

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
"ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"



Матеріали

IV Всеукраїнської науково-практичної конференції

**"ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ТА
ПРИКЛАДНА ЛІНГВІСТИКА"**

ХАРКІВ 2015

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
"ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"

Матеріали

IV Всеукраїнської науково-практичної конференції

**"ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ТА
ПРИКЛАДНА ЛІНГВІСТИКА"**

ХАРКІВ 2015



УДК 004.9:81

Матеріали IV Всеукраїнської науково-практичної конференції "Інтелектуальні системи та прикладна лінгвістика".

Харків, 16 квітня 2015 р. : матеріали конференції. – Харків: Національний технічний університет "Харківський політехнічний інститут", 2015. – 116 с.

В матеріалах розглядаються проблеми та перспективи розвитку інтелектуальних комп'ютерних систем та різних галузей прикладної лінгвістики, а саме корпусної лінгвістики, комп'ютерної лексикографії, машинного перекладу, лінгвістики Інтернету; питання використання інформаційних технологій в лінгвістиці, з метою дослідження та обробки мови.

Редакційна колегія:

д.т.н. **Гамаюн І.П.** – декан факультету інформатики і управління НТУ "ХПІ";

д.т.н. **Шаронова Н.В.** – завідувач кафедри інтелектуальних комп'ютерних систем НТУ "ХПІ";

Петрасова С.В. – аспірант кафедри інтелектуальних комп'ютерних систем НТУ "ХПІ".

© Національний технічний університет "Харківський політехнічний інститут",
2015



ПРИВІТАННЯ КОНФЕРЕНЦІЇ

Вітаю шановних спікерів та гостей на відкритті IV Всеукраїнської науково-практичної конференції "Інтелектуальні системи і прикладна лінгвістика"!

Від імені компанії NIX Solutions хочу подякувати викладацькому складу кафедри «Інтелектуальні комп'ютерні системи» і завідуючій кафедрою Наталії Валеріївні Шароновій особисто за Вашу величезну щоденну працю з виховання наступних поколінь відмінних фахівців. Ви даєте своїм студентам різнобічну підготовку високого рівня. Завдяки цьому багато випускників вашої кафедри стали незамінними співробітниками нашої компанії.

Так наприклад, ще вчорашня студентка Цибізова Юлія після навчального проекту нашої компанії працює QA-інженером.

Керівництво NIX Solutions завжди надавало велике значення тісній співпраці з харківськими вузами, особливо виділяючи при цьому НТУ «ХПІ» та кафедру «Інтелектуальні комп'ютерні системи». Розуміння основних потреб сфери, бажання надати підтримку молодим ІТ-фахівцям Харкова, взаємодія з провідними вузами нашого міста і підвищувати якість ІТ-освіти в Україні – все це більше двох років тому привело компанію NIX Solutions вперше на вашу кафедру. Я можу сміливо сказати, що за цей час ми змогли налагодити ефективну співпрацю. І, сподіваюся, будемо її продовжувати в рамках коригування навчальних планів і програм.

Всім студентам важливо розуміти, що їх чекає за стінами alma mater. Практичний досвід особливо важливий для представників ІТ-професій. Наша співпраця з вашою кафедрою допомагає хлопцям і дівчатам вчасно побачити свої перспективи і найбільш ефективно використовувати знання, отримані під час навчання.

І пам'ятайте, за Вами майбутнє ІТ-сфери. А це означає, що і майбутнє країни.

Моргун Вікторія
Координатор навчального
центру
NIX Solutions Ltd.



АЛФАВІТНИЙ ПОКАЖЧИК АВТОРІВ

А		М	
Аджит Пратап Сингх Гаутам	9	Медведська А.В.	76
Б		Мусієнко О.П.	90
Бабкова Н.В.	71	Н	
Борзенкова А.В.	68	Ніконов С.І.	42
Борисова Н.В.	52	О	
Булатнікова Т.С.	63	Оліфенко І.В.	110
Бутенко Ю.І.	28	П	
В		Пасенко К.В.	31
Вальчук Х.І.	37	Петрасова С.В.	21
Величко А.М.	40	Печенікова Л.М.	18
Верховець М.В.	108	Пилипенко А.А.	80
Г		Піскова Н.О.	96
Гайдено Т.В.	102	Прокопчук Ю.О.	54
Главчева Ю.М.	14	Р	
Гончар І.О.	50	Рижкова В.В.	40
Гончарова С.О.	88	С	
Д		Семененко Л.П.	14
Дашкевич О.С.	65	Скуменко С.І.	85
Детинич К.Г.	115	Стребкова О.О.	92
Дорошенко А.Ю.	35	Т	
Ж		Терещенко В.І.	26
Жукова К.А.	59	Тулюлюк К.В.	45
З		Тупікова Н.С.	83
Золотько Д.В.	112	Ф	
І		Федосєєв П.І.	94
Іщенко О.С.	23	Х	
К		Хайрова Н.Ф.	11
Каніщева О.В.	16	Ч	
Кліменкова К.Г.	104	Чалюк Г.Є.	78
Козуля М.М.	57	Чухненко М.В.	106
Колесник А.С.	99	Ш	
Коновалова В.Б.	61	Шаронова Н.В.	9
Кудоярова О.В.	59	Шостак І.В.	28
Кузьміна М.О.	71	Шостак О.І.	50
Купріянов Є.В.	48	Ю	
Л		Юрьєва К.О.	74
Лисенко О.І.	50		
Лутай Н.В.	61		



ЗМІСТ

Аджит Прагап Сингх Гаутам, Шаронова Н.В. ОСНОВНЫЕ ПРОБЛЕМЫ ОБРАБОТКИ ТЕКСТОВ В ИНТЕГРИРОВАННЫХ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ.....	9
Хайрова Н.Ф. СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ.....	11
Семененко Л. П., Главчева Ю. М. МОБІЛЬНІСТЬ У ВИЩІЙ ШКОЛІ: ПОГЛЯД БІБЛІОТЕКАРЯ.....	14
Канищева О.В. ФOLKSONOMIA КАК ИСТОЧНИК ДЛЯ ПОСТРОЕНИЯ ОНТОЛОГИЙ И АССОЦИАТИВНЫХ СЛОВАРЕЙ. ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ.....	16
Печенікова Л.М. ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ У ЛІНГВІСТИЦІ.....	18
Петрасова С.В. E-COMMUNITIES КАК ИСТОЧНИК ДАННЫХ В НЕФОРМАЛЬНОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ.....	21
Іщенко О.С. ПРО ОДИН ПРОГРАМНИЙ ЗАСІБ КОМП'ЮТЕРНОГО АНАЛІЗУ ФОНОТЕКСТІВ.....	23
Терещенко В.І. ДОСВІД ЗАСТОСУВАННЯ ЗНАНЬ З ПРИКЛАДНОЇ ЛІНГВІСТИКИ У СТВОРЕННІ СИСТЕМИ, ЩО ВИКОРИСТОВУЄ МЕТОДИ OPINION MINING ТА COREFERENCE RESOLUTION.....	26
Бутенко Ю.І., Шостак І.В. ДОСЛІДЖЕННЯ ВЛАСТИВОСТЕЙ МОВИ СТАНДАРТІВ ЯК ЕКЗЕМПЛЯРУ КЛАСУ МОВ ДЛЯ СПЕЦІАЛЬНИХ ЦІЛЕЙ В КОНТЕКСТІ АВТОМАТИЗАЦІЇ ПРОЦЕДУРИ СЕРТИФІКАЦІЇ.....	28
Rasenko K.V. CORPUS-BASED STUDIES IN LINGUISTICS.....	31
Дорошенко А.Ю. ТЕХНОЛОГИИ ИЗВЛЕЧЕНИЯ ЗНАНИЙ НА ОСНОВЕ ФАКТОГРАФИЧЕСКИХ ДАННЫХ ЭВМ В СТРУКТУРИРОВАННЫХ КОНТЕНТАХ.....	35
Valchuk K.I. PROBLEMS OF MACHINE TRANSLATION.....	37
Величко А.М., Рижкова В.В. ЗАСТОСУВАННЯ МУЛЬТИМЕДІА В НАВЧАННІ ФАХОВІЙ АНГЛІЙСЬКІЙ МОВІ (НА ПРИКЛАДІ ДИСЦИПЛІНИ «ІНЖЕНЕРНІ ОСНОВИ АЕРОКОСМІЧНОЇ ТЕХНІКИ»).....	40



Nikonorov S.I. TECHNICAL COMMUNICATION SKILLS IN THE CONTEXT OF TRAINING SPECIALISTS FOR SOFTWARE OUTSOURCING FIELD.....	42
Тулюлюк К.В. ПРОЯВ ГЕНДЕРА В МОВЛЕННЄВІЙ ДІЯЛЬНОСТІ.....	45
Купріянов Є.В. АНТРОПОЦЕНТРИЧНИЙ ПІДХІД ДО УКЛАДАННЯ ЕЛЕКТРОННОГО ПЕРЕКЛАДНОГО СЛОВНИКА ТЕХНІЧНОЇ ЛЕКСИКИ.....	48
Гончар І.А., Лысенко А.И., Шостак Е.И. БИМАТРИЧНАЯ ИГРА КАК ОСНОВА ФУНКЦИОНИРОВАНИЯ ПРОИЗВОДСТВЕННОЙ МУЛЬТИАГЕНТНОЙ СИСТЕМЫ.....	50
Борисова Н.В. ОЦІНКА ЕФЕКТИВНОСТІ ВИДОБУВАННЯ ТЕРМІНІВ ПРЕДМЕТНОЇ ОБЛАСТІ З ТЕКСТІВ.....	52
Прокопчук Ю.А. ВЫСОКОУРОВНЕВЫЙ ЯЗЫК ЗАПРОСОВ К БАЗЕ ДАННЫХ НА ОСНОВЕ «КОННЕКТОМА» КОГНИТИВНОЙ СИСТЕМЫ.....	54
Козуля М.М. ИНТЕЛЛЕКТУАЛЬНАЯ ОБРАБОТКА ДАННЫХ О СОСТОЯНИИ ПРИРОДНО- ТЕХНОГЕННЫХ ОБЪЕКТОВ.....	57
Кудоярова О.В., Жукова К.А. СУЧАСНІ ТЕНДЕНЦІЇ СТВОРЕННЯ ТЕРМІНОЛОГІЧНИХ ПЕРЕКЛАДНИХ СЛОВНИКІВ (НА ПРИКЛАДІ СТВОРЕННЯ ЕЛЕКТРОННОГО СЛОВНИКУ У ГАЛУЗІ GPS SOFT У ТОПОГРАФІЇ ТА КАРТОГРАФІЇ).....	59
Lutai N.V., Konovalova V.B. LEARNING GRAMMAR ACTIVITIES TO STUDENTS OF LINGUISTIC DEPARTMENTS.....	61
Булатнікова Т.С. КЛАСИФІКАЦІЯ ТЕКСТІВ ЕКОНОМІЧНОГО НАПРЯМКУ НА ОСНОВІ СЕМАНТИЧНИХ МЕРЕЖ ГІПОНІМІЧНИХ ВІДНОШЕНЬ.....	63
Дашкевич О.С. АЛГОРИТМ СИСТЕМИ АВТОМАТИЗОВАНОГО РЕФЕРУВАННЯ КОРПУСУ ТЕКСТІВ.....	65
Борзенкова А.В. ЗАДАЧА КЛАСИФІКАЦІЇ ТЕКСТІВ РОСІЙСЬКОЮ МОВОЮ ЗА ГЕНДЕРНИМИ ОЗНАКАМИ.....	68
Кузьмина М.А., Бабкова Н.В. О РЕФЕРЕНЦИИ И КОРЕФЕРЕНЦИИ В ЛИНГВИСТИКЕ.....	71
Юрьева К. О. АНАЛІЗ ІСНУЮЧИХ СИНОНІМАЙЗЕРІВ ТА ОГЛЯД ПРОБЛЕМ ЩОДО ЇХ СТВОРЕННЯ.....	74



