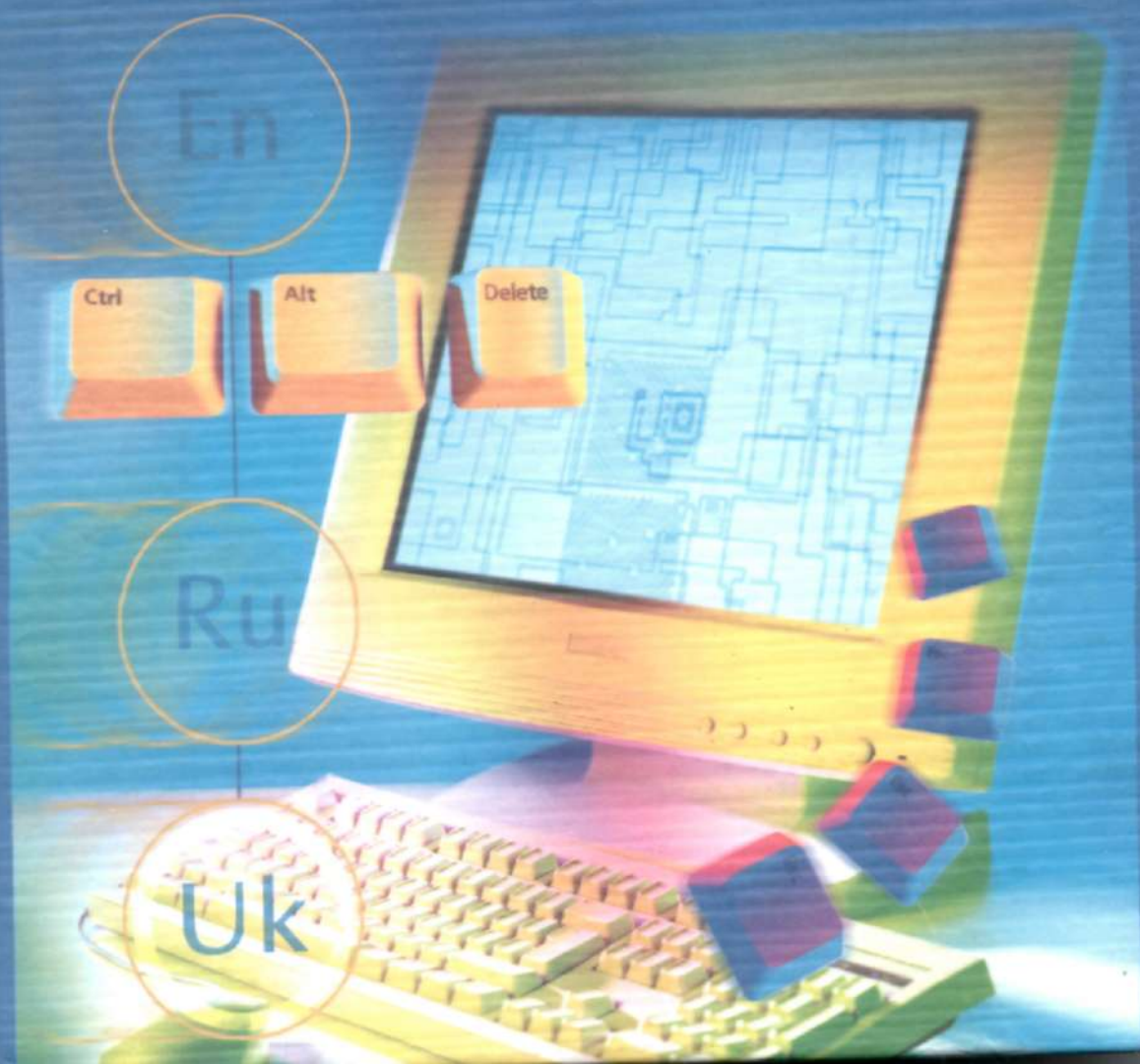


004.43
В68

ДОЛШИН

КОМП'ЮТЕРНА ЛІНГВІСТИКА

Навчальний посібник



В.Г. Волошин

004.43
B68

КОНТРОЛЬНИЙ
ПРИМІРНИК

Комп'ютерна лінгвістика

Навчальний посібник

Рекомендовано Міністерством освіти і науки України
як навчальний посібник для студентів вищих
навчальних закладів



