменти веб-сторінок) на терміни запиту та на запитання природною мовою, а також зробити умовиводи про ймовірні наміри користувачів, наприклад, про купівлю певного товару чи послуги.

Інтерфейси бази даних

▷ Database front-ends

Інтерфейси природною мовою для баз даних вже давно, починаючи із таких систем, як LUNAR (Woods et al., 1972) і REL (Thompson et al., 1969; Thompson, & Thompson, 1975), вважаються привабливим застосуванням технологій обробки природної мови.

Ця привабливість полягає в тому, що пошук та оперування інформацією, отриманої із реляційної (або іншої одноманітно структурованої) бази даних, описується існуючою мовою запитів для баз даних, що різко обмежує типи запитань природною мовою, які можна очікувати від користувача, наприклад, запитання, спрямовані на пошук об'єктів або груп об'єктів, що задовольняють заданим реляційним обмеженням, або надання їхніх істотних або визначальних властивостей (найдовші річки, найнижчі витрати тощо). Також, значно спрощується процес інтерпретації запитання та надання відповіді на нього, оскільки питомі логічні форми - формальні запити до бази даних - мають заздалегідь визначений, точний синтаксис та виконуються автоматично системою управління базою даних, залишаючи лише роботу з відображення обчислених результатів у відповідній мовній, табличній чи графічній формі.

Із того часу було створено велику кількість систем, призначених для різноманітних сфер людського життя, як-от судноводіння (Ladder: Hendrix et al., 1978), землекористування (Damerau, 1981), географія (Chat-80: Pereira, & Warren, 1982), пошук записів про компанії та продукти для страхових компаній, нафтових компаній, виробників, роздрібних торговців, банків тощо (Intellect: Harris, 1984), збір статистичних даних про клієнтів, послуги, активи компанії (Cercone et al., 1993) та багато інших (наприклад, див. Androutsopoulos, & Ritchie, 2000). Утім, комерційний вплив таких систем залишався незначним, оскільки їм, як правило, не вистачало надійності та деяких функціональних можливостей традиційного доступу до баз даних.

Інференційні (засновані на знаннях) відповіді на запитання

▷ Inferential (knowledge-based) question answering

Деякими обмеженими можливостями умовиводу, або інференції, у текстових системах відповідей на запитання та інтерфейсах природною мовою для баз даних є здатність підтверджувати зв'язки між варіантами відповідей та запитаннями, використовуючи прості види семантичних зв'язків між

відповідними термінами, та здатність сортувати чи класифікувати набори даних з баз даних та обчислювати середні значення або навіть створювати статистичні діаграми.

Утім, такі обмежені, спеціалізовані методи інференції не зовсім відповідають тому виду загальних суджень, що ґрунтуються на символічних знаннях і що вже деякий час визначаються розробниками як мета у створенні систем штучного інтелекту, здатних відповідати на запитання. Однією з ранніх спроб створити справді інференційну систему відповідей на запитання був проєкт ENGLAW Л. Стівена Коулза (Coles, 1972). ENGLAW був задуманий як прототип системи, що могла б використовуватися вченими та інженерами для отримання інформації про фізичні закони. Вона включала базу знань аксіом (викладених у мові логіки першого порядку) для 128 важливих фізичних законів, закодованих вручну за допомогою довідкового тексту. Запитання (наприклад, “In the Peltier Effect, does the heat developed depend on the direction of the electric current?”) записувалися мовою логіки за використання парсера та набору правил (викладених різними функціями мови Lisp), які порівнювали шаблони словосполучень природною мовою і записи, зроблені мовою логіки. Система не була розроблена до кінця, але те, як у ній інтегрувалися алгоритми формування суджень, технології обробки природної мови і методи вибіркового пошуку аксіом для інференційної відповіді на запитання постало помітним внеском у прикладну комп’ютерну лінгвістику.

Однією із пізніших, більш масштабних систем, орієнтованих на практичні цілі, була система JANUS компанії BBN (Ayuso et al., 1990). Вона призначалася для управління військово-морськими боями та могла відповідати на запитання про місцезнаходження, готовність, швидкість та інші атрибути кораблів, враховуючи зміну часових параметрів і можливість перебування об’єктів у можливих світах.

Ще одним прикладом подібних систем постав веб-двигун відповідей (англ. *web-based answer engine*) Wolfram|Alpha (або WolframAlpha), розроблений компанією Wolfram Research у складі 15 мільйонів рядків коду Mathematica для баз даних, моделей та алгоритмів у тисячах різних областей. (Mathematica - це математично орієнтована мова програмування високого рівня, розроблена британським ученим Стівеном Вольфрамом). Двигун орієнтований в основному на кількісні запитання англійською мовою (наприклад, “What is the GDP of France?” або “What is the surface area of the Moon?”) і часто надає графіки та діаграми разом із прямими відповідями. Утім, розуміння англійської мови у Wolfram|Alpha на момент написання цитованої статті не є особливо надійним. Наприклад, запитання “How old was Lincoln when he died?”, “At what age did Lincoln die?” та інші варіанти Wolfram|Alpha не зрозумів до кінця. Пов’язаний із

цим недолік полягає в тому, що кількісні навички Wolfram|Alpha не доповнені розвиненими навичками якісних суджень. Наприклад, “Was Socrates a man?” (знову ж таки, на момент написання статті) виводить на екран коротку інформацію про Сократа, включаючи зображення, але не дає прямої відповіді на запитання. Однак кількісні здібності Wolfram|Alpha цікаві не тільки в автономному режимі, але і корисні як доповнення для пошукових двигунів, як-от Microsoft Bing, і персональних голосових помічників, як-от Siri від Apple.

Ще однією системою відповідей на запитання є суперкомп'ютер Watson компанії IBM (Ferrucci, 2012; Ferrucci et al., 2010; Baker, 2011), який здобув широке визнання завдяки своїй перемозі у телевізійній вікторині Jeopardy!. Як і Wolfram|Alpha, Watson є у певному сенсі програмою грубої сили (англ. *a brute force program*), адже створювався групою 20 дослідників й інженерів-програмістів протягом трьох років і складається з близько мільйона рядків коду Java, C++, Prolog та інших мов. Програма виконує 3000 паралельних процесів на дев'яноста серверах IBM Power 750 і має доступ до 200 мільйонів сторінок контенту з таких джерел, як Wordnet, Wikipedia (і її структурованих похідних YAGO та DBpedia), тезаурусів, статей зі стрічок новин та літературних текстів, що становить кілька терабайт людського знання. (Це дорівнює приблизно 10^{10} клаузальних чанків - число, яке більше, ніж на 2 порядки перевищує кількість фактів, якими володіє будь-яка людина).

Замість того, щоб покладатися на якийсь один метод лінгвістичного чи семантичного аналізу або метод оцінки релевантності вилучених уривків та текстових “самородків” у них, Watson застосовує одночасно велику кількість методів до запитань та варіантів відповіді, включаючи методи класифікації запитань, виявлення ключових сутностей, парсингу, чанкування, лексичного аналізу, обчислення логічних форм, визначення референтів, виявлення семантичних зв'язків, часовий аналіз, а також спеціальні методи для пар запитання-відповідь, що включають каламбури, анаграми та інші зачіпки, поширені у Jeopardy!. Різні методи аналізу запитань використовуються окремо для отримання відповідних документів, а також для отримання, аналізу та оцінки потенційних відповідей з абзаців і речень у цих документах. У цілому, на одне запитання видаються численні варіанти відповідей, а їхній аналіз дає сотні характеристик, значущість яких для отримання ранжованих відповідей з відповідними рівнями достовірності визначається методами машинного навчання, які застосовуються до корпусу вже використовуваних у Jeopardy! запитань та відповідей. Watson зважує варіанти відповідей на основі рівнів їхньої достовірності, використовуючи складну регресійну модель.

Чи *розуміє* Watson поставлені йому запитання і ті відповіді, які він на них надає? Попри свої вражаючі результати в Jeopardy! у порівнянні із

гравцями-людьми, Watson мислить та розуміє англійську мову дуже обмежено. Він використовує той факт, що метою запитання Jeopardy! зазвичай є іменованій об'єкт, як-от *Jimmy Carter*, *Islamabad* або *Black Hole of Calcutta*. Watson здатен знайти кілька речень, у яких згадується цей об'єкт, і визначити, чи збігаються синтаксичні та семантичні властивості цих речень із властивостями запитання: у такий спосіб, він обирає правдоподібну відповідь, але реальне розуміння запитання у нього відсутнє. Наприклад, запитання із тем “нової історії” про президента, за якого США повністю визнали комуністичний Китай (Ferrucci, 2012), цілком може звестися до таких речень, як

Although he was the president who restored full diplomatic relations with China in 1978, Jimmy Carter has never visited that country ... (New York Times, June 27, 1981)

або

Exchanges between the two countries' nuclear scientists had begun soon after President Jimmy Carter officially recognized China in 1978. (New York Times, Feb. 2, 2001).