Медведська А. В. ВИРІШЕННЯ ЗАВДАННЯ SENTIMENT ANALYSIS ЗА ДОПОМОГОЮ СЛОВНИКОВОГО ПІДХОДУ (НА ПРИКЛАДІ МАТЕРІАЛУ ФОРУМІВ СМАРТФОНІВ).....	76
Чалюк Г. Є. АВТОМАТИЧНЕ ВИЛУЧЕННЯ ДІЕСЛІВНИХ КОЛОКАЦІЙ АНГЛІЙСЬКОЇ МОВИ.....	78
Пилипенко А.А. СОЗДАНИЕ ЭЛЕКТРОННОГО СЛОВАРЯ АНТОНИМОВ.....	80
Тупікова Н.С. РОЗРОБКА СЛОВНИКА СИНОНІМІВ ДЛЯ СИНОНІМАЙЗЕРА НОВИННИХ ПОВІДОМЛЕНЬ.....	83
Скуменко С.І. РОЗРОБКА ЕЛЕКТРОННОГО ТЛУМАЧНОГО СЛОВНИКА ТЕРМІНІВ З КОМП'ЮТЕРНОЇ ЛІНГІСТИКИ НА ОСНОВІ КОРПУСУ ТЕКСТІВ.....	85
Гончарова С. О. ЛІНГВІСТИЧНИЙ АСПЕКТ УКЛАДАННЯ ЕЛЕКТРОННОГО ПАРАЛЕЛЬНОГО СЛОВНИКА ФРАЗЕОЛОГІЗМІВ.....	88
Мусієнко О.П. MOODLE – ВІРТУАЛЬНЕ НАВЧАЛЬНЕ СЕРЕДОВИЩЕ ДЛЯ КЕРУВАННЯ ДИСТАНЦІЙНИМ НАВЧАННЯМ.....	90
Стребкова О.О. РОЗРОБКА СПЕЦІАЛІЗОВАНОГО ПЕРЕКЛАДНОГО УКРАЇНСЬКО- АНГЛІЙСЬКОГО СЛОВНИКА З ТЕКСТИЛЬНИХ МАТЕРІАЛІВ.....	92
Федосєєв П.І. РОЗРОБКА СИСТЕМИ ПЕРЕВІРКИ І ОЦІНЮВАННЯ РІВНЯ ЗНАНЬ СТУДЕНТІВ СПЕЦІАЛЬНОСТІ "ПРИКЛАДНА ЛІНГВІСТИКА".....	94
Піскова Н.О. ДИНАМІЧНЕ ФОРМУВАННЯ СТОРІНКИ WEB-САЙТУ НА ОСНОВІ ВИКОРИСТАННЯ ОНТОЛОГІЇ ПРЕДМЕТНОЇ ОБЛАСТІ.....	96
Колесник А. С. ПРИНЦИПЫ РАБОТЫ СИСТЕМ МАШИННОГО ПЕРЕВОДА.....	99
Гайденко Т.В. ЭЛЕКТРОННАЯ И ТРАДИЦИОННАЯ ЛЕКСИКОГРАФИЯ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ.....	102
Клименкова Е.Г. ПОСТРОЕНИЕ ЧАСТОТНОГО СЛОВАРЯ СЛОВОФОРМ НЕСКОЛЬКИХ ТЕКСТОВ.....	104
Чухненко М.В. О МЕТОДАХ СОКРЫТИЯ ИНФОРМАЦИИ.....	106



Верховець М.В. ДО ПРОБЛЕМИ ІНШОМОВНИХ ЗАПОЗИЧЕНЬ ТЕРМІНІВ.....	108
Олифенко І.В. ЛОЖНЫЕ ДРУЗЬЯ ПЕРЕВОДЧИКА.....	110
Золотько Д.В. ГЕНДЕРНИЙ АСПЕКТ СЕМАНТИКО-СТИЛІСТИЧНИХ ОСОБЛИВОСТЕЙ РЕКЛАМНОГО ІНТЕРНЕТ-ДИСКУРСУ.....	112
Детинич К.Г. ЛЕКСИЧНІ ТРУДНОЩІ ПРИ ПЕРЕКЛАДІ (НА МАТЕРІАЛІ ПОВІСТІ Г.Ф. ЛАВКРАФТА «ПОШУК НЕЗВІДАНОГО КАДАТУ»).....	115



ОСНОВНЫЕ ПРОБЛЕМЫ ОБРАБОТКИ ТЕКСТОВ В ИНТЕГРИРОВАННЫХ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Аджит Пратап Сингх Гаутам, Шаронова Н.В.
*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,
e-mail: nvsharonova@mail.ru*

Перед автоматизированными системами извлечения знаний из текста сегодня встают насущные практические задачи, появление которых стимулировано развитием Интернета, содержащего огромное количество текстовой информации – реальные элементы утилитарного знания, полученные людьми в результате их деятельности. К таким задачам, как показывает анализ литературных источников, относятся: поиск и извлечение элементов знания, явно присутствующих в текстовой коллекции в виде: а) утверждения; б) факта; порождение сложного знания путем обработки элементов знания следующими способами: а) генерация нового знания как цепочки логического вывода из элементарных утверждений и/или фактов; б) эксплицирование обобщенного знания, скрытого в совокупности частных утверждений и/или фактов [1, 2].

Большой объем информации, циркулирующей в интегрированных корпоративных информационных системах, а также ее низкая структурированность обуславливают повышение роли такой процедуры, как извлечение информации из текста. В Украине также ведутся разработки по интеграции данных, в том числе и семантической интеграции, и интеграции, основанной на онтологии.

Современные процедуры извлечения информации, использующие методы обработки текстов на естественном языке, как правило, направлены лишь на решение узкого класса задач (отбор ограниченного набора тем (вопросов, проблем), а зачастую и только на одну тему). Это привело к тому, что разработанные эффективные процедуры обработки текста невозможно применить в качестве универсального метода обработки текста.

Традиционно в системах анализа текстов для представления знаний используется четыре типа моделей: продукционная, формально-логическая, фреймовая и семантико-сетевая модели. На базе этих моделей описываются решения и основные перспективы их использования [2]. В литературе представлен систематизированный анализ методов, позволяющих обнаруживать и извлекать из текста конструктивные элементы. Исследования таких ученых, как Ю.П. Шабанов-Кушнарченко, М. Ф. Бондаренко, Н. Ф. Хайрова и др. доказывают, что именно использование лингвистических методов позволяет существенно улучшить качество автоматического анализа текста. Следует отметить перспективность использования алгебрологического аппарата для решения задач анализа текста. При этом ядром таких систем могут стать интегрированные интеллектуальные системы, включающие элементы искусственного интеллекта,



основанные на методах и средствах теории интеллекта, развиваемых на кафедре интеллектуальных компьютерных систем НТУ «ХПИ» [2, 4]. Существующее многообразие частных задач обработки текстовой информации позволяет сгруппировать их в следующие крупные классы, связанные с анализом текстовой информации [5]:

1. Распознавание именованных элементов (сущностей), например, имён людей, названий организаций, географических названий и пр.
2. Разрешение анафоры и кореференций: поиск связей, относящихся к одному и тому же объекту.
3. Выделение терминологии: нахождение для данного текста ключевых слов и словосочетаний (коллокаций).
4. Автореферирование: выделение из текста смысловой, эмотивной, оценочной и пр. информации.
5. Корпусная лингвистика: создание и использование корпусов текстов.
6. Создание электронных словарей, тезаурусов, онтологий.
7. Автоматический перевод текстов.
8. Автоматическое извлечение фактов из текста.
9. Построение систем управления знаниями.
10. Создание вопросно-ответных систем.
11. Информационный поиск.

Информационные системы представляют широкий класс программного обеспечения, используемого различными предприятиями для автоматизации их работы. Поскольку объем обрабатываемой информации огромен, уже в каждой организации существует, как правило, несколько информационных систем. Часто в этих системах обрабатывается одна и та же информация. В связи с этим возникает проблема интеграции данных из различных систем. Под интеграцией данных понимается процесс объединения данных из различных источников для получения их согласованного представления, а в широком смысле – процесс организации регулярного обмена данными между различными информационными системами предприятия [5].

Список литературы:

1. Ермаков А.Е. Автоматизация онтологического инжиниринга в системах извлечения знаний из текста / А.Е.Ермаков // Труды Международной конференции Диалог 2008. – С.137.
2. Лингвотехнологии идентификации знаний в информационных системах : монография / О. В. Канищева, Н. В. Шаронова. – Saarbrücken, Deutschland : LAP LAMBERT Academic Publishing, 2013. – 173 с. – На рус. яз.
3. Бондаренко М. Ф. Мозгоподобные структуры: Справочное пособие. / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнарченко. Том первый. Под редакцией акад. НАН Украины И.В. Сергиенко. – К.: Наукова думка, 2011. – 460 с.
4. Хайрова Н.Ф., Шаронова Н.В. Лингвистические технологии экстракции и идентификации знаний // Тези доповідей Міжнародної науково-технічної конференції "Інтелектуальні технології лінгвістичного аналізу" (м. Київ, 22-23 жовтня 2013 р.). – К.: НАУ, 2013. – С. 7.
5. Павленко М.А. Анализ методов решения задачи извлечения информации из текстов / М.А.Павленко // Системи обробки інформації. – 2013. – Т.1.– С.29.



СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

Хайрова Н.Ф.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–64–60,
e-mail:nina_khajrova@yahoo.com*

В пятидесятые годы прошлого столетия на стыке компьютерных наук, искусственного интеллекта и лингвистики выделилось новое научное направление Natural language processing (NLP), изучающее взаимодействие между компьютером и естественным языком. Но активное развитие компьютерной лингвистики началось после 1954 года, когда ученые университета Georgetown и компании IBM впервые показали систему машинного перевода, которая успешно осуществила автоматический перевод нескольких предложений с русского языка на английский.

Такой успех Джоржтаунского эксперимента (Georgetown experiment) позволил ученым утверждать, что в ближайшие несколько лет произойдет полный переход от интеллектуального к машинному переводу. И хотя до настоящего времени это не произошло, и вряд ли произойдет в ближайшем будущем, сегодняшние успехи прикладной и компьютерной лингвистики в области разработки систем автоматической обработки текстов на естественном языке очевидны. Среди задач, связанных с использованием NLP следует выделить:

- классификацию и кластеризацию текстов;
- информационный поиск текстовых документов;
- тематическое индексирование и рубрицирование;
- извлечение фактов и понятий;
- автоматическое реферирование и квазиреферирование и др.

Существующий уровень реализаций приложений NLP можно разделить на три больших категории: направления, по которым есть коммерческие продукты и системы; направления, в реализации которых задействованы работающие алгоритмы; и практически не решенные на сегодня задачи лингвистического процессора.

Так, например, существует достаточное количество коммерческих приложений по определению спама в получаемом пользователем потоке e-mail сообщений. И, хотя эти приложения продолжают принимать достаточное количество ошибочных решений, они включены в современные почтовые приложения. Аналогичным образом, достаточно коммерциализированы системы синтеза речи (speech synthesis), имеющие успешно разработанные приложения, подобно спецификации SALT (Speech Application Language Tags) [1].

В то же время, сегодня существует большая группа задач обработки текстов, в которых достигнут достаточный прогресс, но полностью они пока не решены. К таким задачам относятся системы извлечения информации



(Information extraction), включаючи задачі Opinion Mining. Такі системи використовуються, в том числі при маркетинговому дослідженні, для извлечения позитивної или негативної інформації о том или іншому продукті или сервисі в Веб. Методи Opinion Mining, позволяющие класифікувати тексти по тональності, появились сравнительно недавно, и во многом используют подходы Text Mining и Information Retrieval [2].

К этой же группе задач относится уже упомянутое приложение лингвистического процессора — машинный перевод, подразумевающий полностью автоматический переводчик. В автоматически получаемых переводах, несмотря на достигнутый в последнее время прогресс в развитии технологии парсинга и статистических параллельных переводных баз данных, до сих пор встречается достаточное количество ошибок.

Последнее, третье направление задач лингвистического процессора является на сегодня практически не решенным. Так, вопросно-ответные системы, которые используются для автоматического ответа на вопросы любого вида, имеют реальные алгоритмы только для формулировки самых простых фактографических вопросов. Простые, но общие вопросы, по-прежнему остаются тяжелой проблемой.

Аналогичным образом вопросно-ответные системы, используемые в интеллектуальном интерфейсе современных автоматизированных информационных систем, существенно отстают от уровня практической востребованности [3] и относятся к третьей группе практически не решенных задач лингвистического процессора.

Еще одним, практически не решенным направлением лингвистического процессора, является перефразирование (paraphrases). Система, использующая перефразирование, должна обладать мощными средствами семантического смыслового анализа, практически не существующими на сегодняшний день. Еще одна задача, которая к сложностям перефразирования добавляет проблемы смыслового обобщения — задача автоматического реферирования (summarization). В настоящее время по-прежнему на рынке присутствуют только системы квазиреферирования, использующие по большей части статистико-позиционные подходы, а алгоритмов «истинного» смыслового обобщения практически нет [4].

Наибольшую сложность и наименьшее количество решений представляют диалоговые системы, которые могут отвечать на вопросы и интерпретировать имеющуюся ситуацию. Это системы, для работы которых существуют только первичные варианты алгоритмов [5].

