

КОМП'ЮТИНГ

Пасічник В.В., Щербина Ю.М.,
Висоцька В.А., Шестакевич Т.В.

Математична лінгвістика

Навчальний посібник



Computing

Міністерство освіти і науки, молоді та спорту України

**В. В. Пасічник, Ю. М. Щербина, В. А. Висоцька,
Т. В. Шестакевич**

МАТЕМАТИЧНА ЛІНГВІСТИКА

Книга 1. Квантитативна лінгвістика

НАВЧАЛЬНИЙ ПОСІБНИК

СЕРІЯ «КОМП'ЮТИНГ»

За науковою редакцією д.т.н., професора В. В. Пасічника

Рекомендовано Міністерством освіти і науки, молоді та спорту України

**Львів
Новий Світ - 2000
2012**

ББК 22.11я73
УДК 519 76(075 8)
М34

Відтворення цієї книги або будь-якої її частини заборонено без письмової згоди видавництва. Будь-які спроби порушення авторських прав будуть переслідуватися у судовому порядку.

*Гриф надано Міністерством освіти і науки України
(Лист № 1/11-Г-9785 від 22.10.2011 р.)*

Рецензенти:

Цегелик Г.Г., доктор фіз.-мат. наук, професор. Львівський національний університет імені Івана Франка;

Камінський Р.М., доктор технічних наук, професор. Буковинський університет;

Пелейцишин А.М., доктор технічних наук, Національний університет «Львівська політехніка».

М 34 В.В. Пасічник, Ю.М. Щербина, В.А. Висоцька, Т.В. Шестакевич

Математична лінгвістика. [Книга 1. Квантитативна лінгвістика] : [Навчальний посібник]. - Львів : «Новий світ -2000», 2012. -359 [1] с.

ISBN 978-966-418-191-1

Навчальний посібник містить матеріал, необхідний для вивчення основних теоретичних засад, функціональних можливостей та практичного застосування теорії математичної лінгвістики в дослідженні мов, розроблення прикладних засобів та інформаційних систем аналізу та опрацювання природних мов. У книзі першій розглянуто питання квантитативної лінгвістики. Викладення матеріалу супроводжується значною кількістю прикладів, що полегшує його сприйняття і засвоєння. Подається перелік запитань і тестів для самоконтролю, а також завдання для самостійного виконання (трьох рівнів складності) та довідкова інформація для розв'язування задач.

Навчальний посібник призначається для студентів, що навчаються за напрямами галузей знань «Інформатика та обчислювальна техніка» (напрямок 6.050101 «Комп'ютерні науки»), «Системні науки та кібернетика» (напрямок 6.040303 «Системний аналіз») і споріднених галузей знань, пов'язаних з вивченням прикладної лінгвістики та інформаційних технологій. Може бути використаний аспірантами як підґрунтя для наукових досліджень і викладачами як дидактичний матеріал, а також для самостійного вивчення і підвищення кваліфікації.

ББК 22.11я73
УДК 519.76(075.8)
М 34

ISBN 978-966-418-191-1

© В.В. Пасічник, Ю.М. Щербина,
В.А. Висоцька, Т.В. Шестакевич, 2011
© «Новий світ-2000», 2012

З М І С Т

ПЕРЕДМОВА НАУКОВОГО РЕДАКТОРА СЕРІЇ ПІДРУЧНИКІВ ТА НАВЧАЛЬНИХ ПОСІБНИКІВ «КОМП'ЮТИНГ».....	7
ВСТУПНЕ СЛОВО АВТОРІВ.....	12
РОЗДІЛ 1.....	18
МЕТОДИ КВАНТИТАТИВНОЇ ЛІНГВІСТИКИ.....	18
1.1. Комбінаторний аналіз.....	18
1.1.1. Множина. КORTEЖ. Декартів добуток.....	19
1.1.2. Операції над множинами. Доведення рівностей з множинами.....	21
1.1.3. Взаємно однозначна відповідність.....	23
1.1.4. Основні правила комбінаторики. Розміщення і сполучення.....	24
1.1.5. Обчислення кількості розміщень і сполучень.....	26
1.1.6. Перестановки.....	28
1.1.7. Задача про цілочислові розв'язки.....	29
1.1.8. Розв'язування комбінаторних задач.....	30
1.1.9. Контрольні питання.....	30
1.1.10. Тести для самоконтролю.....	30
1.1.11. Задачі для самостійної роботи.....	32
1.2. Елементи теорії ймовірностей.....	53
1.2.1. Елементарні та складні лінгвістичні події.....	54
1.2.2. Операції над лінгвістичними подіями.....	54
1.2.3. Ймовірність елементарної лінгвістичної події.....	56
1.2.4. Ймовірності складних лінгвістичних подій.....	58
1.2.5. Контрольні питання.....	65
1.2.6. Тести для самоконтролю.....	65
1.2.7. Задачі для самостійної роботи.....	67
1.3. Основи математичної статистики.....	68
1.3.1. Первинне статистичне опрацювання тексту.....	69
1.3.2. Методи організації статистичного спостереження над текстом.....	70
1.3.3. Варіаційні ряди лінгвістичних ознак.....	71
1.3.4. Порядкова ознака в лінгвістичних одиницях.....	77
1.3.5. Якісна ознака в лінгвістичних одиницях.....	77
1.3.6. Графічне зображення лінгвістичних варіаційних рядів.....	78
1.3.7. Статистичні характеристики лінгвістичних варіаційних рядів.....	81
1.3.8. Середнє арифметичне і дисперсія для декількох сукупностей.....	90
1.3.9. Довжина слововживання як статистично-стилістичний параметр.....	92
1.3.10. Статистична однорідність тексту.....	93
1.3.11. Двовимірний статистичний розподіл вибірки і його числові характеристики.....	94
1.3.12. Контрольні питання.....	96
1.3.13. Тести для самоконтролю.....	96
1.3.14. Задачі для самостійної роботи.....	98