Хоча зв'язки між такими реченням та правильною відповіддю є непрямими (наприклад, залежать від співвіднесення *he* та *who* з *Jimmy Carter*, *restored diplomatic relations* - з *recognized*, а *Communist China* - з *China*), правильний аналіз цих зв'язків не є обов'язковою умовою успіху - достатньо, аби кластер речень, які віддають перевагу відповіді Jimmy Carter (через наявність у цих реченнях відповідних слів і словосполучень, а також кількох інших синтаксичних і семантичних ознак), надав більшу вагу саме цій відповіді у порівнянні із іншими конкуруючими кластерами. Таке комбінування статистичних показників навряд чи дорівнює тому осмисленню запитань, яке демонструють навіть першокласники, відповідаючи на прості питання, як-от “How do people keep from getting wet when it rains?” або “If you eat a cookie, what happens to the cookie?” Водночас ті величезні банки даних, що використовуються в Watson, можуть компенсувати недоліки інференції в інших системах. Так, наприклад, ІВМ активно розвиває Watson як ресурс для лікарів, у перспективі здатний надати нові можливості діагностики та лікування.

Отже, мета створити систему, здатну відповідати на відкриті запитання й при цьому демонструвати справжнє осмислення і розуміння цих запитань, залишається значною мірою нереалізованою.

Голосові веб-сервіси та помічники

▷Voice-based web services and assistants

Голосові сервіси, особливо на мобільних пристроях, є сферою застосування, яка швидко розвивається. Послуги варіюються від органайзерів (для списків покупок, розкладу зустрічей, нагадувань, списків контактів тощо), автомобільних інформаційно-розважальних систем (прокладання маршруту, визначення дорожніх умов, попередження про небезпеку, вибір iTunes, пошук найближчих ресторанів та інших закладів тощо) до можливості використання низки інших, витонченіших програм, як-от диктування електронної пошти, набір контактів, фінансові операції, бронювання та розміщення замовлень, доступ до Вікіпедії, послуги довідкової служби, консультації з питань здоров'я та відповіді на загальні питання. Деякі з цих послуг (наприклад, набір номера та вибір iTunes) потрапляють у категорію управління “вільні руки,” яке стає все більш важливим у транспорті (включаючи транспортні засоби без водія чи пілота), логістику (розгортання ресурсів) та виробництво. Основою для таких голосових сервісів виступає технологія чат-ботів та агентів підтримки діалогу.

Ключовим у цих сервісах є розпізнавання мовлення, точність та адаптивність якого поступово підвищуються. Найдешевші, вузькоспрямовані системи (наприклад, прості органайзери) мають жорстко визначені очікування щодо тих даних, що їх має ввести користувач, для розпізнавання, інтерпретації та реагування на них. Універсальніші системи, як-от автомобільні співрозмовники, які можуть опрацьовувати маршрути, музичні запити, пошук місць тощо, покладаються на більш просунуті можливості управління діалогом. Вони дозволяють перемикає теми та потенційно враховують стан уваги користувача (наприклад, відкладають відповідь на запитання водія, якщо йому потрібно повернути). Найбільший ажітаж на ринку технологій пов'язаний із передовими голосовими помічниками, зокрема з Siri для iPhone (за нею слідують Iris для Android, Evi для True Knowledge, Google Now та інші). Хоча попередні системи голосового управління та диктування, такі як Vlingo для Android, мали багато таких само функцій, Siri додає індивідуальність, покращує обробку діалогів та інтеграцію сервісів - користувачі відчувають, що вони взаємодіють із живим синтетичним персонажем, а не з програмою. Крім технології Nuance SR, Siri використовує складні технології, які до певної міри були запроваджені проектом Calo (Cognitive Assistant that Learns and Organizes), реалізованим SRI International та кількома університетами у 2003-2008 роках (Ambite et al., 2006). Ці технології базуються на методах розуміння природної мови, методах машинного навчання, методах умовиводу у невизначених умовах,

онтологіях, плануванні та делегуванні завдань іншим подібним програмам, як-от Wolfram|Alpha. Незважаючи на складність залучених технологій, здатність Siri до загального розуміння все ще дуже поверхова, що незабаром і виявляють користувачі. Анекдотичний приклад серйозного непорозуміння: у відповідь на запит користувача “Call me an ambulance” Siri відповіла “From now on I will call you ‘an ambulance.’” Утім, великий інтерес і попит у спільноті користувачів, викликаний цими ранніми, інтелектуальними (певною мірою) й досить універсальними помічниками, ймовірно, активізує та прискорить дослідження у напрямку все більш життєподібних віртуальних агентів, чия здатність до порозуміння із користувачем зростатиме (Schubert, 2020).



Завдання 21. Наведіть приклади сучасних користувацьких інтерфейсів природною мовою. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.³⁹

Узагальніть лінгвістичні аспекти створення і застосування користувацьких інтерфейсів природною мовою.

³⁹ Завдання середнього рівня складності.

Проектні роботи до Розділу 8



Проектна робота 25. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Пушкар, О.І., & Грабовський, Є.М. (2019). Методика розробки інтелектуального користувацького інтерфейсу навчальних видань в системі e-learning. *Business Perspectives. Development management*, 17(3), 23-34. [http://dx.doi.org/10.21511/dm.17\(3\).2019.03](http://dx.doi.org/10.21511/dm.17(3).2019.03)



Анотація. У статті наведено аналіз характерних особливостей і принципів побудови користувацького інтерфейсу та обґрунтування власного авторського методичного підходу до розробки інтелектуального користувацького інтерфейсу навчальних видань в системі e-learning. В роботі проаналізовано зміст поняття “користувацький інтерфейс”, розглянуті види інтерфейсів користувача. Для представників кожної репрезентативної системи в статті вказано перелік характерних способів сприйняття інформації з навколишнього середовища, які варто враховувати в процесі розробки користувацького інтерфейсу навчальних видань. Розглянуто основні особливості процедурно-орієнтованих та об’єктно-орієнтованих видів інтерфейсу користувача та наведено їх основні параметри. Впровадження інноваційного компонента в розроблювальний інтерфейс навчальних видань в системі e-learning в статті пропонується здійснювати шляхом додавання до інтерфейсу користувача властивості адаптивності. Визначено набір положень, які визначають концепцію інтерфейсу користувача, на основі чого сформовані принципи побудови адаптивних інтерфейсів. Запропоновано основні принципи візуального дизайну ментальних карт електронних навчальних видань для системи e-learning. Подано у вигляді візуальної

схеми основні можливості, що надаються ментальною картою. Розглянуто дидактичний інтерфейс розробленого електронного навчального видання та наведено схему взаємодії користувача з виданням. За допомогою методу фокус-групи здійснено оцінку якості користувацького інтерфейсу створеного електронного навчального видання.

Ключові слова: *електронні навчальні видання, користувацький інтерфейс, користувач, система e-learning, ментальна карта.*



Проектна робота 26. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Фуртат, Ю.О. (2013). Функціональні основи засобів адаптації користувацьких інтерфейсів до когнітивних особливостей користувачів в автоматизованих системах. *Математичне та комп'ютерне моделювання. Серія: Технічні науки*, 8, 120-131.



Анотація. Стаття присвячена задачі підвищення ефективності взаємодії користувачів складних автоматизованих систем з інформацією, що надходить від системи. Для вирішення задачі пропонується технологія управління користувацькими інтерфейсами, яка дозволяє проектувальникам та користувачам модифікувати інтерфейс системи в залежності від когнітивних особливостей взаємодії користувача з системою. В статті розглянуті принципи реалізації та запропонована функціональна модель системи настройки користувацьких інтерфейсів.

Ключові слова: автоматизована система, користувацький інтерфейс, форма представлення інформації, когнітивний стиль.



Проектна робота 27. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 8.
Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 9

Системи спільного вирішення проблем та інтелектуальні навчальні системи

▷ Collaborative problem solvers and intelligent tutors

Системи спільного вирішення проблем та інтелектуальні навчальні системи, ключову роль в яких відіграє діалог, можуть бути розглянуті під спільним заголовком, оскільки і ті й інші залежать від тих репрезентацій або моделей доменів, на які вони спираються у роботі, а також від психічного стану користувачів, які із ними взаємодіють.

Системи спільного вирішення проблем, на відміну від інтелектуальних навчальних систем, зазвичай мають справу з набагато менш передбачуваними ситуаціями, що їх надають для вирішення користувачі, й, відповідно, набагато більше уваги у таких системах приділяється гнучкій обробці діалогу. Наприклад, ситуація учасників спільної роботи з евакуації в надзвичайних ситуаціях, як це описується у роботі (Ferguson, & Allen, 1998, 2007), динамічно змінюється - на відміну від ситуації викладання фізики (наприклад, Jordan et al. 2006), де учня за допомогою авторських інструкцій можна провести через мережу цілей навчання й у відповідності до цих цілей розробити моделі діалогу, які класифікуватимуть вхідні дані від учня в кожній точці діалогу і генеруватимуть підготовлену відповідь, яка підходить для цього запиту.

Тому інтелектуальні навчальні системи мають більший комерційний успіх: такі системи були створені для багатьох областей та великої кількості потенційних клієнтів, починаючи від предметів шкільної програми та закінчуючи комп'ютерною грамотністю та програмуванням для початківців, якісною та кількісною фізикою, аналізом схем, експлуатацією машин, фізіологією серцево-судинної системи, боротьбою з пожежами на кораблях, навичками ведення переговорів та багатьма іншими царинами людської діяльності і знання (наприклад, див. Boyer et al., 2009). Серед найбільш успішних інтелектуальних навчальних систем - викладачі читання (наприклад, Cole et al., 2007), оскільки в цьому випадку матеріали, що їх потрібно представити учню, є відносно простими в розробці, а відповіді учня, особливо коли вони складаються в основному з читання представленого тексту вголос, відносно легко оцінити. Для більш амбітної мети - розвитку розуміння прочитаного - центральною проблемою залишається розробка діалогів таким

чином, аби зробити запит учня передбачуваним, а взаємодію - ефективною з освітньої точки зору (наприклад, Aist, & Mostow, 2009).