Проведенный анализ показывает, что, несмотря на то, что работы в направлении автоматизации обработки естественного языка продолжаются более 50 лет и интенсифицировались в последние годы, когда накоплены огромные полнотекстовые информационные массивы, и лингвистические технологии из средств, занимающихся разработкой формальных моделей языка, переходят в фактор производства, в настоящий момент можно выделить четко очерченный класс приложений лингвистического процессора, которые не решены или решены лишь частично (для узких предметных областей).



Невозможность быстрого решения проблемы автоматической обработки текстов на естественном языке обусловлена двумя основными (фундаментальными) причинами:

- задача разработки лингвистического процессора относится к сложно формализуемым задачам, связанным с неопределенностью;
- практически все существующие проблемы обработки текстов на сегодня связаны с проблемой смыслового анализа и необходимостью формализации понимания смысла текстовой информации.

Для решения задач NLP кроме лингвистических знаний необходимо обладать знаниями о мире и уметь формализовать объединение этих знаний. С этим связана необходимость подхода к задачам NLP с точки зрения системного анализа, рассматривая язык как сложно формализуемую иерархическую систему, в которой для формализации семантики нужно моделировать функции человеческого интеллекта по пониманию и идентификации знаний в естественно-языковых текстах.

Список литературы

1. *Daniel Jurafsky, James H. Martin.* Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Second Edition. – Prentice Hall, 2008. – 988 p.
2. *Nozomi Kobayashi, Kentaro Inui and Matsumoto Yuji.* Opinion Mining from Web Documents: Extraction and Structurization. Journal of Japanese society for artificial intelligence, special issue on data mining and statistical science, 2007. – Vol. 22. – № 2. – P. 227–238.
3. *Юсупов Р. М.* Научно-методологические основы информатизации / Р. М. Юсупов, В. П. Заболотский. – СПб.: Наука, 2000. – 455 с
4. *Ландэ Д.* Глубинный анализ текстов технология эффективного анализа текстовых данных [Электронный ресурс] //СНІР Ukraine. – №10. – 2003. – Режим доступа: <http://www.visti.net/~dwl/art/dz/>.
5. *Кокорева Л.В.* Диалоговые системы / Л. В. Кокорева, О. Л. Перевозчикова, Е. Л. Ющенко. – Ин-т кибернетики АНУ. – К.: Наук. думка, 1993.– 448с



МОБІЛЬНІСТЬ У ВИЩІЙ ШКОЛІ: ПОГЛЯД БІБЛІОТЕКАРЯ

Семененко Л.П., Главчева Ю.М.

*Національний технічний університет «Харківський політехнічний інститут»,
м. Харків, вул. Червонопрапорна, 16. Тел.: 707-63-61,
e-mail: semenenkoL12@gmail.com, glavcheva@khpі.edu.ua*

Принцип мобільності у вищій школі передбачає не лише можливість навчання за кордоном, вивчення досвіду викладання та змісту освіти у вищих навчальних закладах, знання мов, а й комплекс вмінь та навичок користування віддаленими ресурсами у будь-який час у будь-якому місці, мобільними пристроями та сервісами, які вони надають. Технічна база пристроїв та сервісів розвивається швидко, а сервіси, які вони надають, динамічно змінюються і удосконалюються. Оперативно відстежувати появу науково цінних інформаційних ресурсів, ринок технічних засобів та пристроїв, визначати доцільність застосування того чи іншого сервісу, можуть бібліотекари, оскільки вони мають фахову підготовку у формуванні та використанні інформаційних ресурсів, сервісів, володіють навичками аналітико-синтетичної обробки документів та лінгвістичного супроводу генерованих ресурсів. Для швидкого інформування своїх користувачів про нові сервіси та навчання їх застосування бібліотека, як ніхто інший має необхідні ресурси: кадри, програмно-технічне середовище та матеріально-технічну базу. Застосування різних форм групового навчання, зручний час, індивідуальні графіки навчання для окремих користувачів роблять такі заняття популярними та ефективними. Лише семінари бібліотеки щорічно відвідують понад 600 осіб.

Бібліотекари науково-технічної бібліотеки Національного технічного університету «Харківський політехнічний інститут» у своїй роботі з обслуговування користувачів керуються принципами: повноти, оперативності, доступності, диференційованого підходу у обслуговуванні читачів; прагнуть максимально повно задовольнити читацькі запити.

Наша робота спрямована на удосконалення існуючих та пошук нових форм і методів роботи, що стало можливим лише завдяки постійній підтримці інноваційного клімату в колективі. Система управління спрямовується на перебудову професійного мислення, розвиток творчої ініціативи співробітників. Підтримуючи розвиток університетських свобод, розвиток методів навчання, заснованих на інтересах студентів, ми активно приєдналися до реалізації гасла «від навчання — до освіти; від навчання, заснованого на програмі, — до освіти на базі спільного рішення проблем». Намагаємося виявляти нові ідеї, підтримувати систему електронної освіти на основі концепції відкритості.

Пріоритетними напрямками роботи бібліотеки є: формування комплексу власних електронних наукових та навчальних ресурсів, підвищення рівня та оперативності бібліотечно-бібліографічного та інформаційного обслуговування всіх категорій читачів, забезпечення доступу користувачам бібліотеки, незале-



жно від їх місця роботи, навчання та проживання, як до традиційних документів з фонду, так і до інформації, що існує виключно в електронній формі.

Максимально реалізувати увесь комплекс завдань бібліотеки можливо лише за умови розширення співпраці з іншими бібліотеками та науковими установами, оскільки це дозволяє залучити додаткові інформаційні та документні ресурси на умовах взаємовигідного партнерства, тому бібліотека ініціює участь у роботі міжнародних організацій та їх проектів.

Завдяки високому рівню якості власних інформаційних продуктів та сервісів бібліотека стала членом п'яти міжнародних організацій, залучена до реалізації одинадцяти корпоративних проектів. Доступ лише до баз даних 9 міжнародних та регіональних корпоративних проектів забезпечує користувачам бібліотеки оперативний пошук релевантної інформації, формування списків джерел із власних ресурсів та ресурсів партнерів по корпоративній взаємодії загальним обсягом понад 14 млн. записів; електронну доставку документів з ресурсів близько 700 бібліотек та наукових установ України та зарубіжжя. Щорічно до світових авторитетних наукових ресурсів забезпечується понад 20 сесій тимчасового доступу.

Для забезпечення комфортного доступу користувачам бібліотеки до документних та електронних фондів упроваджуються та розвиваються сервіси бібліотеки:

- попереднє замовлення та бронювання видань;
- перегляд та друк інформації з електронного формуляру;
- електронна книговидача;
- електронна доставка документів;
- хмарні сервіси від Microsoft (Office 365);
- дистанційний курс «Куратор змісту» (разом з проблемною лабораторією дистанційного навчання НТУ «ХП»);
- модуль «Книгозабезпеченність» автоматизованої системи ІРБІС-64 та інші.

Додаткові зручності отримали користувачі бібліотеки завдяки упровадженню єдиної картки читача бібліотек вищих навчальних закладів м. Харкова (надає право безкоштовного користування ресурсами 24 бібліотек учасниць проекту), профілям бібліотеки у соціальних мережах: Facebook, Twitter, Pinterest, ВКонтакте, інтеграції інформації про 900 науковців НТУ «ХП» до реєстру вчених України Системи наукометричного моніторингу суб'єктів наукової діяльності, інтеграції ресурсів університету до світових авторитетних наукових баз даних та спеціалізованих каталогів.

Динамічно розвиваються ресурси бібліотеки: електронний каталог — близько 500 тис. записів; інституційний репозитарій — понад 13 тис. документів; повнотекстові бази даних — понад 8 тис. документів (доступні зареєстрованим користувачам за авторизацією відповідно до законодавства про авторське право). Web-сайт <http://library.kpi.kharkov.ua> щорічно відвідують понад 115 тис. зовнішніх користувачів, завантажують близько 350 тис. документів.



ФОЛКСОНОМИЯ КАК ИСТОЧНИК ДЛЯ ПОСТРОЕНИЯ ОНТОЛОГИЙ И АССОЦИАТИВНЫХ СЛОВАРЕЙ. ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

Канищева О.В.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,
e-mail: olya-kanisheva@rambler.ru*

С развитием концепции семантического веба – web 2.0 появились и активно развиваются такие социальные ресурсы как Flickr [1], Instagram [2], Picasaweb [3], Photobucket [4] и другие. В основе таких систем лежит методика проектирования, в результате которой с помощью сетевых взаимодействий, системы становятся тем лучше, чем больше людей ими пользуются. Таким образом пользователи (не профессионалы) добавляют свои собственные ключевые слова (признаки) к информационным объектам (фото, видео и др.). Информации (тегов) становится все больше, однако вопросы о её надёжности, достоверности и объективности подвергаются сомнениям.

В результате создания таких ресурсов и их разметки происходит так называемая *фолксономия* (англ. folksonomy, от folk – народный + taxonomy таксономия, от гр. расположение по порядку + закон) – народная классификация, практика совместной категоризации информации (текстов, ссылок, фото, видео клипов и т. п.) посредством произвольно выбираемых меток, называемых тегами [5].

Системы, реализующие маркировку, состоят из трех главных элементов: пользователи, ресурсы и признаки (теги). Хотя в большинстве систем теги не являются обязательными, но они являются очень важными, благодаря этим меткам у пользователей есть возможность осуществлять поиск, сортировать и объединять в группы сходные объекты.

Однако никакими правилами или словарями при разметке ресурсов пользователи не пользуются. Это влечет за собой орфографические ошибки, мультязычное написание слов (*muenchen, munich*), использование различных символов при написании (*ny, new_york, "new york", newyork*), синонимы, флективные формы слов (множественное число и др.), эмоциональную лексику (*pretty, nice*), местоимения и др. Весь этот шум мешает качественно проводить кластеризацию информационных объектов и обработку ключевых слов.

Примеры тегов для произвольного изображения 1) *niagara, falls, buffalo, new, york, ny, city, nature, water, steam, haze, park, state, erosion, historic, tourist, boat*; 2) *panda, giant, wildlife, nature, bear, panda bear, national zoo, The Perfect Photographer, AnimalKindomElite, SpecAnimal*.

Особенностью фолксономии является то, что правил для маркировки ресурсов для пользователей как таковых нет. Пользователи могут как присвоить ресурсу одно ключевое слово, так и 10 слов. Если ресурсы похожи и относятся к одному пользователю, то он копирует теги и меняет несколько слов, а в неко-



торых случаях так их и оставляет. В работе [2] авторы показали, что при анализе сервиса Delicious за 2004 год самыми популярными тегами были: *software, design, programming, music, politics, web, news, blog, css, linux, art, osx, java, mac, blogs, reference, fun, python, games, tech, photography, humor, tools, delicious, rss, firefox, toread, comics*. Многие из них являются техническими тегами, которые отражают технологические интересы (*rss, firefox, python, java, linux*). Некоторые из них описывают жанр или носят описательный характер (*comics, humor, fun, photography*). Одним из интересных тегов является *toread*, он используется для самоорганизации и напоминания. Тег *wishlist* (<http://del.icio.us/tag/wishlist>) использовался пользователями для того, чтобы выдвинуть на первый план потребительские группы, которыми они были заинтересованы [6].

Существует достаточно много работ, связанных с фолксономией, часть из них посвящено классификации в фолксономии, созданию онтологий, анализу лексики, которую пользователи используют при аннотировании ресурсов и т.д.

В связи со спецификой тегов появляются программные разработки, которые позволяют произвести автоматическую разметку изображения. Однако и они не лишены недостатков, так как напрямую связаны с задачей распознавания образов.

Основные подходы, которые используются для обработки информации в фолксономиях – это статистические методы, методы вычисления семантической близости (расстояние Левенштейна, коэффициент Жаккара и др.) [7, 8]. В большинстве подходов используются различные лингвистические ресурсы: WordNet, Wikipedia, DBpedia, YAGO, ConceptNet и др. Однако эти ресурсы не совершенны для решения проблем обработки ключевых слов, так как не охватывают весь спектр лексики, представленной в фолксономиях.

Особое место занимает очистка тегов и приведение их к унифицированному виду. Для этого используют расстояния Левенштейна, стемминг, лемматизацию и словари-тезаурусы. После этапа очистки можно находить семантически близкие теги и проводить кластеризацию ресурсов.