1557822



Суми
«Університетська книга»
2004

УДК 004.43(075.8)
ББК 32.973-018.1я73
В68

Рекомендовано до друку вченою радою факультету романо-германської філології ОНУ ім. І.І. Мечникова. Протокол № 6 від 19.03.2002

Рецензенти:

А.Й. Багмут, доктор філологічних наук, професор;
Т.М. Корольова, доктор філологічних наук, професор;
В.Г. Таранець, доктор філологічних наук, професор

Гриф надано Міністерством освіти і науки України.
Лист № 14/18.2-2168 від 12.12.03

Волошин В.Г.

В68 Комп'ютерна лінгвістика: Навчальний посібник. – Суми: ВТД «Університетська книга», 2004. – 382 с.

ISBN 966-680-134-5

Посібник містить систематичний виклад основ комп'ютерної лінгвістики – одного з новітніх напрямів інформаційної технології сучасного мовознавства. Описано методологію та інструментальні засоби організації процесу набуття знань.

Основна увага приділяється питанням добору і формування лінгвістичних структур, що супроводжують процес викладу наукових положень комп'ютерної лінгвістики, яка лежить на стику багатьох наук – від технічної кібернетики до прикладної лінгвістики.

Для студентів відділень теоретичної і прикладної лінгвістики університетів і вузів, студентів усіх спеціальностей, що цікавляться проблемами комп'ютерної обробки текстів і усного мовлення. Може бути корисним широкому колу фахівців, які вирішують завдання побудови систем прикладного характеру в галузі штучного інтелекту, машинного перекладу, аналізу і синтезу мови, створення інформаційно-пошукових систем, евристики, експертних систем та інших технічних додатків лінгвістики.

ББК 32.973-018.1я73

ISBN 966-680-134-5

© Волошин В.Г., 2004
© ТОВ «ВТД «Університетська книга», 2004

ЗМІСТ

Вступ	6
Розділ 1. Основні наукові напрями комп'ютерної лінгвістики і її зв'язок з іншими науками	11
1.1. Взаємозв'язок комп'ютерної лінгвістики з іншими науками	11
1.2. Комп'ютерна лінгвістика і прикладні завдання мовознавства	14
1.3. Інформатика і комп'ютерна лінгвістика	19
Розділ 2. Моделювання мовленнєвої діяльності та основні інформаційні елементи створення тексту	22
2.1. Моделювання мовленнєвої діяльності в лінгвістиці	22
2.2. Інформаційні елементи тексту в усній та письмовій формі	36
Розділ 3. Автоматичний інформаційний пошук	43
3.1. Поняття інформаційно-пошукової системи	43
3.2. Автоматичне індексування, реферування й анотування	51
3.3. Загальні принципи пошуку та одержання інформації в пошукових системах Internet	54
Розділ 4. Акустична структура звуків мови	64
4.1. Загальні принципи організації звукової мовленнєвої комунікації	64
4.2. Організація і принцип роботи мовленнєвого тракту людини	67
4.3. Основні положення теорії сприйняття звуків мови людиною	73
Розділ 5. Автоматичне введення звуків мови і його розпізнавання комп'ютером	80
5.1. Аналого-цифрове перетворення мовленнєвого сигналу	80
5.2. Проблеми автоматичного розпізнавання звуків усного мовлення ..	83
5.3. Машинні методи аналізу і розпізнавання елементів усного мовлення	87
5.3.1. Первинна обробка мовленнєвого сигналу для розпізнавання звуків усного мовлення при введенні в комп'ютер	87
5.3.2. Машинні методи розпізнавання окремих звуків усного мовлення	89
5.3.3. Звуковий спектрограф і видима мова	91
5.3.4. Методи спектрально-часового аналізу усного мовлення	92
5.3.5. Метод двійкового добору (бінарної селекції) звукових елементів мови	96
5.3.6. Метод формантного аналізу мови	100