Деякі навчальні системи, особливо ті, що призначені для дітей, використовують анімованих персонажів, щоб посилити почуття залучення учня. Такі вдосконалення фактично необхідні для систем, призначених для учнів з такими порушеннями, як глухота (коли рухи рота та язика віртуального агента, за якими спостерігає учень, можуть допомогти з артикуляцією), аутизм або афазія (Massaro et al., 2012; Cole et al., 2007). Крім того, якщо викладання спрямоване на навчання міжособистісним навичкам, створення життєподібних персонажів (віртуальних людей) стає невід'ємною частиною розробки системи (наприклад, Campbell et al., 2011).

Моделювання душевного стану користувача в навчальних системах є насамперед питанням визначення того, які з цільових понять та навичок вже засвоєні або ще не засвоєні користувачем, а також діагностики непорозуміння, яке могло мати місце й задля виявлення якого залучається протокол поточної користувацької сесії. Деякі останні експериментальні системи можуть також адаптувати свої стратегії до настрою користувача, як-от до розчарування чи нудьги, про що може свідчити поведінка, тон голосу і навіть вираз обличчя або жести, проаналізовані за допомогою комп'ютерного зору. Інші прототипні системи також прагнуть до об'ємнішого моделювання психіки користувача, яке базуватиметься, зокрема, на покращеному розумінні його мови (наприклад, Callaway et al., 2007).

У діалогових системах, орієнтованих на вирішення конкретних завдань, моделювання діалогу є набагато складнішим, оскільки від таких систем очікується не лише внесок у вирішення поставленого завдання, а й розуміння висловлювань, переконань та намірів користувача, а також уміння тримати себе у людиноподібному діалозі зі змішаними ініціативами. Для цього необхідні моделі відповідних доменів, розвинені методи спільного планування, гнучке управління діалогом, яке моделюватиме раціональну комунікативну взаємодію, й повне розуміння мови (особливо розпізнавання намірів) користувача. Прототипи подібних систем мають дуже великі практичні перспективи і вже були успішно створені для таких областей, як планування маршрутів й авіаперельотів, супровід водіїв та пішоходів, керування та робота зовнішніх пристроїв, евакуація у надзвичайних ситуаціях та консультування щодо прийому ліків (наприклад, Bühler, & Minker, 2011). Утім, системи, здатні вирішувати складні завдання, особливо такі, що потребують широких знань про людське мислення та поведінку, сьогодні є недосяжними (Schubert, 2020).



Завдання 22. Наведіть приклади сучасних систем спільного вирішення проблем. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.⁴⁰

Узагальніть лінгвістичні аспекти створення і використання систем спільного вирішення проблем.



Завдання 23. Наведіть приклади сучасних інтелектуальних навчальних систем. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи роботи цих систем, а також підходи і методи, які використовуються в цих системах.⁴¹

Узагальніть лінгвістичні аспекти створення і використання інтелектуальних навчальних систем.

⁴⁰ Завдання середнього рівня складності.

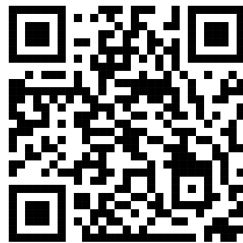
⁴¹ Завдання середнього рівня складності.

Проектні роботи до Розділу 9



Проектна робота 28. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Сорока, М.Ю., Сало Н.А., & Матющенко О.Г. (2020). Інтелектуальна навчальна система підготовки диспетчерів управління повітряним рухом. *Системи обробки інформації*, 2(161), 29-36. <https://doi.org/10.30748/soi.2020.161.04>



***Анотація.** У статті розглянуті основні методи організації інтелектуального навчального середовища підготовки диспетчерів управління повітряним рухом. Обґрунтована необхідність створення інтелектуальної навчальної системи в адаптивних тренажерах диспетчерів управління повітряним рухом. Інтелектуальні навчальні системи повинні базуватись на основі синтезу імітаційно-моделюючих комплексів у вигляді розподілених систем обробки даних для імітації поведінки середовища навчання. В роботі сформувані вимоги, що висуваються до побудови мультиагентного середовища інтелектуальної навчальної системи підготовки диспетчерів управління повітряним рухом. З метою забезпечення моделювання інтелектуального поведіння об'єктів, що входять у віртуальне середовище навчання, запропоновано створення інтелектуальних об'єктів, як елементів мультиагентної системи з використанням методів планування дій. Запропоновано підхід удосконалення та розширення функціональних можливостей системи підготовки диспетчерів управління повітряним рухом. Запропонована архітектура інтелектуального агента навчальної системи підготовки диспетчерів управління повітряним рухом створена на базі елементів InterRRa архітектури, що забезпечує взаємодію агента з зовнішнім середовищем та іншими агентами через модель фізичного представлення*

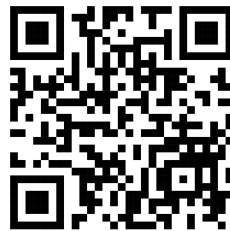
об'єкта. Наведена математична модель інтелектуального агента в якій враховано можливість здійснення впливу на зовнішнє середовище. Розроблена модель мультиагентного середовища інтелектуальної навчальної системи яка забезпечує ідентифікацію ситуації в підсистемі підготовки і прийняття рішень, що виконує передачу управління на відповідний рівень ієрархії системи поводження інтелектуального агента. Особливістю розробленої структури мультиагентного середовища інтелектуальної навчальної системи є використання моделі поведінки інтелектуальних агентів, що забезпечують змінну поведінку і можливість рішення задач підготовки і прийняття рішень своїх подальших дій за допомогою різних моделей поведінки.

Ключові слова: управління повітряним рухом, інтелектуальна система, розподілена система, мультиагентна система, штучний інтелект, модель навчання.



Проектна робота 29. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Сорока, М.Ю. (2020). Метод адаптації поведінки агентів в інтелектуальній навчальній системі підготовки диспетчерів управління повітряним рухом. *Системи управління, навігації та зв'язку. Збірник наукових праць*, 2(60), 17-20. doi:<https://doi.org/10.26906/SUNZ.2020.2.017>



Анотація. В рамках агентно-орієнтованого підходу з метою підвищення варіативності середовища підготовки диспетчерів управління повітряним рухом у статті удосконалено метод адаптації поведінки агентів навчального середовища інтелектуальної навчальної системи, що ґрунтується на знаннях про результати взаємодії агентів. На етапі внесення змін до параметрів підсистеми планування при настоюванні систем моделювання поведінки агентів запропоновано використовувати механізм самонастроювання. Для управління процесом самонастроювання поведінки інтелектуальних агентів в статті розроблено інструментарій адаптації системи планування поведінки агента. Застосування моделей поведінки агентів із методами планування дозволить збільшити показник успішності реалізації очікуваної поведінки агентів при побудові інтелектуальних навчальних систем. Запропонований у статті підхід адаптації поведінки агентів навчального середовища інтелектуальної навчальної системи дозволить підвищити його варіативність. Як наслідок, така інтелектуальна навчальна система дозволить вивести на якісно новий рівень підготовку диспетчерів управління повітряним.

***Ключові слова:** агент, адаптація, диспетчер управління повітряним рухом, середовище інтелектуальної навчальної системи, мультиагентні системи, планування, самонастроювання, теорія нечітких множин, ефективність навчання.*



Проектна робота 30. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 9.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 10

Наділені мовою роботи

▷Language-enabled robots

Сучасні роботи все частіше починають оснащуватися веб-сервісами, здатністю відповідати на питання, технологіями чат-ботів, функціями інтелектуальних навчальних систем тощо, залучаючи у такий спосіб у розробку усі наявні на сьогодні здобутки комп'ютерної науки і, зокрема, прикладної комп'ютерної лінгвістики.

Перенесення таких технологій на роботів відбувається повільно, в основному через дуже складні завдання, пов'язані з оснащенням робота апаратним та програмним забезпеченням, необхідним для базового візуального сприйняття, розпізнавання мови, дослідницької та цільової навігації (для мобільних роботів) та маніпулювання об'єктами. Утім, жвавий суспільний інтерес до розумних роботів та їхній потужний економічний потенціал (допомога по дому, догляд за літніми людьми, медицина, освіта, розваги, сільське господарство, промисловість, пошук та порятунок, військові місії, освоєння космосу тощо), безсумнівно, продовжують стимулювати розробників у їхньому прагненні підвищити рівень інтелекту та мовної компетенції роботів.