Список литературы

1. <https://www.flickr.com/>
2. <https://instagram.com/>
3. <https://picasaweb.com/>
4. <http://photobucket.com/>
5. <http://en.wikipedia.org/wiki/Folksonomy>
6. Adam Mathes (2004) *Folksonomies – Cooperative Classification and Communication Through Shared Metadata*. Access: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
7. Lucia Specia, Enrico Motta (2007) *Integrating Folksonomies with the Semantic Web*. In: *The Semantic Web: Research and Applications Lecture Notes in Computer Science Volume 4519*, 2007, pp. 624-639.
8. Chao Wu, Bo Zhou (2009) *Semantic Relatedness in Folksonomy*. In: *International Conference on New Trends in Information and Service Science*, pp. 760 – 765.



ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ У ЛІНГВІСТИЦІ

Печенікова Л.М.

*Київський національний економічний університет ім. Вадима Гетьмана
e-mail: pechenikova.larisa@yandex.ua*

В останні роки глобальною тенденцією науково-технічного прогресу є розвиток і широке використання інформаційних технологій, що в свою чергу вплинуло на значні зміни у всіх сферах людської діяльності. Необхідно обумовити той факт, що прогресивний рух до інформаційного суспільства — це шлях до нового майбутнього людської цивілізації, зокрема галузі освіти.

Вчені зауважують, що інформатизація суспільства – це глобальний соціальний процес, що онтологічно змінює “картину світу” і активно впроваджує різне програмне забезпечення, глобальні мережі (Інтернет), мультимедійні технології тощо [3; 7; 8]. У такому випадку структура постіндустріального суспільства окреслена інформаційною значущістю, де інфосфера виступає домінуючою вартістю цивілізації.

Доцільність використання інформаційних технологій обґрунтовано в працях А. Ахаян, Т. Гуменникової, Н. Іваницької, К. Карпова, Л. Карташової, С. Коломійця, Ю. Колос, В. Терехової, А. Янковця. Вчені наголошують, що такі технологічні можливості при використанні сучасних інформаційних ресурсів дозволяють організувати різноманітну навчальну діяльність студентів.

Інноваційний характер та висока якість освіти можуть бути забезпечені шляхом впровадження нових освітніх технологій, інтерактивних форм навчання та методів, які дозволяють імітувати реальні ситуації, використання мультимедійних технологій та проектних методів тощо. А. Янковець зауважує, що нові сучасні технології впливають не лише на характер трансформації світової цивілізації, а й зумовлюють масову потребу в самостійному навчанні і постійному підвищенні кваліфікації [9; с. 3].

Інформатизація освіти зумовила нові вимоги до формування професійних якостей та рівня підготовки майбутніх спеціалістів, оскільки сучасний фахівець із вищою освітою повинен вільно орієнтуватися у всесвітній інформаційній мережі, мати відповідні знання для пошуку, обробки та зберігання необхідної інформації з використанням сучасних інформаційних технологій, комп'ютерних систем і т.п.

На сьогодні відбувається процес модернізації вищої освіти, змінюється її зміст, структура, методика і педагогічні технології викладання. Безсумнівно, одним із факторів, який впливає на зазначені зміни є глобальна інформатизація, яка кардинально змінює “обличчя” вищої школи. Ефективність процесу інформатизації прямо залежить від ефективності процесу створення і використання інформаційного потенціалу суспільства, оскільки інформаційний ресурс є не що інше, як сукупність інформації минулого і сучасного досвіду людства, а також підґрунтя для відтворення нової інформації.

На нашу думку, варто звернути увагу та проаналізувати технології в галузі



освіти в інших країнах світу, де різні компанії (KhanAcademy та ін.) передбачають та вкладають значні кошти в розробку освітніх сайтів, зокрема онлайн-школи (Keystone School); коледжі (University of Phoenix, Kaplan University, The Open University, University of the People); програми, які готують до тестів (Top Test Prep, GrockIt, Knewton, RevolutionPrep, TutorJam, BrightStorm), програми, які видають сертифікати (Alison.com) тощо.

В класичному освітньому процесі важливими є не тільки заняття, а й студентське середовище, в якому студент безперервно спілкується та розвивається. Цей фактор зумовив необхідність розробки відповідних сайтів в системі Інтернет, зокрема UnClasses.org та OpenStudy.com дозволяють студентам створювати різноманітні групи і навчатися разом, Quora.com і StackExchange — соціальні сервіси, які дозволяють ставити питання на професійні теми та надають на них відповіді. Також існує багато форумів, які присвячені спільному вивченню різних мов або співтовариства роботи викладачів [4; с.85].

У зв'язку з вищезазначеним, на нашу думку, у вітчизняних вищих школах здобуваючи у процесі вивчення інформаційних технологій знання, вміння та навички студенти спеціальності "Лінгвіст" має набути глибокі знання щодо проблем алгоритмізації, моделювання лінгвістичних задач, сучасних мов програмування, різних лінгвістичних програм та навички роботи з ними.

Варто зауважити, що лекційний матеріал для студентів-лінгвістів має бути побудований таким чином, щоб знання комп'ютера, вміння спілкуватися з ним не вимагало засвоєння основ математики, а необхідно звернути увагу на те, що кожний випускник філологічного факультету може стати вагомим учасником роботи у створенні комп'ютерних систем навчання, систем автоматичного аналізу і синтезу текстів, систем автоматичного пошуку, реферування чи перекладу текстів і т.п. Практичні заняття мають базуватися та охоплювати такі лінгвістичні програми:

- ▲ програми аналізу та лінгвістичної обробки текстів;
- ▲ програми для автоматичної обробки текстів;
- ▲ програми перетворення текстів;
- ▲ психолінгвістичні програми;
- ▲ генератори текстів;
- ▲ системи обробки природних мов і машинного перекладу;
- ▲ каталог і колекції ресурсів;
- ▲ словники та тезауруси;
- ▲ пошукові машини і системи повнотекстового пошуку;
- ▲ системи синтезу та розпізнання мови тощо [5].

Отже, високі технології безперечно змінюють навчальний процес, відбувається зміщення та перехід від традиційної освіти до методів, які побудовані на ІТ, їх використання у вищій школі є одним із пріоритетних напрямів інформатизації суспільства, тому необхідно удосконалювати систему підготовки у використанні інформаційних і комунікаційних технологій у підготовці лінгвістів та їх професійній діяльності.



Література

- 1 Гриценчук О.О. Теоретико-методологічні основи застосування ІКТ в процесі вивчення суспільствознавчих дисциплін в країнах зарубіжжя та Україні: загальні підходи [Електронний ресурс] // Інформаційні технології і засоби навчання. – 2010. – №5 (19). Режим доступу до журналу: <http://www.ime.edu-ua.net/em.html>.
- 2 Евдокимова И.А. Использование современных информационных технологий и методов в процессе обучения студентов / [Електронний ресурс]. – Режим доступу: <http://festival.1september.ru/articles/585473/>
- 3 Захарова И. Г. Информационные технологии в образовании / И. Г. Захарова. – М. : Academia, 2013. – 192 с.
- 4 Кожевникова А.С. Информационные образовательные технологии в России и за рубежом // Современные информационные технологии и ИТ-образование [Электронный ресурс] / Сборник научных трудов VII Международной научно-практической конференции. / под ред. В.А. Сухомлина. - Москва: МГУ, 2012. – Т.1. – С. 83-89.
- 5 Логичев С.В. Каталог лингвистических программ и ресурсов в Сети / С.В. Логичев // Русская виртуальная библиотека [Электронный ресурс]. 1999-2009. Режим доступа: <http://www.rvb.ru/soft/catalogue/index.html>.
- 6 Новые педагогические и информационные технологии в системе образования / Учебное пособие для студентов педагогических вузов и системы повышения квалификации педагогических кадров / Е. С. Полат, М. Ю. Бухаркина – М.:Академия, 2002. – 272 с.
- 7 Полат Е. С. Современные педагогические и информационные технологии в системе образования / Е. С. Полат, М. Ю. Бухаркина. – М. : Academia, 2011. – 368 с.
- 8 Трайнев И. В. Новые информационные коммуникационные технологии в образовании / И. В. Трайнев. – М. : ФЛП Дашков И.К., 2010. – 320 с.
- 9 Янковець А. В. Підготовка майбутніх перекладачів засобами інформаційно-комунікаційних технологій у вищих військових навчальних закладах [текст] : автореф. дис. ... канд. пед. наук : 13.00.04 / А. В. Янковець. – Хмельницький, 2005. – 20 с.



E-COMMUNITIES КАК ИСТОЧНИК ДАННЫХ В НЕФОРМАЛЬНОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ

Петрасова С.В.

*Национальный технический университет
«Харьковский политехнический институт»,
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,
e-mail:svetapetrasova@gmail.com*

Под **информационным пространством** принято понимать совокупность информационных ресурсов, технологий их сопровождения и использования, информационных и телекоммуникационных систем, образующих информационную инфраструктуру [1].

В информационном пространстве средством установления социальных связей являются коммуникации. Среди развивающихся технологий коммуникации следует выделить социальные сети (или неформальное информационное пространство). Под **социальной сетью** понимается социальная структура, состоящая из множества акторов (субъектов) и определенного на нем множества отношений (совокупности связей между акторами). Формально социальная сеть представляет собой граф $G(N, E)$, в котором $N = \{1, 2, \dots, n\}$ — конечное множество вершин (акторов) и E множество ребер, отражающих взаимодействие акторов [2].

Социальные сети предназначены для формирования, отображения и упорядочения социальных взаимоотношений. Другими словами, сайт социальной сети представляет собой автоматизированную социальную среду, позволяющую общаться группе пользователей, объединенных общим интересом.

В последнее время появилось несколько отдельных научных направлений, исследующие социальные сети. В частности, **Social Network Analysis** – направление структурного подхода, основными целями которого являются исследование взаимодействий между социальными акторами и выявление условий возникновения этих взаимодействий. Сеть социальных взаимодействий состоит из конечной совокупности социальных акторов и набора связей между ними [3].

Основными методами анализа социальных сетей являются:

- методы теории графов для изучения структурных взаимосвязей актора;
- методы нахождения локальных свойств субъектов, например, центральности, престижа, положения, принадлежности к некоторым подгруппам;
- методы определения эквивалентности акторов, включая их структурную эквивалентность;
- вероятностные модели;
- топологические методы, представляющие сеть в виде некоторого формализованного комплекса элементов и связей.

Характерными особенностями социальной сети являются:

- предоставление практически полного спектра возможностей для обмена информацией (фото, аудио, видео, сервис блогов и микроблогов, организация сообществ, форумы и чаты, возможность отметить местоположение и т. п.);



- создание профилей, в которых можно указать реальные персональные данные и информацию о себе;
- идентификация друзей (подавляющее большинство друзей пользователя в социальной сети – это не только реальные друзья, родственники, коллеги, одноклассники и однокурсники, но и виртуальные друзья по интересам).

Естественным свойством пользователей социальных сетей является тенденция к их объединению в различные сообщества. В социальных сетях пользователи объединяются явно (используя средства сети для создания групп и взаимодействия внутри них) либо неявно (устанавливая связи на основе общей или похожей деятельности, роли, социального круга, интереса или других свойств) [4].

Главным результатом данной тенденции является появление феномена онлайн-сообщества («e-community»).

E-community – новый тип сообществ, которые возникают и функционируют в электронном пространстве, а также – объединение пользователей сети в группы с общими интересами для работы в электронном пространстве [5].

Можно выделить следующие функции e-community: самовыражение; коммуникация; развлечение; образование (профессиональная): сети предоставляют широкие возможности для размещения разнообразной полезной информации, которая может быть использована в образовательных целях; информационная: сети предоставляют пользователям все возможности для размещения информации любого типа и получения обратной связи [6].

В социальных сетях e-community позволяет активно обмениваться мнениями, получать информацию, размещать текстовые, аудио и видеофайлы и т.д.

Вместе с тем, при работе с данными в неформальном информационном пространстве нужно учитывать такие факторы, как нестабильность качества пользовательского контента (спам и ложные аккаунты), проблемы с обеспечением приватности личных данных пользователей при хранении и обработке. Все это требует постоянного совершенствования алгоритмов решения различных аналитических задач.

Список литературы:

1. Додонов А.Г. Компьютерные сети и аналитические исследования /А.Г. Додонов, Д.В. Ландэ, В.Г. Путьгин. – К. : ИПРИ НАН Украины, 2014. – 486 с.
2. Губанов Д.А. Социальные сети: модели информационного влияния, управления и противоборства / Д.А. Губанов, Д.А. Новиков, А.Г. Чхартишвили. – М. : Издательство физико-математической литературы, 2010. – 228 с.
3. Boyd D. M. Social Network Sites: Definition, History, and Scholarship / D. M. Boyd, N. V. Ellison // Journal of Computer-Mediated Communication. – 2007. – № 13 (1). – PP. 210–230.
4. Коршунов А. Анализ социальных сетей: методы и приложения / А. Коршунов, И. Белобородов, Н. Бузун, В. Аванесов и др. // Труды ИСП РАН. – 2014. – Т.3 (1) – С. 439–456.
5. Глоссарий по информационному обществу: [Электронный ресурс]. – Режим доступа: <http://www.iis.ru/glossary/virtualcommunities.ru.html>.
6. Селезенев Р.С. Социальные сети как феномен информационного общества и специфика социальных связей в их среде / Р.С. Селезенев, Е.И. Скрипак // Вестник Кемеровского государственного университета. – 2013. – №2(54). Т.3 – С. 125–131.