5.3.7. Метод розпізнавання звуків усного мовлення за допомогою використання лінгвістичної інформації	102
5.3.8. Шляхи вирішення проблеми автоматичного розпізнавання усного тексту за допомогою комп'ютерних технологій	105
Розділ 6. Інформаційна оцінка знакових елементів аркуша	111
Розділ 7. Обробка письмової лінгвістичної інформації на рівні словоформ, слів, словосполучень, речення, тексту	124
7.1. Машинна морфологія	124
7.2. Методи автоматичного морфологічного аналізу тексту	126
7.3. Морфологічний аналіз зі словником словоформ	128
7.4. Омографія основ і афіксів	130
7.5. Морфологічний аналіз методом логічних множин	131
7.6. Автоматичний синтаксичний аналіз тексту	134
7.7. Автоматичний семантичний аналіз тексту	142
Розділ 8. Когнітивна наука як наукова дисципліна моделювання людського мислення	154
Розділ 9. Штучний інтелект	165
9.1. Основні принципи створення інтелектуальних систем (ШІ)	165
9.2. Логічні закони рішення систем ШІ	174
9.3. Евристика	180
9.4. Мова програмування PROLOG	183
9.5. Розпізнавання образів	184
9.6. Експертні системи (ЕС)	188
9.7. Опис ЕС аналізу навчання в школі	191
9.8. Комп'ютерні навчальні системи	194
9.8.1. Загальні теоретичні положення систем навчання іноземним мовам	194
9.8.2. Основні напрями розвитку навчальних комп'ютерних систем	202
9.8.3. Програмне і методичне забезпечення систем комп'ютерного навчання іноземній мові	211
9.9. Робототехніка і мовне керування	220
Розділ 10. Мовні технології і синтез мови	228
10.1. Розвиток систем, які синтезують усне мовлення	228
10.1.1. Метод кодування відновлення форми сигналів	233
10.1.2. Фонетичний синтез мовлення за допомогою аналогових і цифрових формантних синтезаторів	234
10.1.3. Синтез мовлення за методом лінійного предиктивного кодування (ЛПК)	239
10.2. Алгоритм синтезу усного мовлення за текстом	244
10.2.1. Загальна структура синтезатора мовлення за текстом	244
10.2.2. Лінгвістичний процесор	246
10.2.3. Просодичний процесор	254

10.2.4. Фонетичний процесор	261
10.2.5. Акустичний процесор	266
10.2.6. Компіляційні синтезатори	271
10.3. Сучасні технології мовлення	274
Розділ 11. Машинний переклад і комп'ютерна лексикографія	286
11.1. Загальні знання про машинний переклад	286
11.2. Оцінка якості перекладу	290
11.3. Типові помилки машинного перекладу	291
11.4. Сучасний стан машинного перекладу	297
11.5. Аналіз деяких теоретичних моделей процесу перекладу	304
11.6. Загальна структура системи машинного перекладу	313
11.6.1. Робота алгоритму аналізу і синтезу	322
11.6.2. Види перекладу та характеристики сучасних машинних програм-перекладачів	325
11.7. Комп'ютерна лексикографія	331
11.7.1. Предмет, методи і теорія обчислювальної лексикографії ..	331
11.7.2. Проблеми комп'ютерної лексикографії	336
11.7.3. Термінографія	342
Розділ 12. Лінгвістичні проблеми термінології	345
Термінологічний словник	361
Список основної літератури	371
Список додаткової літератури	374
Інтернет-адреси для лінгвістів	380