Основні труднощі діалогу людини та робота розглянуто у роботі (Scheutz et al., 2011), де деякі з прикладів діалогу про коробки і блоки, якими маніпулює робот, нагадують SHRDLU Террі Вінограда, але при цьому демонструють на той момент безпрецедентні виклики, пов'язані з реальною взаємодією суб'єктів із об'єктами у фізичному світі, як-от зміна оточення при русі робота, помилки у розпізнаванні мовлення, неоднозначні та складні багатослівні речення, тлумачення яких, наприклад, залежить від перспективи (“Is the red box to the left of the blue box?”) та має враховувати дейксис (“Go down there”). Усе це, крім того, має бути узгоджене з фізичними діями робота, планованими так, аби він виконав інструкції природною мовою у той спосіб, як він їх розуміє. Хоча здатність сучасних роботів долати ці труднощі дає, певною мірою, надію, багато проблем залишаються відкритими, як-от розпізнавання мовлення в присутності шуму, адекватний синтаксичний і семантичний аналіз в обробці діалогів, адаптація робота до змін у середовищі і нових завдань, моделювання у “мисленні” робота його співрозмовників та інших людей, які знаходяться у середовищі, а також формалізація фонових знань із здатністю використовувати ці знання для умовиводів та планування (на рівні як загальної теми, так і конкретного діалогу із цієї теми).

Діалоги роботів, що є орієнтованими на виконання завдань, пов'язані з усіма цими проблемами. Утім, слід зазначити, що деякі потенційно корисні для людей взаємодії з “розмовляючими” роботами не вимагають від останніх особливих мовних компетенцій. Наприклад, робот Rubi, описаний в роботі (Movellan et al., 2009), показував малюкам предмети на своїй “скрині” із екраном, просячи їх доторкнутися і назвати предмети. В результаті, попри простоту взаємодії, малюки краще засвоювали слова. Іншим прикладом дуже успішного розмовляючого робота без реальних лінгвістичних навичок є “музейний гід” Rhino (Burgard et al., 1999). На відміну від Rubi, він міг орієнтуватися серед людей, що рухаються спонтанно, і підтримував цікавість аудиторії своїми заздалегідь записаними повідомленнями природною мовою та відображенням своїх поточних цілей на екрані. Аналогічно, численні людиноподібні роботи (наприклад, Asimo компанії Honda), що розроблялися в минулому і розробляються у цей час по всьому світу, все ще мало розуміють мову і покладаються в основному на скрипти. Безсумнівно, їхня корисність і привабливість зростатимуть завдяки усім технологіям, згаданим вище: тим, що використовуються у створенні ігор, агентів підтримки діалогу, голосових асистентів, інтелектуальних навчальних систем тощо. Ці технології, у свою чергу, сприятимуть моделюванню у штучних системах витонченіших аспектів людиноподібного сприйняття, руху, маніпулювання об'єктами та осмисленого діалогу (Schubert, 2020).



Завдання 24. Наведіть приклади сучасних роботів, наділених мовою. Дослідіть на основі авторитетних джерел та опишіть (своїми словами!) принципи функціонування цих роботів, а також підходи і методи, які покладені в основу створення цих роботів.⁴²

Узагальніть лінгвістичні аспекти створення і використання наділених мовою роботів.

⁴² Завдання середнього рівня складності.

Проектні роботи до Розділу 10



Проектна робота 31. [Усі студенти групи є учасниками цієї проектної роботи]

Оберіть одну із запропонованих тем так, аби між усіма студентами групи та обраними темами встановилося взаємно однозначне відношення. На основі авторитетних джерел, дослідіть обрану Вами тему; дайте темі ємну і привабливу назву. Створіть постерну доповідь для обраної Вами теми. Підготуйтеся до її обговорення; будьте готові висловити і обґрунтувати власну думку.⁴³

1. Витоки ідеї створення роботів. Ранні роботи
2. Походження слова “робот.” Еволюція роботів. Сучасні роботи
3. Механічні роботи vs. штучні агенти-програми⁴⁴
4. Види механічних роботів, підходи до їхнього створення
5. Роботи, які розмовляють: підходи, прототипи, відомі приклади
6. Сучасні сфери використання роботів
7. Відомі роботи й провідні компанії-розробники роботів
8. Роботи і робототехніка в Україні: стан, здобутки і перспективи
9. Роботи і суспільство людей: автономія роботів
10. Роботи і суспільство людей: моральні роботи
11. Роботи і суспільство людей: етичні виклики і дилеми
12. Роботи-андроїди. Антропоморфізм роботів: виміри, підходи й оцінки
13. Роботи-“митці”: література, музика і живопис, створені роботами

⁴³ Семінар(и) намічено як постер-конференцію, де доповіді, що із ними попередньо мають ознайомитися усі учасники, обговорюватимуться у форматі круглого столу.

⁴⁴ англ. *Mechanical robots vs. Software agents*.

14. Література і мистецтво про роботів: спектр і еволюція окреслених проблем
15. Майбутнє роботів і робототехніки: можливі нові функції і прототипи
16. *Запропонуйте власну - доладну і варту - тему*



Проектна робота 32. [Усі студенти групи є учасниками цієї проектної роботи]

Оберіть одну із запропонованих тем так, аби між усіма студентами групи та обраними темами встановилося взаємно однозначне відношення. На основі авторитетних джерел, дослідіть обрану Вами тему; дайте темі ємну і привабливу назву. Створіть постерну доповідь для обраної Вами теми. Підготуйтеся до її обговорення; будьте готові висловити і обґрунтувати власну думку.⁴⁵

1. Історія штучного інтелекту
2. Підходи до визначення штучного інтелекту
3. Загальний і спеціальний штучний інтелект
4. Природний і штучний емоційний інтелект
5. Підходи до створення систем штучного інтелекту
6. Лінгвістичне забезпечення в системах штучного інтелекту
7. Штучний інтелект в Україні: наукові й інжинірингові розробки, провідні установи, компанії та стартапи
8. Прикладні аспекти штучного інтелекту: інтелектуальні гаджети
9. Прикладні аспекти штучного інтелекту: віртуальна реальність
10. Прикладні аспекти штучного інтелекту: Інтернет-пошук
11. Прикладні аспекти штучного інтелекту: розпізнавання образів
12. Прикладні аспекти штучного інтелекту: автономні автомобілі
13. Прикладні аспекти штучного інтелекту: комп'ютерні ігри

⁴⁵ Семінар(и) намічено як постер-конференцію, де доповіді, що із ними попередньо мають ознайомитися усі учасники, обговорюватимуться у форматі круглого столу.

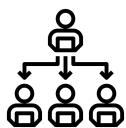
14. Прикладні аспекти штучного інтелекту: відкриті математичні задачі
15. Прикладні аспекти штучного інтелекту: моделювання природної мови
16. Штучний інтелект: історії комерційного успіху
17. Штучний інтелект: філософські передумови, аспекти і виклики
18. Етичний штучний інтелект і моральні роботи
19. Сильний і слабкий штучний інтелект
20. Сингулярність
21. Штучний інтелект: можливості та ризики для *Homo sapiens*
22. Штучний інтелект і технологічне безробіття
23. Суперкомп'ютер Watson від IBM
24. Штучний інтелект AlphaGo від Google DeepMind
25. Штучний інтелект Libratus aka Pluribus від Facebook AI Research
26. Майбутнє штучного інтелекту
27. *Запропонуйте власну - доладну і варту - тему*



Проектна робота 33. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 10.

Будьте винахідливі й розвивайте власне почуття напрямку.

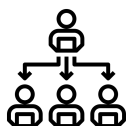
ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ



Завдання 1. Прочитайте статтю “Машинний переклад” в *Українській Вікіпедії*.

На основі авторитетних джерел, оцініть якість і повноту статті. Виявіть недоліки і прогалини у статті та додайте релевантну інформацію із посиланням на використувані Вами джерела.⁴⁶

Створіть власну статтю “*Машинний переклад*,” орієнтовану на Українську Вікіпедію⁴⁷ та академічну доброчесність. Представте статтю на семінарі; по підсумках обговорення, оновіть в Українській Вікіпедії сторінку “Машинний переклад.”⁴⁸



Завдання 2. Онлайн-платформа *Grammarly*: дослідіть витoki, профіль компанії, напрямки діяльності, продукти і сервіси тощо. Будьте винахідливі у своєму дослідженні й здійсніть ґрунтовний, дійсно інтелектуальний веб-пошук, включивши у нього афілійовані ресурси, нещодавні і майбутні події, соціальні мережі (зокрема, Фейсбук) й останні активності компанії у них, контакти, особисті сайти фахівців тощо.

Особливу увагу зверніть на *можливості професійного розвитку* (зокрема, ті, що Grammarly традиційно пропонує *українським студентам* - наприклад, воркшопи і літні школи), кар'єри, вимоги до кандидатів, історії

⁴⁶ Завдання цього розділу намічені як такі, що можуть бути виконані також у форматі проєктної роботи, або бути заплановані для написання підсумкових есе, курсових проєктів (англ. *a term paper*) тощо.

⁴⁷ Студентам пропонується ознайомитися із поняттям і принципами відкритої науки (англ. *Open Science*) і, зокрема, подивитися на Українську Вікіпедію як на просвітницьку громадську ініціативу поширення наукового знання в Інтернет-добу, із розумінням того, що це знання має бути якісним.

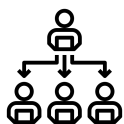
У нашому курсі, студентів запрошено до створення та редагування Вікіпедія-статей на відповідні теми: ця практика, окрім того, що надасть безпрецедентний досвід і дозволить студентам відкрито опублікувати результати своєї роботи, також тренуватиме широкий спектр цифрових і м'яких навичок студентів й сприятиме покращенню якості статей у Вікіпедії. Закликаємо студентів опікуватися своїми статтями у Вікіпедії і після завершення курсу й долучатися до спілкування за інтересами, яке відіграє значну роль як у науковій роботі, так і в професійному житті.