ПРО ОДИН ПРОГРАМНИЙ ЗАСІБ КОМП'ЮТЕРНОГО АНАЛІЗУ ФОНОТЕКСТІВ

Іщенко О.С.

*Інститут української мови НАН України
Київ, вул. Грушевського, 4, тел. 0 (44) 279-20-07
e-mail: o.ishenko@gmail.com*

Фонотекстами називаємо аудіозаписи усного мовлення – передусім ті, що мають ознаки тексту як такого (тобто для яких властиві зв'язність, цілісність, структурність, завершеність тощо). Фонотексти (або аудіотексти) – це одне із цінних джерел сучасних лінгвістичних досліджень. Доказом цього є активне створення і використання так званих мовленнєвих корпусів – бібліотек аудіотекстів.

Для лінгвістичного опрацювання фонотекстів створюються спеціальні програмні засоби (комп'ютерне програмне забезпечення). Про один з них, який призначено для фонетичних досліджень, йдеться у цій розвідці.

Так, у фокусі нашої уваги – програма Praat [3], що широко використовується фонетистами для аналізу різних характеристик мовлення (часокількісних, тембрових, амплітудних тощо). Для цього програмного забезпечення існує низка доповнень у форматі так званих скриптів (сценаріїв)¹, які дозволяють отримати додаткові можливості аналізу мовлення. Зокрема, нашу вагу привернув скрипт Syllable Nuclei [5], який в інтерфейсі Praat реалізує можливість ділити мовлення (фонотексти) на фонетичні склади, тобто виконувати функцію складоподілу (див. рис. 1).

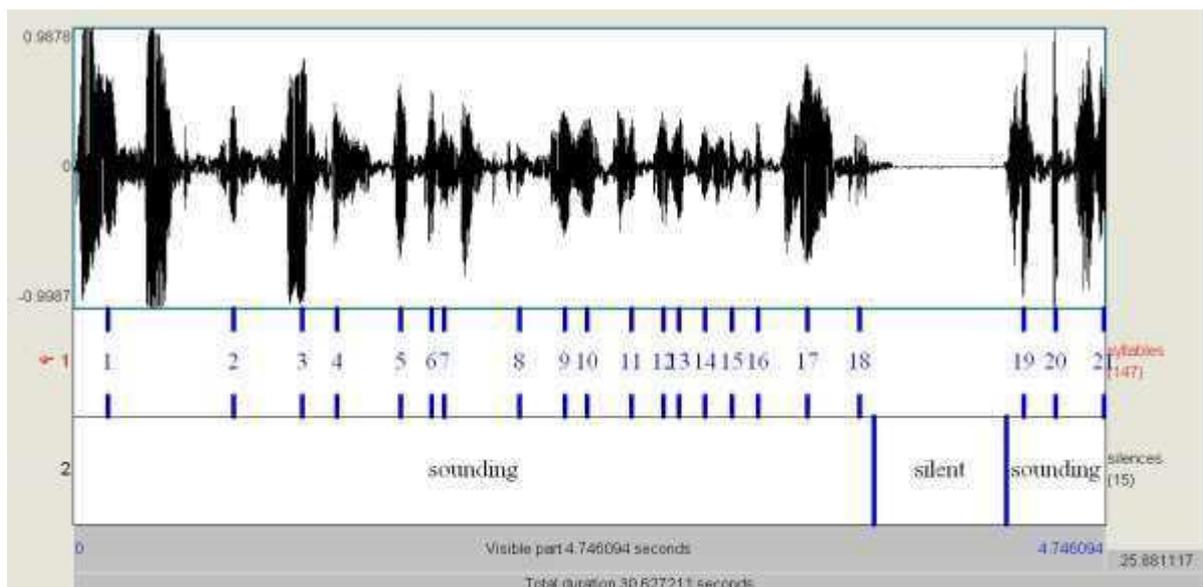


Рис. 1: Складоподіл у програмі Praat
за допомогою Syllable Nuclei (демонстрація)

¹ Окрім Syllable Nuclei, для Praat існують також інші скрипти (сценарії), наприклад, Beat-Extractor, ProsodyPro, які спрямовані на аналіз окремих фонетико-просодичних ознак мовлення.



Цей комп'ютерний інструмент може бути зручним для дослідження таких одиниць усного мовлення, як темп і ритм. Адже, наприклад, для визначення темпу мовлення необхідно знати кількість складів у досліджуваному тексті. Інструмент на зразок Syllable Nuclei створено для суттєвої оптимізації роботи дослідника, оскільки з його допомогою зникає необхідність "ручного" членування тексту на склади. Та чи можна довіряти цьому програмному засобу в дослідженні української мови? Запитання зумовлено, по-перше, тим фактом, що проблема українського складоподілу досі належить до актуальних (цілком не розв'язаних) в українській фонетиці; по-друге, тією обставиною, що механізм (алгоритм) автоматизованої сегментації мовлення на склади в Syllable Nuclei відбувається за принципами теорії складу Мермельштайна², яка, за нашими даними [1], має принципові недоліки (принаймні, з погляду особливостей фонетики української мови).

Мета нашої розвідки – здійснити тестування інструмента Syllable Nuclei на матеріалі української мови та з'ясувати, наскільки надійними є результати, отримані в ньому. Для того, щоб перевірити точність складоподілу в Syllable Nuclei, було проведено експеримент, який полягав у зіставленні результатів підрахунку фонетичних складів у мовленні, отриманих за двома різними методиками. Перша методика – це "ручний" спосіб підрахунку складів, здійснений відповідно до чинних стандартів українського складоподілу (правила складоподілу в українській мові див. у: [2]). Другий спосіб – це автоматизований спосіб підрахунку складів, реалізований за допомогою сценарію Syllable Nuclei в інтерфейсі програми Praat. В експерименті використано три фонотексти, кожен із яких було сегментовано на склади двома описаними вище способами. Три використані в експерименті фонотексти – це зразки мовлення різного темпу – повільного, помірного й швидкого.

Результати експерименту показали, що кількість складів, яку отримано "ручним" підрахунком і за допомогою автоматизованого аналізу, не збігається. Зокрема, різниця між ними на прикладі першого фонотексту (повільного темпу мовлення) становить 9%, другого фонотексту (помірного темпу мовлення) – 6% і третього фонотексту (швидкого темпу мовлення) – 20%. В середньому похибка автоматизованого підрахунку складів – 12%.

Використання фонотекстів різного темпу мовлення дало можливість з'ясувати, що найменш точні (недостовірні) результати комп'ютерної сегментації мовлення на склади засобами Syllable Nuclei варто очікувати при опрацюванні текстів швидкого мовлення.

Отже, здійснена розвідка, з одного боку, підводить до думки, що, у зв'язку з ненадійністю результатів обчислення, програмний засіб Syllable Nuclei, що сегментує мовлення на склади, не може повноцінно замінити "ручний" спосіб складоподілу; а з другого боку – свідчить про те, що склади мають різну фоне-

² Склад у розумінні П. Мермельштайна [6] визначається змінами амплітуди акустичних коливань: максимальну амплітуду асоційовано з вершиною складу, мінімальну – зі складовою межею. Теорія Мермельштайна є власне акустичною теорією фонетичного складу. Окрім неї, існують й інші акустичні концепції складу, зокрема теорія О. Есперсена [4], пов'язана з фізичною якістю звуків (сонорністю).



тичну реалізацію в мовленні залежно від темпу (що становить собою окрему мало досліджену наукову проблему, пов'язану зі складами), інакше б похибка в автоматизованому обчисленні кількості складів збігалася незалежно від швидкості вимовляння складів.

Список літератури

1. Іщенко О.С. Український складоподіл у світлі сегментації Мермельштайна (експериментально-фонетичне дослідження) / О.С. Іщенко // Матеріали III Всеукраїнської науково-практичної конференції "Інтелектуальні системи та прикладна лінгвістика". – Харків: НТУ "Харківський політехнічний інститут", 2014. – С. 38-40.
2. Сучасна українська літературна мова. Фонетика / Відп. ред. М.А. Жовтобрюх. – К.: Наук. думка, 1969. – 436 с.
3. Boersma P. Praat: computer software package for the analysis of speech in phonetics / P. Boersma, D. Weenink. – University of Amsterdam, 2005-2014.
4. Jespersen O. Lehrbuch der Phonetik / O. Jespersen. – Leipzig, 1904. – 258 S.
5. Jong de N. Praat Script Syllable Nuclei / N. de Jong, H. Quené, I. Persoon, T. Wempe. – University of Amsterdam, 2008-2010. – Ver. 02.
6. Mermelstein P. Automatic segmentation of speech into syllabic units / P. Mermelstein // Journal of the Acoustical Society of America. – 1975. – Vol. 58. – P. 880–883.



ДОСВІД ЗАСТОСУВАННЯ ЗНАНЬ З ПРИКЛАДНОЇ ЛІНГВІСТИКИ У СТВОРЕННІ СИСТЕМИ, ЩО ВИКОРИСТОВУЄ МЕТОДИ OPINION MINING TA COREFERENCE RESOLUTION

Терещенко В.І.

Search Concepts AG

4052 Basel, Switzerland, 59 Gartenstrasse,

phone: +41 61 715 95 06

e-mail: contact@searchconcepts.io

В епоху розвитку інформаційних і комунікаційних технологій та їх впровадження в усі сфери людської діяльності, вирішення лінгвістичних проблем та завдань набуває особливої актуальності. Це пов'язано у більшій мірі зі збільшенням кількості інформації, яку можна знайти в мережі Internet, а також з прагненням людства до прискорення якості та швидкості обробки цієї інформації. Результатом таких прагнень є створення лінгвістичних систем обробки та аналізу інформації. І одними з перших проблем, на які натрапляють науковці під час створення таких систем, є Opinion Mining та Coreference Resolution [1].

Coreference – відношення між займенником (анафором) та його антецедентом, за якого і займенник, і його антецедент співвідносяться з одним і тим же предметом об'єктивної дійсності [2]. Нижче наведено приклад речення з кореферентним зв'язком:

«Чистий прибуток Visa[antecedent] склав 1,27 млрд дол, йдеться в повідомленні компанії. За результатами II кварталу 2012 року вона[anaphor] повідомила про прибуток в 1,29 млрд дол.»

Coreference Resolution – процес встановлення кореферентних зв'язків в реченні або між реченнями [3].

Opinion Mining – (Sentiment Analysis, аналіз тональності тексту) метод контент-аналізу в прикладній лінгвістиці, призначений для автоматизованої обробки емоційно забарвленої лексики або емоційної оцінки автора, яка безпосередньо відноситься до об'єктів мови, про які йдеться в тексті.

Результатом розв'язання обох завдань може бути система, здатна до якісної обробки та аналізу природної мови. Створенням саме такої системи займається компанія «Search Concepts AG», з центральним офісом в м. Базель, Швейцарія. Створюваний компанією програмний продукт під назвою «News Board» має на меті автоматизований аналіз та обробку поширюваних в мережі internet новин фінансово-економічної тематики. Гордістю компанії є те, що вже сьогодні «News Board» може робити аналіз текстів на англійській, французькій, німецькій, російській та іспанській мовах. Найбільшу користь від використання програми отримують люди, що цікавляться станом справ в економіці певної країни чи компанії, або ж ті, що бажають інвестувати капітал в ту чи іншу галузь чи компанію. Засновник компанії, Hans-Peter Gier як ніхто інший знайомий

з проблемою автоматичної обробки фінансово-економічної інформації і говорить про те що зазвичай, для того щоб дізнатися реальний стан справ компанії, людина може витратити дуже багато часу та сил на пошук та аналіз новин. Проте, з цією системою той же результат можна отримати за декілька хвилин.



Рис. 1. – Скріншот програми News Board by Search Concepts

Незважаючи на те що центральний офіс компанії знаходиться в Швейцарії, розробка програми починалася в Україні зі співпраці кафедри ІКС та новоствореної компанії Search Concepts AG. Зокрема, великий вклад в розробку лінгвістичної основи програми зробили професор кафедри інтелектуальних комп'ютерних систем НТУ «ХП», доктор технічних наук Хайрова Ніна Феліксівна, а також студенти кафедри: Тетяна Булатнікова, Сільветте Луда, Альона Ковальова, Світлана Петрасова та Віталій Терещенко.