1.4.	Елементи теорії інформації в лінгвістиці.....	100
1.4.1.	Кількісні міри інформації.....	101
1.4.2.	Інформаційні виміри кодування інформації.....	106
1.4.3.	Умовна ентропія та ентропія об'єднання.....	110
1.4.4.	Розрахунок швидкості передавання інформації та пропускної здатності інформаційних каналів зв'язку.....	125
1.4.5.	Визначення надлишковості повідомлень. Оптимальне кодування.....	133
1.4.6.	Контрольні питання.....	137
1.4.7.	Тести для самоконтролю.....	137
1.4.8.	Задачі для самостійної роботи.....	139
РОЗДІЛ 2	149
ЙМОВІРНІСНЕ МОДЕЛЮВАННЯ ТЕКСТУ ТА ЙОГО СКЛАДОВИХ	149
2.1.	Види випробувань у тексті.....	150
2.1.1.	Повторення незалежних випробувань у тексті.....	150
2.1.2.	Вибірki без повторень та з повтореннями.....	151
2.2.	Три схеми незалежних лінгвістичних випробувань.....	151
2.2.1.	Проста схема незалежних випробувань. Формула Бернуллі.....	152
2.2.2.	Поліноміальна схема.....	154
2.2.3.	Пуасонівська схема.....	155
2.3.	Лінгвістична вибірка без повторень та її опис за допомогою формули Бернуллі.....	157
2.4.	Ймовірність появи події в заданому діапазоні кількості появ.....	159
2.5.	Визначення необхідного обсягу вибірки.....	161
2.6.	Метод контент-аналізу текстової інформації.....	162
2.7.	Контрольні питання.....	173
2.8.	Задачі для самостійної роботи.....	174
РОЗДІЛ 3	176
ІНФОРМАЦІЯ ЛІНГВІСТИЧНИХ ПОДІЙ	176
3.1.	Формально-структурні характеристики інформації.....	178
3.1.1.	Трактування поняття «інформація».....	178
3.1.2.	Основи теорії інформації.....	179
3.1.3.	Якісні властивості інформації.....	181
3.2.	Дійсні числа.....	182
3.2.1.	Поняття числа.....	182
3.2.2.	Множина дійсних чисел.....	183
3.3.	Моделювання інформаційної побудови мови.....	183
3.3.1.	Зміна синтаксичної інформації в мові.....	183
3.3.2.	Розподіл інформації в слові.....	183
3.3.3.	Розподіл інформації в тексті.....	184
3.3.4.	Розподіл контекстної обумовленості.....	185
3.4.	Інформаційні моделі слова і тексту.....	186

3.5. Інформаційні виміри в тексті.....	1 ⁸⁸
3.5.1. Кількість інформації.....	1 ⁸⁸
3.5.2. Приклади вимірювання кількості інформації.....	191
3.6. Синтаксична інформація в тексті і словах.....	193
3.6.1. Особливості синтаксичної інформації.....	193
3.6.2. Максимальний обсяг інформації в слові.....	193
3.6.3. Оцінювання максимального обсягу інформації слова.....	195
3.6.4. Ряд та діяхронічний процес.....	196
3.7. Приріст та накопичення інформації в тексті.....	197
3.8. Контекстна обумовленість та надлишковість тексту.....	200
3.9. Вимірювання змістовної інформації в тексті.....	202
3.9.1. Кількісні оцінювання змістовної інформації.....	202
3.9.2. Застосування методів аналізу побудови тексту.....	205
3.9.3. Організація пам'яті для інформаційного пошуку.....	209
3.9.4. Методи визначення емоційної тональності тексту.....	218
3.10. Контрольні питання.....	221
3.11. Задачі для самостійної роботи.....	222
РОЗДІЛ 4.....	224
СТАТИСТИЧНА ЛІНГВІСТИКА.....	224
4.1. Лінгвометрія.....	225
4.1.1. Статистичні параметри мови.....	225
4.1.2. Кількісні оцінки мовлення.....	227
4.2. Стилеметрія.....	230
4.3. Глоттохронологія.....	239
4.4. Контрольні питання.....	247
4.5. Задачі для самостійної роботи.....	248
РОЗДІЛ 5.....	249
КРИПТОЛІНГВІСТИКА.....	249
5.1. Основні визначення криптолінгвістики.....	250
5.2. Лінгвістичні схеми у криптографії.....	253
5.2.1. Кодування інформації.....	253
5.2.2. Моноалфавітні шифри.....	256
5.2.3. Застосування перестановок.....	261
5.2.4. Поліалфавітні шифри.....	262
5.2.5. Книжкові шифри.....	266
5.2.6. Дволітерні шифри.....	267
5.3. Основи криптоаналізу. Лінгвістичне дешифрування.....	270
5.3.1. Розкриття шифру Цезара.....	270
5.3.2. Розкриття шифру заміни.....	271
5.3.3. Розкриття книжкового шифру.....	283
5.3.4. Розкриття шифру Віженера.....	287

5.4. Контрольні питання.....	289
5.5. Задачі для самостійної роботи.....	290
ДОДАТКИ.....	291
Додаток А.....	291
Додаток Б.....	295
Додаток В.....	295
Додаток Г.....	301
Додаток Д.....	303
Додаток Е.....	304
Додаток Ж.....	306
Додаток И.....	337
Додаток К.....	339
Додаток Л.....	341
ПРЕДМЕТНИЙ ПОКАЖЧИК.....	350
СПИСОК ЛІТЕРАТУРИ.....	351