

Впровадження програмних засобів і методів електронного документообігу в комп'ютерній лінгвістиці

Світлана Трішина

Кафедра інформатики, Кременчуцький університет економіки, інформаційних технологій і управління, УКРАЇНА, м.Кременчук, вул.Пролетарська, 24/37, E-mail: mif506@yandex.ru

Scanning questions, recognitions, are considered in the article, translating of documents by development and introduction of programmatic facilities and methods of electronic dokumentooborota in linguistics.

Ключові слова – linguistics, komp'yuternyy translation, WEB-page.

I. Вступ

Прикладна лінгвістика традиційно пов'язана з різноманітними аспектами лінгвістичного забезпечення людської комунікації, які можна звести до єдиної проблеми оброблення інформації, що функціонує у суспільстві у вигляді текстів. Це проблеми лінгводидактики мов, орфографії, транслітерування, перекладу, термінології, дешифрування тощо.

Комп'ютеризація інформаційної сфери стимулює розвиток нового напрямку прикладної лінгвістики – комп'ютерного, до компетенції якого входить лінгвістичне забезпечення функціонування інформації в автоматизованих системах різних типів[1].

На сучасному етапі можна без особливих труднощів навести приклади використання інформаційних технологій у всіх галузях: від освіти і до менеджменту. Сьогодні успіх буде мати та фірма, той заклад, який володіє найсучаснішими комп'ютерними технологіями. Значного прогресу можна досягти і в галузі освіти з впровадженням відповідних інформаційних комп'ютерних технологій, які зможуть зробити процес здобуття освіти більш гнучким, індивідуалізованим і одночасно нададуть змогу студентам використовувати глобальні ресурси для навчання, спілкуватись та обмінюватись досвідом із студентами інших міст, країн тощо.

Комп'ютерні сучасні інформаційні технології передусім мають на увазі технології створення, передачі та обробки даних за допомогою персональних комп'ютерів [2].

Оскільки переклад займає одну з центральних позицій серед тематичних розділів лінгвістичної діяльності, то комплексний підхід до аналізу лексичних особливостей перекладу є досить актуальним за сучасних умов. Одним із прикладів комплексної інформаційної технології є сканування, розпізнавання та переклад текстів з друкованих оригіналів.

II. Автоматизація сканування і розпізнавання тексту на комп'ютерах

Перетворення паперового документа в електронний проходить в три етапи. Кожний із цих етапів програми може виконуватися, як автоматично, так і під контролем користувача. Якщо всі етапи проходять автоматично, то перетворення документа відбувається за один прийом.

Перший етап роботи – сканування. На цьому етапі завжди використовується сканер. Однак зображення з листка паперу може бути перетворено в цифрову форму і за допомогою інших засобів, таких, як наприклад, цифрові фотоапарати і цифрові відеокамери. В результаті сканування створюється графічний файл, в якому зберігається растрове зображення первинного документа.

Другий етап роботи – сегментація тексту, тобто розбивка тексту на блоки.

Останній етап роботи програми – розпізнавання.

Більшість програм розпізнають текст на декількох мовах. Один і той же алгоритм можна використовувати для розпізнавання російського, латинського, арабського і других алфавітів і навіть змішаних текстів. Найбільш розповсюджені програми: Fine Reader і Cunei Form. Програма Fine Reader має ряд можливостей:

- дозволяє об'єднати сканування і розпізнавання в одну операцію;
- розпізнає неправильно надруковані тексти і складні шрифти;
- обробляє пакети документів і бланки;
- редагує текст і перевіряє його орфографію;
- працює з різними моделями сканерів.

III. Автоматичний переклад документів

За допомогою мультимедійних технологій створені електронні енциклопедії та спеціалізовані словники, які містять термінологію пов'язану з різними галузями науки.

Програмні засоби автоматичного перекладу можна умовно розбити на дві основні категорії. Першу категорію представляють комп'ютерні словники, які розробляються спеціальними програмами – базами даних, комп'ютерними картотеками, програмами обробки текстів. Ці програми дозволяють в автоматичному режимі формувати словарні статті, зберігати словникову інформацію та обробляти її.

В теперішній час в комп'ютерній лексикографії під електронним словником прийнято розуміти “словник, процедури укладання якого здійснює комп'ютер” [3].

Створення електронного словника може розбиватися на такі етапи:

1. Формування словника.
2. Складання корпусу прикладів (з використанням електронних корпусів тексту).
3. Написання словникових статей.
4. Копіювання словникових статей в базу даних.
5. Коректура тексту словника в базі даних.
6. Словник.