Ми вдячні доценту Тімо Торстену Шмідту, *who is an active Wikipedian*[®], за натхнення залучити роботу студентів із Вікіпедією у викладання.

⁴⁸ Таке саме завдання може стосуватися й інших прикладних аспектів комп'ютерної лінгвістики, обговорюваних в навчально-методичному посібнику: студентам може бути запропоновано оновити або, за необхідності, створити відповідні сторінки в Українській Вікіпедії.

успіху, особливості корпоративної культури, сучасний стан і тенденції розвитку ринку праці тощо. Чи відгукуються вони Вам особисто? Чи відгукнулися б Ви їм? Чи можуть слугувати певним дороговказом (англ. *a road map*)?

Структуруйте результати своєї розвідки і представте їх на семінарі. Проведіть для колег розгорнуту онлайн-екскурсію.



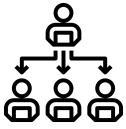
Завдання 3. Дослідіть призначення і функціонал ресурсів, об'єднаних у ***“The Stanford Natural Language Processing Group.”***



Також, дослідіть афілійовані до ресурсу групи, нещодавні і майбутні події, соціальні мережі й останні активності у них, контакти, особисті сайти й персональні сторінки учасників тощо - будьте винахідливі у своєму дослідженні й здійсніть ґрунтовний, дійсно інтелектуальний веб-пошук.

Особливу увагу зверніть на ***кар'єрні можливості*** (зокрема, ті, що пропонуються (міжнародним) студентам), вимоги до кандидатів, особливості корпоративної культури, сучасний стан і тенденції розвитку ринку праці тощо. Чи відгукуються вони Вам особисто? Чи відгукнулися б Ви їм? Чи можуть слугувати певним дороговказом - і з чого цю дорогу розпочати?

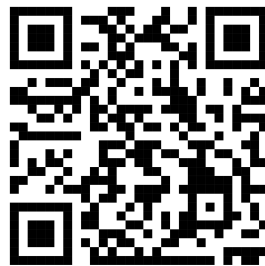
Структуруйте результати своєї розвідки і представте їх на семінарі. Проведіть для колег розгорнуту онлайн-екскурсію.



Завдання 4. На основі авторитетних джерел, дослідіть тему **Мови програмування**. Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними: мова програмування - визначення; класифікація мов програмування; синтаксис і семантика мов програмування; діалекти мов програмування; мова програмування vs. комп'ютерна мова; особливості застосування; затребувані по галузях й поширені у світі мови програмування.

Поміркуйте над співвідношенням мов штучного інтелекту й штучних мов.

Дослідіть призначення і функціонал ресурсів, об'єднаних у **“Github.”**



Дослідіть призначення і функціонал ресурсів, об'єднаних у **“Ontohub.”**



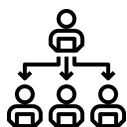
Також, дослідіть афілійовані до ресурсів продукти, послуги, проекти і групи, нещодавні і майбутні події, соціальні мережі й останні активності у них, контакти, особисті сайти й персональні сторінки учасників тощо - будьте винахідливі у своєму дослідженні й здійсніть ґрунтовний, дійсно інтелектуальний веб-пошук.

Особливу увагу зверніть на **кар'єрні можливості** (зокрема, ті, що пропонуються (міжнародним) студентам), вимоги до кандидатів,

особливості корпоративної культури, сучасний стан і тенденції розвитку ринку праці тощо. Чи відгукуються вони Вам особисто? Чи відгукнулися б Ви їм? Чи можуть слугувати певним дороговказом?

Знайдіть досяжні Вам пропозиції **вивчення мов програмування**: в українському та міжнародному освітньому (онлайн-)просторі. Зважте на ці пропозиції, порівняйте їх. Чи відгукуються вони Вам?

Структуруйте результати своєї розвідки і представте їх на семінарі. Проведіть для колег розгорнуту онлайн-екскурсію.



Завдання 5. Дослідіть ресурси *IBM Training* і *Microsoft Learn*.

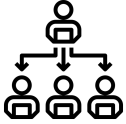


Юні колеги, слова тут зайві. “Як нерозумно випрошувати те, чого можеш сам досягти!” (Г. Сковорода).

Із точки зору користувачів, розгляньте можливості і цифрові рішення, що їх об'єднано у *Microsoft | Osvima*. Як запропоновані інструменти можуть бути застосовані Вами протягом навчання в Університеті?



Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 6. Особливу увагу у посібнику приділено *машинному перекладу*. Запрошуємо Вас ширше - з лінгвофілософської точки зору - подивитися на *переклад*, що є Вашою (майбутньою) професією.

Дослідіть на основі авторитетних джерел сутність наступних понять: метамова, мова смислу, інтерлінгва; природна мова, штучна мова; мова програмування. Обміркуйте (взаємо)зв'язок цих понять, якщо такий є.

На Вашу думку, чи є метамова як універсальний формалізований код семантичного й синтаксичного знакового опису природних мов тією самою метомовою, про яку пише О.О. Кришталь (2020)? Яке тлумачення метамови Вам ближче? Чому?

Чи є інтерлінгва, що використовується у трансферних системах машинного перекладу, метомовою? Чи, може, вона є мовою смислу? Чому?

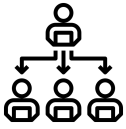
Програмісти кажуть, що, всупереч поширеній думці, програмування - це переклад, а не математика. Яку роль у діалозі “людина-машина(-людина)” відіграють мови програмування?

Зважте на перспективи *перекладу в IT-сфері*. Наш курс може стати в неабиякій пригоді.

Дослідіть досяжний Вам сегмент сучасного українського та міжнародного ринків праці для перекладачів в IT, знайдіть і зважте на можливості такого працевлаштування. Чи відгукуються вони Вам? Які вимоги висуваються до кандидатів? Що треба зробити, аби цим вимогам відповідати?

Протягом вивчення дисциплін перекладознавчого циклу в Університеті, не проминіть можливість розширити свій набір дієвих інструментів перекладача.

Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 7. На основі авторитетних джерел, дослідіть тему **Взаємодія людини і комп'ютера**.⁴⁹ Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними: взаємодія людини і комп'ютера як інтерфейс; принципи і методології комп'ютерного дизайну; соціальні обчислення й інженерія⁵⁰; емоції і комп'ютер; нейрокомп'ютерні інтерфейси⁵¹; напрямки досліджень і прикладні аспекти.

Значне місце у Human-computer interaction, а особливо у Cognitive human-computer interaction, посідає моделювання окремих аспектів взаємозв'язку природної мови і мислення людини. На Вашу думку, чому? Чи відгукується Вам перспектива медіювати взаємодію людини і комп'ютера?

На **(нейро)лінгвістичні аспекти** цієї взаємодії Вас можуть найсучасніше скерувати:

Moens, J. (September 8, 2022). An AI can decode speech from brain activity with surprising accuracy. *ScienceNews. Neuroscience*.



Oxley, T. (June 2, 2022). A brain implant that turns your thoughts into text. *TED2022. TED: Ideas Worth Spreading*.

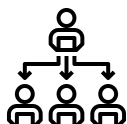
⁴⁹ англ. *Human-Computer Interaction*.

⁵⁰ англ. *Social Computing*.

⁵¹ англ. *Brain-Computer Interfaces*.



Структуруйте результати своєї розвідки і представте їх на семінарі.



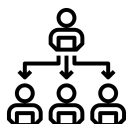
Завдання 8. На основі авторитетних джерел, дослідіть тему *Інформація. Теорія інформації*. Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними: інформація: визначення і властивості, види і функції; теорія інформації⁵²: визначення, витоки, ключові поняття, завдання, міждисциплінарні зв'язки.

Чим (інтелектуальний) інформаційний пошук відрізняється від пошуку інформації? Що таке інформаційний шум?

Осібнo, дайте визначення інформатиці⁵³, квантовій інформатиці⁵⁴ й нейроінформатиці⁵⁵ через зв'язок цих дисциплін із завданням (квантового) обчислення⁵⁶.

Схарактеризуйте відношення між інформацією і **знанням**. Чому говорять, що “знання - сила,” а “той, хто володіє інформацією, володіє світом”? Як співвідносяться знання і **вміння**?

Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 9. На основі авторитетних джерел, дослідіть тему *Експертні системи*. Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними:

⁵² англ. *Information Theory*.

⁵³ англ. *Information Science*.

⁵⁴ англ. *Quantum Information Science*.

⁵⁵ англ. *Neuroinformatics*.

⁵⁶ англ. *(Quantum) Computing*.

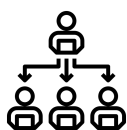
експертна система - визначення; витоки й еволюція експертних систем; сучасні підходи до створення; архітектури програмного забезпечення в експертних системах; застосування і функції експертних систем; сильні й слабкі сторони експертних систем.

Дослідіть і узагальніть *лінгвістичні аспекти* створення і використання експертних систем.⁵⁷ Який аспект людської мовної і мовленнєвої діяльності Ви б вважали за потрібне, й запропонували б, залучити до створення таких систем? Як саме Ваші фахові компетентності можуть бути залучені у таке створення? Які жорсткі та м'які, а також цифрові, навички Вам знадобляться?