В ході співпраці було створено декілька програмних продуктів для вирішення окремих лінгвістичних задач. Такими є програма для встановлення кореферентних зв'язків в російськомовних текстах економічної тематики та програма для швидкого та повного пошуку перекладних еквівалентів в мережі Internet. Всі вони розроблялись як окремі модулі для основної програми, проте можуть використовуватись і самостійно.

Таким чином, результатом плідної співпраці між компанією Search Concepts AG та кафедрою інтелектуальних комп'ютерних систем НТУ «ХП» став програмний продукт, що дозволяє проводити сентімент-аналіз новин фінансово-економічної тематики на декількох мовах. Створений програмний продукт користується досить великою популярністю у користувача. Розроблене в ході співпраці програмно-алгоритмічне забезпечення є винятковим прикладом використання теоретичних навичок та знань з прикладної лінгвістики на практиці.

Список літератури

1. Clark J.H., Gonzalez-Brenes J.P. "Coreference Resolution: Current Trends and Future Directions", 2008. – pp. 11-16.
2. Hobbs J. R. "Pronoun resolution" – California, 1976. – pp. 10-18.
3. Скатов Д. "Разрешение кореференции: обзорная экскурсия" – Н. Новгород, ДИКТУМ, 2012.



ДОСЛІДЖЕННЯ ВЛАСТИВОСТЕЙ МОВИ СТАНДАРТІВ ЯК ЕКЗЕМПЛЯРУ КЛАСУ МОВ ДЛЯ СПЕЦІАЛЬНИХ ЦІЛЕЙ В КОНТЕКСТІ АВТОМАТИЗАЦІЇ ПРОЦЕДУРИ СЕРТИФІКАЦІЇ

Бутенко Ю.І., Шостак І.В.

*Національний аерокосмічний університет ім. М.Є. Жуковського "ХАІ",
м. Харків, вул. Чкалова 17, тел. 057-788-42-61,
e-mail: iuliabutenko@yandex.ru*

Сертифікація програмного забезпечення значною мірою визначає реальні можливості в забезпеченні необхідного рівня безпеки та якості в галузях і соціальних сферах, пов'язаних з використанням систем критичного застосування [1]. Процедура оцінювання програмного забезпечення передбачає вирішення ряду задач, серед яких слід виокремити задачу формування нормативного профілю.

Практичне проведення сертифікації на етапі формування нормативного профілю до тепер значною мірою полягає у ручному аналізі експертами великих обсягів нормативної та проектної документації, представлені текстами природньою мовою [2].

У той самий час в процесі пошуку вимог у колекції складно-структурованих текстів стандартів недостатньо отримати лише список релевантних документів в якості пошукової видачі через значний обсяг і високу складність документів. Підвищення ефективності пошуку у таких документах можна досягнути, якщо у якості пошукової видачі будуть отримані не лише документи, але й цитати з них – точні дослівні витримки з тексту, що мають смислову завершеність. Цитати можна отримати за допомогою аналізу ієрархічної структури текстів стандартів і нормативної документації, а потім можуть бути уточнені через використання семантичного аналізу. У результаті буде отримано компактну пошукову видачу, де відсічено значний обсяг інформації, нерелевантний запиту [3].

Для автоматизації формування нормативного профілю, що представляє собою набір вимог до об'єкта сертифікації, необхідно провести формалізацію шляхом розробки моделей, що з певним ступенем адекватності відображають усі особливості мови стандартів, якою написано всю нормативну документацію предметної області сертифікації програмного забезпечення систем критичного застосування.

Необхідно встановити загальну структуру мови стандартів, як сукупності елементів. Структуру сукупності може бути виявлено в результаті проведення ретельної класифікації досліджуваних одиниць, тобто їхнього ієрархічного розподілу за певною ознакою на основні розділи, які далі розпадаються на підрозділи, пункти, підпункти, вимоги, які у свою чергу розбиваються на окремі більш дрібні базисні одиниці - речення. Все це разом утворює номенклатуру - повний докладний перелік окремих елементів досліджуваної сукупності.



Такий підхід породжує подальшу можливість аналізувати за допомогою математичних методів як ймовірнісну, так і логічну структуру всього досліджуваного тексту в цілому. Номенклатурна класифікація створюється не довільно, а за аналогією з існуючими моделями в деяких інших «родинних» підмовах [4]. Елементи формалізації, притаманні лінгвостилістичним підсистемам, можна підрозділити на структурні і семантичні. Ці елементи знаходять своє відображення і в модельованому мовою стандартів.

Таким чином, для вибору адекватної моделі представлення стандартів доцільно проведення моделювання з подальшою конкретизацією розробленої системної моделі формування нормування профілю для сертифікації систем критичного застосування з допомогою математичного апарату теорії графів. Для комп'ютерної реалізації пропонується використання онтологічних моделей.

Лінгвістичний аналіз доцільно розпочинати з визначення притаманних текстам нормативної бази рис. У сучасній лінгвостилістиці статус стильової приналежності текстів нормативної бази, зокрема стандартів, є невизначеним. Саме тому, необхідно провести лінгвістичний аналіз текстів стандартів, визначити принципи для автоматизації процедури формування нормативного профілю для сертифікації програмного забезпечення систем критичного застосування моменти, та виокремити саме ті, які матимуть суттєвий вплив на якість інформаційного пошуку у колекції таких текстів.

Аналіз текстів мови стандартів показав, що його основними рисами є максимальна ясність викладу, стислість і лаконічність висловлювань, стислість і чіткість формулювань, що не допускають різних тлумачень, визначеність інформації, динамічність та експресивність її передачі, однозначність її сприйняття.

Основною особливістю організації мовних засобів у мові стандартів є їх узагальнено-абстрактний характер на лексичному і граматичному рівнях мовної системи. Лексику мови стандартів становлять три основних шари: загальноживані слова, загальнонаукові слова і терміни [5].

Тексти мови стандартів мають чітко виражену композиційну структуру, яка проглядається в чіткому розподілі структурних елементів тексту стандартів, його розподіл на розділи, підрозділи, пункти, підпункти. У свою чергу, такий поділ породжує особливість: значення нижніх елементів стандартів можна встановити однозначно тільки з урахуванням значення верхніх елементів.

Текст стандарту доцільно розділити на три частини [6]: попередню частину, рекомендації та вимоги й інформативну частину. До попередньої частини відноситься вступ, призначення, нормативні посилання, терміни та визначення, загальні відомості. До частини рекомендацій та вимог відносять розділи стандарту, в яких безпосередньо описано вимоги. Інформативна частина містить додаткову інформацію у вигляді додатків, що не увійшла в частину рекомендацій і вимог стандарту.

Прагненням до інформаційної насиченості обумовлено відбір найбільш емних і компактних синтаксичних конструкцій. У мові стандартів переважають прості поширені і складнопідрядні речення. Серед перших найбільш уживані невизначено-особисті з прямим доповненням до початку речення, синонімічні пасивним конструкціям. Абстрактність і узагальненість мови стандартів на



синтаксичному рівні виражається насамперед у широкому використанні пасивних конструкцій, безособових речень різних типів.

Модальні дієслова – це дієслова зі значенням можливості, повинності, бажання, тобто дієслова, що виражають ставлення повідомлюваного до реального здійснення. Одна частина розглянутих слів виражає реальну, стверджувальну модальність, достовірність, безперечність, очевидність; інша - гіпотетичну модальність, припущення.

Таким чином, найбільш суттєві для вирішення задачі автоматизації процедури сертифікації систем критичного застосування на етапі формування нормативного профілю лінгвістичні особливості текстів стандартів виявлені в їх композиційній структурі, синтаксичних особливостях речень з позиції виявлення їхньої семантики і ступеня повинності, виражений модальними дієсловами. Подання тексту стандарту у вигляді графа дає можливість під час інформаційного пошуку врахувати місце і взаємозв'язок заголовка/підзаголовка в тексті стандарту, а також дає однозначне уявлення про зміст цього фрагмента тексту. Моделювання семантичної структури речення дозволить виявляти речення ідентичні за семантикою, але різні за синтаксичною структурою та парадигматичними особливостями лексики мови стандартів. Модальність, широко виражена в модельованій мові стандартів, виражає ступінь обов'язковості дотримання вимоги до конкретного об'єкту сертифікації.

Результати моделювання мови стандартів є основою створення інформаційної технології підтримки діяльності аудитора сертифікаційного центру на етапі формування нормативного профілю. У подальшому передбачається втілити розроблену інформаційну технологію у формі інтелектуальної діалогової системи підтримки прийняття рішень аудитора сертифікаційного центру на етапі формування нормативного профілю.

Список літератури

1. Харченко, В.С. Анализ рисков аварий для ракетно-космической техники: эволюция причин и тенденций [Текст]/ В.С. Харченко, В.В. Скляр, О.М. Тарасюк // Радиоэлектронні і комп'ютерні системи. – Харків: НАКУ «ХАІ».– 2003. – Вип. 3. – С.135–149.
2. Липаев, В.В. Сертификация программных средств. Учебник [Текст] / В.В. Липаев. – М.: СИНТЕГ, 2010. – 344с.
3. Маннинг, К. Введение в информационный поиск: Пер. с англ. [Текст] / К. Маннинг, Пр. Рагхаван, Х. Шютце. – М.: ООО “И.Д. Вильямс”, 2014. – 528с.
4. Нелюбин, Л.Л. Перевод и прикладная лингвистика [Текст] / Л.Л. Нелюбин. – М.: Высш. школа, 1983 – 207с.
5. Даниленко, В.П. Лингвистический аспект стандартизации терминологии [Текст] / В.П. Даниленко. – М.: Наука, 1993. – 280с.
6. Ястребенецкий, М.А. Управление старением критических систем [Текст] / М.А. Ястребенецкий // Радиоэлектронные и компьютерные системы, 2008, №6, С.114-121.



CORPUS-BASED STUDIES IN LINGUISTICS

Pasenko K.V.

*Cherkasy Institute of Banking of the University of Banking of the National Bank of
Ukraine (Kyiv),
Cherkasy, 164 V. Chornovil Street, tel. (+380472) 71-99-51,
e-mail: k.pasenko@mail.ru*

In recent years a lot of investigations have been devoted to how computers can facilitate language learning. In this work we have made an attempt to examine the basic questions: what corpus is and what corpus linguistics is, how it can be applied to teaching English. Over the past twenty years, an approach to the study of language referred to as corpus linguistics has largely become accepted as an important and useful mode of linguistic enquiry [1]. Corpus linguistics is the study of language data on a large scale – the computer-aided analysis of very extensive collections of transcribed utterances or written texts [7].

The advent of computers led to the creation of what we consider to be modern-day corpora. The first computer-based corpus, the Brown corpus, was created in 1961 and comprised about 1 million words. Today, generalised corpora are hundreds of millions of words in size, and corpus linguistics is making outstanding contributions to the fields of second language research and teaching [2, p. 2].

Electronic (digital) language corpus is a new thing. It has a history of nearly half a century. Therefore, we are yet to come to a common consensus as to what counts as corpus, and how it should be designed, developed, classified, processed and utilised. The uniqueness corpus linguistics lies in its way of using modern computer technology in collection of language data, methods used in processing language databases, techniques used in language data and information retrieval, and strategies used in application of these in all kinds language-related research and development activities [4].

Different problems of corpus studies were investigated in great detail by G. R. Bennett, D. Biber, N. S. Dash, G. Leech, T. McEnery, N. Nesselhauf, T. Virtanen and A. Wilson.

The term "**corpus**" is derived from Latin corpus "body". At present it means representative collection of texts of a given language, dialect or other subset of a language to be used for linguistic analysis. Theoretically, corpus is

- (C)apable
- (O)f
- (R)epresenting
- (P)otentially
- (U)nlimited
- (S)elections of texts.

It is compatible to computer, operational in research and application, representative of the source language, processable by man and machine, unlimited in data, and systematic in formation and representation [3]. The term "corpus" when used in



the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for. These may be considered under four main headings: sampling and representativeness; finite size; machine-readable form; a standard reference [6, p. 29].

There are a number of areas where language corpus is directly used as in *language description, study of syntax, phonetics and phonology, prosody, intonation, morphology, lexicology, semantics, lexicography, discourse, pragmatics, language teaching, language planning, sociolinguistics, psycholinguistics, semiotics, cognitive linguistics, computational linguistics, etc.*

In fact, there is hardly any area of linguistics where corpus has not found its utility. This has been possible due to great possibilities offered by computer in collecting, storing, and processing natural language databases. The availability of computers and machine-readable corpora has made it possible to get data quickly and easily and also to have this data presented in a format suitable for analysis corpus as knowledge resource:

- 1) corpus is used for developing multilingual libraries;
- 2) designing course books for language teaching;
- 3) compiling monolingual dictionaries and thesaurus (printed and electronic);
- 4) developing bilingual and multilingual dictionaries (printed and electronic);
- 5) various reference materials (printed and electronic version);
- 6) developing machine readable dictionaries (MRDs);
- 7) developing multilingual lexical resources, electronic dictionary [4].