Найбільш розповсюджені електронні словники - Lingvo, ProlingOffice, Multilex. Перевага комп'ютерних словників складається з швидкості доступу і вигоди автоматичного пошуку виділеного слова. На відміну від книжкового словника, в електронному є можливий варіант лексикографічного моделювання, при якому поява на екрані тієї чи іншої зони словникової статті виконується за вибором користувача.

До другої категорії відносяться програми, які дозволяють виконувати автоматичний переклад тексту. Програми перекладу (системи машинного перекладу) з'явилися у відповідь на потреби користувачів в оперативному перекладі різної комерційної, технічної або INTERNET-інформації, яка подана в електронному вигляді. Крім перекладу з іноземних мов, важливе значення має переклад з української (російської) мови іншими мовами, зокрема англійською.

При перекладі програми використовують словники, набори граматичних правил та засоби, які забезпечують найкращу якість перекладу. Словник може складатися не тільки з виділеного слова, але й з словосполучення. Використовуючи ці засоби, програми аналізують граматичну структуру тексту, знаходять зв'язок між словами і намагаються побудувати правильний переклад фрази на іншій мові. Найбільш широке розповсюдження отримали програми: Systran, Promt, Pragma, Language Master. Для перекладу невеликих фрагментів тексту доцільно використовувати програму QTrans.

Розглянемо критерії, за якими порівнювались вказані вище програмні продукти:

1. Інсталяція/Деінсталяція характеризує можливість установки та знищення програм машинного перекладу з комп'ютера користувача.
2. Швидкість перекладу визначає, наскільки швидко можуть бути перекладені текстові документи великих розмірів.
3. Кількість неперекладених слів характеризує словники, які супроводжують програми машинного перекладу.
4. Якість перекладу визначає граматичну коректність перекладу.
5. Зручність настройки програми та словників визначає дружелюбність інтерфейсу користувача.
6. Використання у мережі засвідчує можливість застосування програми-перекладача для колективної роботи.
7. Документація дозволяє отримати первісні поняття та теоретичні знання про програму.
8. Робота з Web-браузером дає можливість виконувати оперативний переклад Web-сторінок.

Використання програм-перекладачів може бути організована двома способами:

1. Автоматичний переклад усього тексту з обов'язковим наступним його вичитуванням і коригуванням вручну.
2. Переклад текстів в інтерактивному режимі з вибором варіантів перекладу з альтернатив, запро-

понованих програмою, або уведенням власного варіанту перекладу.

В будь-якому випадку завершений текст слід добре вичитати з метою виправлення стилістичних помилок, оскільки ні одна програма стилістики перекладеного тексту не контролює.

Більшість користувачів розглядають Web – сторінки Інтернету, як документи, які потребують швидкого перекладу без особливих вимог до якості. Більшість Web – сторінок в світі написана на англійській мові, і програма Promt має спеціальне положення для швидкого перекладу Web – сторінок – програму Web View. Web View представляє собою браузер, який проводить синхронний переклад Web – сторінок під час роботи в Інтернеті. Важливою особливістю програми являється також автоматичний переклад головних слів, використаних при пошуку потрібної інформації в мережі. Головні слова запиту автоматично перекладаються в заданому напрямленні перекладу. Сформований запит відображається на екрані. Цей запит можна направити в будь-яку із декількох доступних програм пошукових систем.

Переклад документів звичайно являє собою проміжну операцію, тому система пропонує багато різних варіантів збереження документа. Вибір варіанта збереження залежить від використання документа в майбутньому.

За допомогою електронної пошти будь-який адресат може отримати передану та оброблену інформацію.

Висновок

Комп'ютерний переклад текстів - одна з найцікавіших і потрібніших інформаційних технологій, основним призначенням якої є така обробка даних, після якої, користувач отримує у своє розпорядження інформацію, яку можна використати для прийняття різноманітних рішень на всіх рівнях діяльності. Тим більше, вона потрібна в офісах, де доводиться мати справу з великою кількістю документів іноземними мовами. Тому основними питанням є створенням і впровадження на базі лінгвістичного забезпечення інформаційних систем різних типів.

References

- [1] Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. Изд. 2-е, исправленное. – М.: Едиториал УРСС, 2003. – 360 с.
- [2] Зубов А.В. Информационные технологии в лингвистике: Учеб. пособие для студ. лингв. фак-тов высш. учеб. заведений/ А.В.Зубов, И.И.Зубова – М.: Издательский центр “Академия”, 2004. – 208 с.
- [3] Карпіловська Є.А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика: Підручник. – Донецьк: ТОВ “ Юго-Восток, Лтд”, 2006. – 188 с.