На Вашу думку, наскільки ефективними є тести на коефіцієнт інтелекту?⁵⁸ Дослідіть історію, проблематику і підходи до їхнього створення. Чи співвідносяться тести зовнішнього незалежного оцінювання (ЗНО) із тестами на коефіцієнт інтелекту? Проаналізуйте свій досвід складання ЗНО.

Експерт - хто це? Якими якостями має володіти людина, щоб бути визнаною експертом? Поміркуйте над критеріями такого визнання. Людина-експерт vs. експертна система: чи зможуть експертні системи свого часу повністю замінити людей-експертів? Чому?

Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 10. Тема *Природна мова: у межах і поза межами розуміння*: подивіться і законспекуйте лекцію Н. Чомскі “Language and the Limits of Understanding” від 30 вересня 2021 року, присвячену Міжнародному дню перекладу.

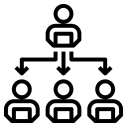
⁵⁷ Див. також (Біскуб, 2015).

⁵⁸ англ. *IQ tests*.



Особливу увагу зверніть на фрагмент лекції з 24 по 30 хвилини. Схарактеризуйте креативний, або генеративний, аспект використання мови, у його витлумаченні Н. Чомскі. Як відбувається генерування думок? Як одна думка обирається із багатьох у процесі мовлення? Чим генерування думки відрізняється від продукування мовлення? Що саме знаходиться поза межами людського розуміння?

Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 11. Тема *Мозок і свідомість: українська нейронаука у медіа-контексті*: на основі відкритих джерел, дослідіть наукову (і творчу) діяльність академіка О.О. Кришталя, прочитайте і подивіться кілька визначних інтерв'ю Олега Олександровича українським медіа-виданням.

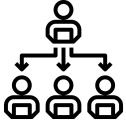
Що Ви дізналися про свій мозок? Про свої мислення і свідомість? Про мову? Хто такий Гомункулус, і якою є мова думки на відміну від мови слів?

Як влаштований і як працює мозок людини? Якою є його робота? Як людині дбайливо й ефективно використовувати свій мозок? Що шкодить мозкові?

Як роздуми, читання і письмо *стимулюють мозок*, сприяючи появі і розвиткові нових нейронів? Чому “багатство людини має вимірюватися не грошима, а кількістю нейронних зв'язків” (О.О. Кришталь)?

Поясніть, будь ласка, як термін “штучна нейронна мережа” виконує свої функції наукової метафори?

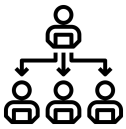
Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 12. Знайдіть і поспілкуйтеся (наприклад, у соціальних і професійних мережах, на відповідних форумах тощо) із **випускниками** Вашої спеціальності (як варіант, суміжних або цікавих Вам особисто спеціальностей) про те, як вони влаштувалися і як зрештою влаштувалися у професії.

Узвіть їх поділитися досвідом працевлаштування, професійного (само)розвитку, побудови кар'єри тощо. Як цьому сприяли здобуті в Університеті знання, вміння і навички? Якою була роль власного досвіду, завзяття, мотивації і відповідальних рішень? *Додайте цікаві Вам питання.*

Структуруйте результати своєї розвідки і представте їх на семінарі.



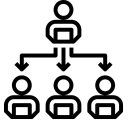
Завдання 13. Чергування на семінарах (англ. *Seminar duty*): протокольний супровід семінарів (англ. *minutes*), рекап (англ. *a recap*) на початку і підведення підсумків (англ. *a take-home message*) у кінці кожного семінару, формулювання питань, організаційна комунікація з учасниками семінару.

Можливість додати тему для обговорення, питання, напрямок, ідею, релевантну літературу й електронні ресурси, які б Ви рекомендували колегам, тощо.

Чергування розраховане як таке, що триває протягом усього семестру вивчення курсу й сприяє розвитку Ваших лідерських якостей і навичок управління проєктами.⁵⁹

Наприкінці семестру, звітуйте про свій досвід.

⁵⁹ Зацікавлених студентів закликаємо ознайомитися на дозвіллі із методологіями управління проєктами: Agile, Kanban і Scrum (зверніться також до описів вакансії (англ. *a job description*) Scrum-master на українському та міжнародному ринках праці - якими є вимоги?); а також, ознайомитися із (цифровими) інструментами управління часом і управління знанням, використовуваними для командної роботи у бізнес-середовищі, наприклад, Trello, Jira, Slack, Sharepoint, Notion тощо.



Завдання 14. Шановні студенти, це завдання відкрите для Ваших ідей відповідно до тематики курсу.

Будьте винахідливі й розвивайте власне почуття напрямку.

ПІСЛЯМОВА

Цей посібник є спробою схопити академічний дух тих найкращих практик вищої освіти, що ми їх засвоїли протягом свого навчання, стажувань і викладання у можних університетах України, Німеччини і Польщі та використовуємо, зокрема, у викладанні курсу комп'ютерної лінгвістики, що його покликаний супроводжувати цей посібник.

Курс базується на нашому досвіді вивчення компютерної лінгвістики в Інституті когнітивної науки Оснабрюцького університету і надихнутий перебуванням саме у цьому академічному середовищі. Шанобливо вдячні Вчителю і нашим викладачам!

До наших студентів звертаємося із тим, що розуміємо і поділяємо Вашу охоту до якостей європейської освіти і прагнемо Вам усіляко у набутті цих якостей сприяти. Через це прагнення, дозволимо собі наголосити на тій активній, відповідальній ролі, яку студенти європейських університетів радо беруть на себе у навчанні й яку, на нашу думку, українські студенти поки що не зовсім налаштовані помічати: оспівуючи академічні свободи, варто було б оспівати - чи не у першу чергу - й академічні доброчесності. Однією із таких чеснот вважаємо відповідальність, що її доросла людина несе за своє навчання, і мотивацію до такого навчання: навіть *-ся* у слові “вчитися” означає, що вчать себе. Закликаємо студентів на дозвіллі поцікавитися етимологією слова “студент” і звернутися до витоків й перекладів “*Gaudeamus igitur*” - потверджуємо, юні колеги, ту неабияку мудрість, що там міститься.

Курс комп'ютерної лінгвістики запрошує Вас ширше подивитися на обрану спеціальність, на її сутність і можливості застосування здобутих в Університеті знань і вмінь, й зважити на перспективи самореалізації, які ця спеціальність Вам надає - у вимірі фахових і загальних компетентностей. Вибудова цих компетентностей - у фундаментальній університетській освіті - передує вбудовуванню свого професійного талану у життя, у той реальний контекст, що оточуватиме Вас на момент працевлаштування й у подальшій трудовій діяльності. Протягом курсу, закликатимемо Вас усвідомлювати особливості цього контексту: дивитися на себе як на кандидата очима HR, складати CV, набувати досвід із work.ua і dou.ua, заглядати на michaelpage.com, проходити job interviews, шукати professional networking у LinkedIn, долучатися до освітніх і профорієнтаційних подій, навідуватися до coursera.org і udemy.com, визначатися з пріоритетами розвитку себе, пам'ятати “отже, веселімося” і зрештою впевнитися, що на момент випуску з Університету Ви встигли

створити для себе достатньо можливостей набуття практичного досвіду за спеціальністю, аби не лише все знати, але й дещо вміти...

“З усіх утрат втрата часу найтяжча” (Г. Сковорода). Юні колеги, бажаємо Вам пожити час навчання із щонайбільшою користю!

О.В. Ваховська

ЛІТЕРАТУРА

ОСНОВНА

- Балог, В. (2005). Сучасний стан української комп'ютерної лінгвістики. *Лексикографічний бюлетень*, 11, 28-35.
- Біскуб, І. (2015). Мова експертних систем: штучна чи природна? *Науковий вісник Східноєвропейського національного університету імені Лесі Українки. Серія : Філологічні науки. Мовознавство*, 4, 171-177.
- Бурдаєв, В.П. (2020). Використання чат-бота @es_economy_karkas_bot для онлайн консультації з експертною системою. *Системи обробки інформації*, 1(160), 100-106. <https://doi.org/10.30748/soi.2020.160.13>
- Вавіленкова, А. (2017). Структура системи порівняльного аналізу електронних текстових документів за змістом. *Вісник Національного університету "Львівська політехніка." Комп'ютерні науки та інформаційні технології*, 864(1), 277-285.
- Гирич, О.В. (2020). Основні проблеми систем обробки природної мови. *Філологічні студії. Збірник наукових праць*, 14, 41-45.
- Завадська, В. (2013). Коли "вікно" не є вікном, або ще раз про сучасну українську ІТ-термінологію. *Українське мовознавство*, 43, 20-27.
- Карпіловська, Є.А. (2006). *Вступ до прикладної лінгвістики: Комп'ютерна лінгвістика: Підручник*. Донецьк: ТОВ «Юго-Восток, Лтд.»
- Карпіловська, Є.А. (2014). Комп'ютерна лінгвістика. *Енциклопедія Сучасної України : енциклопедія [електронна версія] ; ред.: І.М. Дзюба, А.І. Жуковський, М.Г. Железняк та ін.; НАН України, НТШ. Київ : Інститут енциклопедичних досліджень НАН України. Т. 14. URL: <https://esu.com.ua/article-4396> (дата перегляду: 28.11.2022)*
- Коломієць, В.О. (2022). *Короткий англо-український словник з комп'ютерної лінгвістики*. Київ: Видавничий центр КНЛУ.
- Комарницька, О. (2018). Методи автоматизованого семантичного аналізу природномовної інформації. *Філологічний дискурс*, 7. 92-100.
- Кришталь, О.О. (2020). *До співу пташок: приватна подорож до себе*. Київ: Паливода А.В.
- Кришталь, О.О. (2022). *Гомункулус*. (Готується до публікації українською мовою).
- Лавренчук, С.В., & Товстенюк, Б.С. (2020). Інформаційний чат-бот для сервісу обміну повідомленнями Telegram в навчальній сфері. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*, 41, 180-185. <https://doi.org/10.36910/6775-2524-0560-2020-41-28>