According to some researchers, in order to conduct a study of language which is corpus-based, it is necessary to gain access to a corpus and a ***concordancing program***. A corpus consists of a databank of natural texts, compiled from writing and / or a transcription of recorded speech. A concordance is a software program which analyses corpora and lists the results. The main focus of corpus linguistics is to discover patterns of authentic language use through analysis of actual usage [5].

Corpus linguistics is a comparatively new field of language research and application. Corpus linguistics is perhaps best described for the moment in simple terms as a study of language based on examples of real life language use. It has a long and interesting history. Yet the term corpus linguistics is a relatively modern term [6]. Corpus linguistics is not a branch of linguistics in the same sense as syntax, semantics, sociolinguistics and so on. Corpus linguistics does, however, allow us to differentiate between approaches taken to the study of language and, in that respect, it does define an area of linguistics or, at least, a series of areas of linguistics. Hence we have corpus-based syntax as opposed to non-corpus-based syntax, corpus-based semantics as opposed to non-corpus-based semantics and so on. So, while corpus linguistics is not an area of linguistic enquiry in itself, it does, at least, allow us to discriminate between methodological approaches taken to the same area of enquiry by different groups, individuals or studies [6].

It is generally accepted that the invention and advancement of computer technology in the last century has eventually added a new dimension to the field of linguistics. In recent times, as a result of this innovation, a comparatively new field, namely, the Computational Linguistics has evolved as an important area of Artificial



Intelligence, which aims at looking at language as an essential instrument of human communication directly linked with human cognition [9]. It is a method of carrying out linguistic analyses. As it can be used for the investigation of many kinds of linguistic questions and as it has been shown to have the potential to yield highly interesting, fundamental, and often surprising new insights about language, it has become one of the most wide-spread methods of linguistic investigation in recent years [8].

Corpus linguistics is a multidimensional area. It is an area with a wide spectrum for encompassing all diversities of language use in all domains of linguistic interaction, communication, and comprehension. The introduction of corpus in language study and application has incorporated a new dimension to linguistics.

Corpus linguistics is an approach that aims at investigating language and all its properties by analysing large collections of text samples. Corpora, and other (non-representative) collections of machine readable text, now mean that the lexicographer can sit at a computer terminal and call up all the examples of the usage of a word or phrase from many millions of words of text in a few seconds. This means not only that dictionaries can be produced and revised much more quickly than before – thus providing more up-to-date information about the language – but also that the definitions can (hopefully) be more complete and precise [6].

Corpus examples are important in language learning as they expose students at an early stage in the learning process to the kinds of sentences and vocabulary which they will encounter in reading genuine texts in the language or in using the language in real communicative situations.

Corpora have been used not only in language teaching but also in the teaching of linguistics. The link between findings of corpus-based research and (foreign) language teaching is that corpus evidence suggests which language items and processes are most likely to be encountered by language users (what is frequent and typical) and may thus deserve more time in classroom instruction. Corpora and corpus-data 1) help teachers and students make better informed decisions and improve teaching material to become more authentic, i.e. representative of contemporary usage; 2) help students to develop their own descriptive and analytical skills for improving language awareness.

By their nature, corpus activities are for more advanced levels, but they can be adapted for students at lower levels. By focusing on words which have a high frequency of occurrence and by concentrating on the usual rather than the exceptional, teachers can help learners acquire the language more efficiently, especially at elementary and intermediate levels.

To summarize, corpus linguistics is the foundation and an integral part of most linguistic studies. While it may be argued that corpus linguistics is not really a domain of research but only a methodological basis for studying language, one can in fact use corpora as the basis for an empirical approach to linguistics. Corpus linguistics adherents believe that reliable language analysis best occurs on field-collected samples, in natural contexts and with minimal experimental interference.

References

1. Baker, P. (2010) *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press. <http://www.ling.lancs.ac.uk/staff/paulb/socioling.htm>



2. Bennett, Gena R. (2010) *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. Ann Arbor, Michigan: University of Michigan. 144 p. <http://www.press.umich.edu/pdf/9780472033850-part1.pdf>
3. Dash, N. S. (2005) *Corpus Linguistics and Language Technology: With Reference to Indian Languages*. New Delhi: Mittal Publications.
4. Dash, N. S. (2010) *Corpus Linguistics: A General Introduction*. Proceedings of the Workshop on Corpus Normalization, LDCIL, CIIL, Mysore, 25 August, pp. 1-25. <http://www.ldcil.org/download/Corpus%20Linguistics.pdf>
5. Krieger, D. (2003) "Corpus Linguistics: What It Is and How It Can Be Applied to Teaching". <http://iteslj.org/Articles/Krieger-Corpus.html>
6. McEnery T., Wilson A. (2004) *Corpus Linguistics: An introduction* (2nd ed.). Edinburgh: Edinburgh University Press, 247 p.
7. McEnery T., Hardie A. (2012) *Corpus Linguistics: Method, Theory and Practice* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press, 312 p.
8. Nesselhauf, N. (2004) "Learner corpora and their potential for language teaching." In John Sinclair, ed., *How to Use Corpora in Language Teaching*. Amsterdam & Philadelphia: Benjamins, pp. 125-152.
9. Nesselhauf, N. (2011) "Corpus Linguistics: A Practical Introduction". <http://www.as.uniheidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf>



ТЕХНОЛОГИИ ИЗВЛЕЧЕНИЯ ЗНАНИЙ НА ОСНОВЕ ФАКТОГРАФИЧЕСКИХ ДАННЫХ ЭВМ В СТРУКТУРИРОВАННЫХ КОНТЕНТАХ

Дорошенко А.Ю.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,
e-mail: doroshenkoanastasiia@gmail.com*

В работе обсуждаются новые технологии извлечения знаний на основе фактографических данных, предлагается подход построения семантической сети, позволяющий работать с большими структурированными корпусами текста.

Цель работы – сравнение наиболее актуальных средств интеллектуального анализа структурированных контентов на основе построения семантической сети фактографической информации ЭВМ с помощью алгебры предикатов.

В ближайшем будущем наиболее востребованными станут системы с максимально автоматизированными ETL-процессами структурирования контента (extract, transfer, load- «извлечение, преобразование, загрузка»). Важной чертой таких систем будет функция оперативного анализа информации, полученной по запросу для выбора дальнейшего направления исследования документов (автопилотирование направления исследования), выполняемой с помощью методов интеллектуального анализа текста [4].

К актуальным средствам интеллектуального анализа текстов относятся технологии выделения фактографической информации об объектах с учетом анафорических ссылок на них (ссылочные местоимения на объект, поименованный в тексте); нечеткий поиск; тематическое и тональное (точное и полное) рубрицирование; кластерный анализ хранилищ и подборок документов; выделение ключевых тем; построение аннотаций; построение многомерных частотных распределений документов и их исследование с помощью OLAP-технологий; использование методов анализа текста для определения направления исследования больших подборок документов и извлечения новых знаний [1,2].

Факты выделяются из предложений, содержащих упоминания объектов или ссылки на них. Технология выделения фактов основана на использовании специальных семантико-лингвистических методов, которые дают возможность получить точность и полноту фактов, сравнимую с экспертными. Зачастую факты содержат информацию о взаимосвязях объектов и классифицируются как прямые (имеется факт о связи двух объектов); нечеткие (нет фактов); общего места и времени (для пары различных фактов различных объектов); косвенные, или транзитивные (через общий третий объект-связь у пары фактов различных объектов); рефлексивные (между парой атрибутов досье, связанных семантически). Если в одном из них появляется факт с определенным объектом-



связью, то в симметричном атрибуте для объекта-связи также появляется этот факт.

Все эти свойства необходимы в системах аналитической разведки, немислимых без следующих сервисов: автоматическое выявление прямых и косвенных связей объекта; автоматическое выявление связей объектов по месту и времени; типизация связей, представленных различной лексикой; формирование групп объектов, связанных между собой общностью фактов; построение карты связей объектов для различных типов связей, визуализация и фильтрация связей; поиск оптимальных связей между заданными объектами; построение многомерных частотных распределений фактов. Сегодня системы извлечения фактов являются наиболее эффективным инструментом выделения нужной для принятия решений информации, заменяя ее поиск [3].

Заключение. На основе анализа и сравнения наиболее актуальных средств интеллектуального анализа структурированных контентов, была построена модель семантической сети ЭВМ с использованием фактографических данных и алгебры предикатов. Поиск в семантической сети определенного факта, дает возможность считать, что подсеть установлена, после чего производится извлечение сущностей и их маркировка ролями, заданными в соответствующих узлах лингвистических описаний [1,4]. Таким образом, результатом семантического поиска на основе интеллектуального анализа структурированного контента, является имя факта и набор указателей на сущности семантической сети с указанием соответствующих им ролей в лингвистическом описании.

Список литературы

1. Алисейко З. А. Использование алгебры предикатов и предикатных операций для формализации декларативной и процедурной составляющих знаний / З. А. Алисейко, В. И. Булкин, О. В. Канищева, Н. В. Шаронова // Біоніка інтелекту. – Харків : ХНУРЕ, 2006. – № 1(64). – С. 59-63.
2. Бондаренко М. Ф. О мозгоподобных ЭВМ / М. Ф. Бондаренко, З.В. Дударь, И.А. Ефимова, В.А. Лещинский, С.Ю. Шабанов-Кушнаренко // Радиоэлектроника и информатика. – Харьков : ХНУРЭ, 2004. – № 2. – С. 89-105.
3. Булкин В.И. Математические модели знаний и их реализация с помощью алгебропредикатных структур / В. И. Булкин, Н.В. Шаронова: монография. – НТУ «ХПИ», МЭГИ. : Донецк, 2010. – 304 с.
4. Sussna M. Word sense disambiguation for free-text indexing using a massive semantic network // The Second International Conference on Information and Knowledge Management. – 1993. – P. 67-74.



PROBLEMS OF MACHINE TRANSLATION

Valchuk K.I.

*National Technical University of Ukraine
"Kyiv Polytechnic Institute",
Kyiv, prospekt Pobedy, 37, tel. 044–336–7989,
e-mail: krvalchuk@gmail.com*

The idea of machine translation appeared in the 17th century. In 1629, René Descartes proposed a universal language, with equivalent ideas in different tongues sharing one symbol. The field of "machine translation" appeared in Warren Weaver's Memorandum on Translation (1949). The first researcher in the field, Yehosha Bar-Hillel, began his research at MIT (1951). A Georgetown University MT research team followed (1951) with a public demonstration of its Georgetown-IBM experiment system in 1954. MT research programs popped up in Japan and Russia (1955), and the first MT conference was held in London (1956). Researchers continued to join the field as the Association for Machine Translation and Computational Linguistics was formed in the U.S. (1962) and the National Academy of Sciences formed the Automatic Language Processing Advisory Committee (ALPAC) to study MT (1964). Real progress was much slower, however, and after the ALPAC report (1966), which found that the ten-year-long research had failed to fulfill expectations, funding was greatly reduced.

Since early 80s, when PCs have started to conquer the world, time of their work got down in a value and the machine translation finally has become economical. In addition, improvement of the programs for these and next years allowed to translate many types of text quite accurately, however, some problems of machine translation still have been staying unresolved.

Sometimes we replace term "machine" with "automatic", which does not affect the sense. Although the term "automatic translation" has different meaning: it is a program that helps a person to translate some kind of text.

Using automatic translation, we expect the following forms of interaction:

- Partly automatic translation: for example, when a translator uses computer dictionaries.

- Systems with division of labor: computer is educated to translate only phrases with a rigidly set structure (the result is not required to be corrected afterwards), but everything beyond the scheme should be worked up by a person.

Nowadays, there are two different terms in the English terminology: machine translation, MT (completely automatic translation) and machine-aided or machine-assisted-translation, MAT (automated). If there is a need to determine both, M(A)T is used.

Moreover, there are two fundamentally different ways to construct an algorithm for machine translation: rule-based and statistical-based.

The current stage of machine translation development can be described as a stage of syntactical word translation. A fundamental unit of the sense is a word, but



the definition of grammatical forms and the sequence of words in a generated text based on syntactical relationships between the words in the original text. Similarly, to the formed rules the structure of dictionaries for translations reminds a word dictionaries' structure, in which case the percentage of collocations is quite low.