- Нога, Р., & Шаховська, Н. (2011). Аналітичний огляд методів та засобів опрацювання текстової інформації. *Вісник Національного університету "Львівська політехніка."* *Інформаційні системи та мережі*, 715(2), 323-332.
- Погорілий, С.Д., & Крамов, А.А. (2018). Автоматизована екстракція структурованої інформації з множини веб-сторінок. *Проблеми програмування. Науковий журнал*, 2(3), 149-158. <http://dx.doi.org/10.15407/pp2018.02.149>
- Погорілий, С.Д., & Крамов, А.А. (2018). Метод розрахунку когерентності українського тексту. *Реєстрація, зберігання і обробка даних. Інформаційно-аналітичні системи обробки даних*, 20(4), 64-75. <http://dx.doi.org/10.35681/1560-9189.2018.20.4.178945>
- Пушкар, О.І., & Грабовський, Є.М. (2019). Методика розробки інтелектуального користувачького інтерфейсу навчальних видань в системі e-learning. *Business Perspectives. Development management*, 17(3), 23-34. [http://dx.doi.org/10.21511/dm.17\(3\).2019.03](http://dx.doi.org/10.21511/dm.17(3).2019.03)
- Романюк, А., Романишин, М. (2013). Тональний словник української мови на основі сентимент-анотованого корпусу. *Українське мовознавство*, 43, 63-75.
- Сорока, М.Ю. (2020). Метод адаптації поведінки агентів в інтелектуальній навчальній системі підготовки диспетчерів управління повітряним рухом. *Системи управління, навігації та зв'язку. Збірник наукових праць*, 2(60), 17-20. doi:<https://doi.org/10.26906/SUNZ.2020.2.017>
- Сорока, М.Ю., Сало Н.А., & Матющенко О.Г. (2020). Інтелектуальна навчальна система підготовки диспетчерів управління повітряним рухом. *Системи обробки інформації*, 2(161), 29-36. <https://doi.org/10.30748/soi.2020.161.04>
- Трач, Ю. (2017). VR-технології як метод і засіб навчання. *Освітологічний дискурс*, 3(18), 309-322.
- Фуртат, Ю.О. (2013). Функціональні основи засобів адаптації користувачьких інтерфейсів до когнітивних особливостей користувачів в автоматизованих системах. *Математичне та комп'ютерне моделювання. Серія: Технічні науки*, 8, 120-131.
- Шведова М.О. (2022). Застосування корпусу у викладанні української мови як іноземної. *Слов'янські мови, культури, переклад і викладання: сучасний стан і перспективи*, 227-232.
- Шипнівська, О. (2013). Визначення типів синтаксичної неоднозначності у знаннєорієнтованій системі машинного перекладу. *Українське мовознавство*, 43, 104-113.

- Шомко, А.В. (2017). Генеративна грамати́ка як галузь лінгвістичної науки. Проблеми та перспективи її розвитку. *Сучасні філологічні дослідження та навчання іноземної мови в контексті міжкультурної комунікації*, 10, 355-359.
- Штанько, В.І. (2012). Віртуальний комунікативний простір і проблеми самоідентифікації особистості. *Вісник Харківського національного університету імені В.Н. Каразіна. Серія: Теорія культури і філософія науки*, 1029(1), Вип. 47, 5-12.
- Ялова, К.М., Яшина, К.В., Говорущенко, Т.О., & Тарасюк, О.С. (2021). Сентимент аналіз засобами нейронної мережі. *Математичне моделювання*, 1(44), 30-37.
- Bolshakov, I., & Gelbuch, A. (2004). *Computational Linguistics: Models, Resources, Applications*. Electronic edition IPN-UNAM-FCE.
- Clark, A., Fox, C., & Lappin, S. (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. Blackwell Publishing Ltd.
- Dunlop, C., & Fetzer, J. (1993). *Glossary of Cognitive Science*. Saint Paul, MN: Paragon House.
- Gallego, A.J., & Martin R. (2020). *Language, Syntax, and the Natural Sciences*. Cambridge University Press.
- Gillis, A.S. (2022). *What is Computational Linguistics?* Режим доступу: <https://www.techtarget.com/searchenterpriseai/definition/computational-linguistics-CL>
- Grishman, R. (1986). *Computational Linguistics: An Introduction*. Cambridge University Press.
- Hammond, M. (2020). *Python for Linguists*. Cambridge University Press.
- Hausser, R. (2014). *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Heidelberg: Springer.
- Hunston, S. (2022). *Corpora in Applied Linguistics*. Cambridge University Press.
- McCulloch, G. (2020). *Because Internet: Understanding how Language is Changing*. Vintage Publishing.
- Moens, J. (September 8, 2022). An AI can decode speech from brain activity with surprising accuracy. *ScienceNews. Neuroscience*.
- Oxley, T. (June 2, 2022). A brain implant that turns your thoughts into text. *TED2022. TED: Ideas Worth Spreading*.
- Peters, S., & Shrobe, H.E. (2003). Using semantic networks for knowledge representation in an intelligent environment. *IEEE International Conference on Pervasive Computing and Communications (PerCom'03)*, 1, 323-337. doi: 10.1109/PERCOM.2003.1192756
- Poole, G. (2011). *Syntactic Theory*. Bloomsbury Publishing PLC.

- Powell, M. (2019). *Grammar Geek*. Octopus Publishing Group.
- Powell, M. (2019). *Word Nerd*. Octopus Publishing Group.
- Schubert, L. (2020). Computational Linguistics. *The Stanford Encyclopedia of Philosophy*, E.N. Zalta (Ed.). Режим доступу: <https://plato.stanford.edu/archives/spr2020/entries/computational-linguistics/>
- Shvedova, M.O. (2020). The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorus.org): Architecture and Functionality. *The 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020)*. Lviv. P. 489-506.
- Vakhovska, O.V. (2021). Metaphors of Emotions: Towards a Data-driven Formalization. *Вісник Київського національного лінгвістичного університету. Серія Філологія*, 24(1), 33-48. doi: 10.32589/2311-0821.24%20(1).2021.236030
- Vakhovska, O.V. (2022). Must the Pot of Seven Holes be a Translation Theorist's Riddle? *KELM Knowledge, Education, Law, Management*, 4(48), 117-125. DOI <https://doi.org/10.51647/kelm.2022.4.19>

ДОДАТКОВА

- Карпенко, Ю.О. (2006). *Вступ до мовознавства*. Київ: Академія.
- Кочерган, М.П. (2006). *Вступ до мовознавства: Підручник*. Київ: Академія.
- Селіванова, О.О. (2008). *Сучасна лінгвістика: напрями та проблеми: Підручник*. Полтава: Довкілля-К.
- Aist, G., & Mostow, J. (2009). Predictable and Educational Spoken Dialogues: Pilot Results. In *Proceedings of the 2009 ISCA Workshop on Speech and Language Technology in Education (SLaTE 2009)*. Birmingham, UK: University of Birmingham.
- Akmajian, A., Deemers, R.A., Farmer, A.K., & Harnish, R.M. (1995). *Linguistics. An Introduction to Language and Communication*. MIT Press.
- Ambite, J.-L., Chaudhri, V.K., Fikes, R., Jenkins, J., Mishra, S., Muslea, M., Uribe, T., & Yang, G. (2006). Design and Implementation of the CALO Query Manager. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, July 16-20, Boston, MA; Menlo Park, CA: AAAI Press, 1751-1758.
- Andersen, P.M., Hayes, P.J., Huettner, A.K., Schmandt, L.M., Nirenburg, I.B., & Weinstein, S.P. (1992). Automatic Extraction of Facts from Press Releases to Generate News Stories. In *Proceedings of the 3rd Conference on Applied Natural Language Processing (ANLC '92)*, Trento, Italy, March 31-April 3. Stroudsburg, PA: Association for Computational Linguistics (ACL), 170-177.

- Androutsopoulos, I., & Ritchie, G. (2000). Database Interfaces. In Dale, R., Somers, H., & Moisl H. (eds.), *Handbook of Natural Language Processing*, Chapter 9. Boca Raton, FL: CRC Press.
- Ayuso, D., Bates, M., Bobrow, R., Meteer, M., Ramshaw, L., Shaked, V., & Weischedel, R. (1990). *Research and Development in Natural Language Understanding as Part of the Strategic Computing Program*. BBN Report No. 7191, BBN Systems and Technologies, Cambridge, MA.
- Baker, S. (2011). *Final Jeopardy*. Boston: Houghton Mifflin Harcourt.
- Bickmore, T., Schulman, D., & Sidner, C. (2011). Modeling the Intentional Structure of Health Behavior Change Dialogue. *Journal of Biomedical Informatics*, 44, 183-197.
- Bourne, C.P., & Hahn, T.B. (2003). *A History of Online Information Services, 1963-1976*. Cambridge, MA: MIT Press.
- Boyer, K.E., Ha, E.Y., Wallis, M.D., Phillips, Vouk, R.M.A., & Lester, J.C. (2009). Discovering Tutorial Dialogue Strategies with Hidden Markov Models. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED 2009)*, Brighton, U.K.: IOS Press, 141-148.
- Bühler, D., & Minker, W. (2011). *Domain-Level Reasoning for Spoken Dialogue Systems*. Boston, MA: Springer.
- Burgard, W., Cremers, A.B., Fox, D., Hahnel, D., Lakemeyer, G., Schulz, D., Steiner, W., & Thrun, S. (1999). Experiences with an Interactive Museum Tour-Guide Robot. *Artificial Intelligence*, 114(1-2): 3-55.
- Callaway, C., Dzikovska, M., Farrow, E., Marques-Pita, M., Matheson, C., & Moore, J. (2007). The Beetle and BeeDiff Tutoring Systems. In *Proceedings of the 2007 Workshop on Spoken Language Technology for Education (SLaTE)*, Farmington, PA, Oct. 1-3. Carnegie Mellon University and ISCA Archive.
- Campbell, J., Core, M., Artstein, R., Armstrong, L., Hartholt, A., Wilson, C., Georgila, K., Morbini, F., Haynes, E., Gomboc, D., Birch, M., Bobrow, J., Lane, H.C., Gerten, J., Leuski, A., Traum, D., Trimmer, M., DiNinni, R., Bosack, M., Jones, T., Clark, R.E., & Yates, K.A. (2011). Developing INOTS to Support Interpersonal Skills Practice. In *Proceedings of the 32nd Annual IEEE Aerospace Conference (IEEEAC)*, Big Sky, MT, March 5-12, Institute of Electrical and Electronics Engineers (IEEE), 3222-3235.
- Cercone, N., McFetridge, P., Popowich, F., Fass, D., Groeneboer, C., & Hall, G. (1993). *The SystemX Natural Language Interface: Design, Implementation, and Evaluation*. Tech. Rep. CSS-IS TR 93-03, Centre for Systems Science, Simon Fraser University, Burnaby, BC, Canada.
- Cole, R., Wise, B., & van Vuuren, S. (2007). How Marni Teaches Children to Read. *Educational Technology*, 47(1), 14-18.

- Coles, L.S. (1972). *Techniques for Information Retrieval Using an Inferential Question-Answering System with Natural-Language Input*. Technical Note 74, SRI Project 8696, SRI International.
- Damerau, F.J. (1981). Operating Statistics for the Transformational Question Answering System. *American Journal of Computational Linguistics*, 7(1), 30-42.
- DeJong, G.F. (1982). An overview of the FRUMP system. In W.G. Lehnert and M.H. Ringle (eds.), *Strategies for Natural Language Processing*, Erlbaum, 149-176.
- Ferguson, G., & Allen, J.F. (1998). TRIPS: An integrated intelligent problem-solving assistant. In *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98)*. Menlo Park, CA: AAAI Press, 567-573.
- Ferrucci, D. (2012). This is Watson. *IBM Journal of Research and Development*, 56, 3-4.
- Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A.A., Lally, A., Murdock, J.W., Nyberg, E., Prager, J., Schlaefter, N., & Welty, C. (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine*, 31(3), 59-79.
- Fromkin, V. (2000). *Linguistics. An Introduction to Linguistic Theory*. Blackwell. Oxford.
- Harris, L.R. (1984). Experience with INTELLECT: Artificial Intelligence Technology Transfer. *AI Magazine*, 5(2), 43-50.
- Hendrix, G.G., Sacerdoti, E.D., Sagalowicz, D., & Slocum, J. (1978). Developing a Natural Language Interface to Complex Data. *ACM Transactions on Database Systems*, 3(2), 105-147.
- Hobbs, J.R., Appelt, D.E., Bear, J., Kameyama, M., Stickel, M.E., & Tyson, M. (1997). FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text. In E. Roche and Y. Schabes (eds.), *Finite-State Language Processing*, Cambridge, MA: MIT Press, 383-406.
- Jordan, P., Makatchev, M., Pappuswamy, U., VanLehn, K., & Albacete, P. (2006). A Natural Language Tutorial Dialogue System for Physics. In G.C.J. Sutcliffe and R.G. Goebel (eds.), *Proceedings of the 19th International Florida Artificial Intelligence Research Society (FLAIRS-06)*. Menlo Park, CA: AAAI Press.
- O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2006). *Contemporary Linguistics. An Introduction. 5th edition*. Bedford/St.Martin's, Boston & New York.
- Massaro, D.W., Cohen, M.M., Tabain, M., Beskow, J., & Clark, R. (2012). Animated Speech: Research Progress and Applications. In G. Bailly, P. Perrier, and E. Vatiokis-Bateson (eds.), *Audiovisual Speech Processing*, Cambridge, UK: Cambridge University Press, 309-345.

- Maybury, M.T. (ed.). (2004). *New Directions in Question Answering*. Cambridge, MA: AAAI and MIT Press.
- Movellan, J.R., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable Robot Improves Toddler Vocabulary Skills. In *Proceedings of the International Conference on Human Robot Interaction (HRI2009)*, San Diego, CA, March 11-13. New York: ACM Digital Library.
- Pinker, S. (1994). *The Language Instinct*. London: Penguin.
- Pereira, F.C.N., & Warren, D.H.D. (1982). An Efficient Easily Adaptable System for Interpreting Natural Language Queries. *American Journal of Computational Linguistics*, 8(3-4), 110-122.
- Pulman, S., Boye, J., Cavazza, M., Smith, & Santos de la Cámara, C.R. (2010). "How Was Your Day?" In *Proceedings of the 2010 Workshop on Companionable Dialogue Systems*, Uppsala, Sweden, July. Stroudsburg, PA: Association for Computational Linguistics, 37-42.
- Radford, A., Atkinson, M., Britain, D., Clahsen, H., & Spencer, A. (1999). *Linguistics. An Introduction*. Cambridge UP.
- Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*. Boston, MA: Addison-Wesley.
- Scheutz, M., Cantrell, R., & Schermerhorn, P. (2011). Toward Human-Like Task-Based Dialogue Processing for HRI. *AI Magazine*, 32(4), 77-84.
- Thompson, F.B., Lockemann, P.C., Dostert, B.H., & Deverill, R. (1969). REL: A Rapidly Extensible Language System. In *Proceedings of the 24th ACM National Conference*, New York: ACM Digital Library, 399-417.
- Thompson, F.B. & Thompson, B.H. (1975). Practical Natural Language Processing: The REL System as Prototype. In M. Rubinoff and M.C. Yovits (eds.), *Advances in Computers*, vol. 13, New York: Academic Press, 109-168.
- Wilks, Y. (2010). Is a Companion a Distinctive Kind of Relationship with a Machine? In *Proceedings of the 2010 Workshop on Companionable Dialogue Systems, (CDS '10)*, Uppsala, Sweden. Stroudsburg, PA: Association for Computational Linguistics (ACL), 13-18.
- Woods, W.A., Kaplan, R.M., & Nash-Webber, B.L. (1972). *The Lunar Sciences Natural Language Information System: Final Report*. BBN Report No. 2378, Bolt Beranek and Newman Inc., Cambridge, MA.
- (Доповніть список тією літературою, яку Ви мали нагоду додатково опрацювати протягом вивчення курсу.)

ЕЛЕКТРОННІ РЕСУРСИ

ОСНОВНІ

[Вікіпедія](#)

[Головна | Освіта Microsoft](#)

[ГРАК](#)

[Корпус української мови](#)

[AlphaGo](#)

[Coursera](#)

[DOU](#)

[Facebook, Carnegie Mellon build first AI that beats pros in 6-player poker](#)

[Grammarly](#)

[GitHub](#)

[IBM Training](#)

[IBM Watson](#)

[Language and the Limits of Understanding - International Translation Day 2021](#)

[LinkedIn](#)

[Michael Page](#)

[Ontohub.org](#)

[Stanford NLP Group](#)

[Training | Microsoft Learn](#)

[Udemy](#)

[Work.ua](#)

ДОДАТКОВІ

(Доповніть список тими електронними ресурсами, із якими Ви мали нагоду додатково ознайомитися протягом вивчення курсу.)



A series of horizontal lines for writing, consisting of 28 evenly spaced lines that span the width of the page.

Підписано до друку 28.03.2023 р. Формат 60x84 1/16
Спосіб друку офсетний. Умовн. друк. арк. 9,52
Умовн. фарбо-відб. 9,63 Обл.-вид. арк. 9,63
Тираж 100. Зам. № 23 - 11

Видавничий центр КНІУ
Свідоцтво: серія ДК 1596 від 08.12.2003 р.

Віддруковано "Видавництво Ліра-К"
03115, Київ, вул. Василя Стуса 22/1
Свідоцтво про внесення до державного реєстру
Серія ДК № 3981.