In spite of the fact that the word is not the only unit of the sense in native languages, there are a small number of levels of semantic units. Among them morphemes level, word level, collocation level, sentence level, etc. What is more, the sense of the higher level units not always may be determined from the sense of units of the lower level (most meaningfully it might be traced in the morpheme and word layers relationship). Accordingly, the dictionaries for an automatic translation must contain units of various layers.

Depending on style and purpose of the text, the same word could have had different meanings. In some extent this feature is being considered in the systems of machine translation, for instance, removable dictionaries are provided sometimes even for all types of text.

Besides established rules of sentence construction, there are several unpre-scribed rules in each language called the pulchritude of a language. For example, a literal translation of a sentence "This is my pen" in Russian is "Это есть моя ручка" which is formally correct, but Russian speakers do not talk in such way. In this case, it looks like someone weird or foreigner has constructed the sentence.

In contrast, capital letters and reductions are fraught with another trick. When a word starts with a capital letter, its translation also starts with a capital letter. The effective abbreviations, read as a unit word, are frequent in English literature. A computer will also translate such abbreviations as unified words.

Sometimes it is better to perform a word-processing before translation. Thus, machine translation systems can make mistakes because of reductions ended with a point. If a word with a capital letter follows such a point, computer perceives the point as the end of the sentence. As a result, we get an erroneously parsed and incorrectly translated sentence at the output. Therefore, it is better to remove the points after reductions. There must not be any word wraps in a translatable text. Text redactor Microsoft Word can easily provide it. The texts resulting in the identification of a scanned image needs a special attention because remained errors often are being the reasons that some words left untranslated in a case of one wrong letter.

Thus, automated dictionary service should take an important place in the systems of machine translation, whose tasks include creation and management (supplementation, correct). The role of a person in the automated dictionary service is marking foreign texts in order to segregate dictionary elements and to set appropriate equivalents in the output language. A machine can also perform a text marking.

To conclude, a computer still cannot fully replaced a translator. Is there any sense to apply machine translation systems at all? The answer is absolutely there are. If a computer translates literary texts, we can get a rough version of texts, which grow to the masterpiece under the hand of a person who does not have strong language skills but is just as good as literary editor. Talking about technical text translation, in a point of using an appropriate dictionary, a quiet passable is not requiring the following treatment result may get at the output. Actually, the need of redacting a machine



translation appears due to the described problems. Overall, the systems of machine translation must have their own editing tools.

The constructing of a qualitative machine translation of the scientific, technical and business texts should base on using of dictionary complexes including semantic units of different word, collocation and sentence layers. The leading place should take collocations because collocations are more usual for naming concepts than separate words. It is important to abide the rule of preference during translation. According to this rule, the algorithm should isolate and translate semantic units of the highest layer firstly but appeal to the units of lower layers only in a pinch if there is no other possibility to translate the text.

List of references

1. *Aristov N.V.* Osnovy perevoda. Moscow: is-vo literatury na inostrannyh yazikah, 1959 (in Russian)
2. *Cohem J.M.* Translation, 1986
3. *Vannikov Y.V.* Yazikovaya slozhnost` texta kak factor trudnosti perevoda (Metodycheskoye posobie). Moscow: Vsesoyuzniy centr perevodov, 1988 (in Russian)
4. Wikipedia, the free encyclopedia. Machine translation



ЗАСТОСУВАННЯ МУЛЬТИМЕДІА В НАВЧАННІ ФАХОВІЙ АНГЛІЙСЬКІЙ МОВІ (НА ПРИКЛАДІ ДИСЦИПЛІНИ «ІНЖЕНЕРНІ ОСНОВИ АЕРОКОСМІЧНОЇ ТЕХНІКИ»)

Величко А.М., Рижкова В.В.

*Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»
м. Харків, вул. Чкалова, 17, тел.. 788-47-06,
e-mail: foxenfoxen@mail.ru, k706@khai.edu*

В умовах глобалізації та удосконалення науки і техніки виникає потреба у підготовці майбутніх професіоналів, які не лише володіють іноземними мовами на найвищому рівні, але й мають обширні знання в певних технічних галузях. Оскільки забезпечення такої підготовки часто ускладнюється значними розбіжностями та невідповідностями у навчальних програмах технічних і мовних спеціальностей, наша робота з адаптації навчального матеріалу з дисципліни «Інженерні основи аерокосмічної техніки (спец. термінологія)» для студентів спеціальності «Прикладна лінгвістика» в авіаційних та аерокосмічних ВНЗ є актуальною.

Метою нашого дослідження є створення комплексу завдань на розуміння та засвоєння авіаційної термінології з дисципліни «Інженерні основи аерокосмічної техніки».

Об'єктом дослідження є способи підвищення мовної компетенції студентів під час перекладу текстів або англомовного спілкування на теми, пов'язані з аерокосмічною технікою.

Предметом дослідження є шляхи презентації англомовного навчального матеріалу із залученням елементів невербальної комунікації в технічних текстах та мультимедіа.

Матеріалом дослідження слугували англомовні тексти за темами «Історія розвитку авіації», «Класифікація літаків за їх призначенням», «Загальні принципи теорії польоту ЛА, конструкція ЛА», «Основні частини літака, їх призначення та конструкція», «Основні енергоємні системи ЛА», «Авіаційні двигуни», «Кабіна екіпажа», зкомпоновані з навчальних матеріалів за цією тематикою.

Мовна компетенція передбачає здатність людини розуміти і продукувати необмежену кількість правильних у мовному відношенні пропозицій, використовуючи засвоєні мовні знаки та правила їх з'єднання. Існує три рівні розвитку мовної компетентності: низький, середній і високий.

Досягнення мовної компетенції є основною метою сучасної освіти під час навчання іноземній мові в межах професійної діяльності.

На даному етапі розвитку вітчизняної педагогічної науки процес навчання іноземній мові за професійним спрямуванням має свої недоліки, і, таким чином, потребує значних змін і удосконалень.



Людині властиве асоціативне мислення, а отже застосування технічних засобів навчання є необхідним компонентом викладання іноземної мови, оскільки вони полегшують сприйняття та засвоєння студентами інформації.

Наразі такі технічні засоби навчання як презентації, аудіо- та відеозаписи активно використовуються у багатьох сучасних вищих навчальних закладах з метою оптимізації роботи студентів.

Досить популярним ресурсом формування лексичної компетенції студентів на сьогоднішній день є електронні словники.

Електронні словники сприяють кращому сприйняттю та засвоєнню інформації шляхом впливу на різні органи почуттів користувача. Таким чином, використання електронних словників значно прискорює процес навчання та робить його більш ефективним.

Відповідно до поставленої задачі, нами було опрацьовано англійські, українськомовні та російськомовні тексти за темами «Історія розвитку авіації», «Класифікація літаків за їх призначенням», «Загальні принципи теорії польоту ЛА, конструкція ЛА», «Основні частини літака, їх призначення та конструкція», «Основні енергоємні системи ЛА», «Авіаційні двигуни», «Кабіна екіпажа». До текстів було складено російсько-українсько-англійські словники, до яких увійшло 395 авіаційних термінів. Для оптимізації процесу засвоєння студентами лексики, нами було переведено ці словники у формат .xls. Завдяки цьому, словники можна підключити до програми ABBYY Lingvo Tutor та постійно працювати з навчальними картками. Також до кожного тексту було дібрано навчальні відеоматеріали та розроблено по три блоки завдань на перевірку засвоєння авіаційної термінології та розуміння пройденого матеріалу з вищезазначених тем. Кожен блок містить різні типи вправ, за результатами виконання яких можна оцінити знання студентів.

Перший блок пропонує до виконання завдання на визначення поданих англійською мовою слів та словосполучень, заповнення пропусків у реченнях згідно прочитаного тексту, вибір правильного терміну з кількох наведених. Більш складні завдання другого блоку передбачають визначення правильних та неправильних тверджень, визначення зайвого за смислом слова, розміщення подій у хронологічній послідовності, опис наведених зображень, співвіднесення зображень приладів з їх назвами. У завданнях третього – найскладнішого – блоку студенти мають перекласти речення відповідно до теми кожного тексту англійською мовою з української або російської; цей блок також передбачає завдання на заповнення пропусків у схемах, заповнення таблиць, розшифровку аббревіатур, малювання схем, усний опис запропонованих ситуацій, що можуть виникнути у професійній діяльності.

Таким чином, студентам пропонуються різноманітні типи завдань різних рівнів складності, за результатами виконання яких можна буде оцінити як рівень засвоєння студентами теоретичного матеріалу за певними темами з авіаційної галузі, так і їхні знання англійської термінології з цієї галузі.



TECHNICAL COMMUNICATION SKILLS IN THE CONTEXT OF TRAINING SPECIALISTS FOR SOFTWARE OUTSOURCING FIELD

Nikonorov S.I.

National Technical University

"Kharkiv Politechnic Institute"

Kharkiv city, 79/2, Pushkinskaya str., tel. 707-63-60,

e-mail: sivn@mail.ru

Eight years ago, in February 2007 in our University the Intelligence Computer Systems Department was established with the aim of training specialists in the Applied Linguistics field. This event coincided remarkably, with Ukraine's intensive entering the international outsourcing business in the area of software development. Within the last ten years, we have observed a growing number of companies in the Ukrainian IT market specializing in developing high quality software products and providing quality assurance (QA) services for customers in the USA, Canada and Western Europe. Together with the growth of these businesses, the demand for software developers and QA specialists is constantly increasing. In spite of the general economic recession in Ukraine, the IT industry is growing persistently. This fact is favored by western companies' striving for cost reduction by means of outsourcing.

Joe Welinske, a well-known expert in Technical Communication and User Assistance in California, cites some US technical writers sharing their concerns in his 2014 Skills and Technologies Survey: "... Company wants fewer employees and more outsource", "Losing US jobs, hiring in India" and so on [4].

The specialists trained by our department are unique. Combining teaching programming and foreign languages, the department staff strives for educating professionals who would work both as software developers and as translators. At the same time, training in such promising fields as technical communication and user assistance has dropped out of sight for quite a long time. It is an astonishing fact. After all, these are the careers where Applied Linguistics graduates can find self-fulfillment as IT specialists if they cannot or do not want, due to various reasons, to become programmers or translators.

The following information can emphasize actuality and high degree of development of these trades, and of the technical writer profession in particular, in the USA. In his research, Joe Welinske asked 444 respondents working in the area of technical communication to give the titles of their positions in the company. The result of this survey was a list of 158 titles. The most frequent ones were Documentation Manager (10), Documentation Specialist (7), Information Developer (10), Lead Technical Writer (5), Principal Technical Writer (12), Senior Information Developer (10), Senior Technical Writer (62), Technical writer (101). The collocation of words "technical writer" occurred in 26 titles [4].

Are good technical writing skills important in software development business?

In his book "Technical Writing made easier" Bernhard Spuida says, "When talking about algorithms, or sequences of events in a program, absolute clarity of writing



is not only needed in the code discussed; but also in documenting this particular program for our fellow programmers and users. We need to attain the same level of clarity of expression in both cases, otherwise readers will turn to other programs, which are more accessible on the level of understanding and therefore easier to use or extend" [1].

For technical documents to be useful, readers must be able to understand and employ them without having to decode wordy and ambiguous prose. Good technical writing clarifies technical jargon; it presents useful information that is clear and easy to understand for the intended audience. Poor technical writing may increase confusion by creating unnecessary technical jargon or by failing to explain unavoidable technical terms that readers would not be expected to know.

"If a technical text is unreadable in the reader's eye, he will quite probably assume that the product described in this text also is of inferior quality. Code is a language, just as the language of the documentation is. Not writing well in documentation implies faults in coding style. Therefore, readability is an absolute requirement for documentation of successful products," says Bernhard Spuida [1].

Most IT professionals are required to have some technical writing skills. In the information technology realm, project and product managers, business (system) analysts and solution (system) architects write a large number of technical documents that must be clear and concise. Articles and essays need to be well thought out and ordered. How the writer introduces the piece, builds on the introduction through the body, and concludes will largely determine how the information is accepted. Step by step, the writer must present main ideas, supporting evidence, analyses and conclusions in a logical and organized manner. The writing must not wander, but keep to its task of presenting the writer's information in the clearest possible way [2].

Technical writers use computers and other electronic communications equipment extensively in performing their work. They also work regularly with publishing software and various authoring environments to prepare material directly for the Internet. Technical writers frequently work with word processing, graphic design, page layout, and multimedia software. The nature of technical writing is evolving, and modern technical writers combine text, graphics, images, and sound into their work.

And here we are coming to a very significant distinction between two terms – **Technical Writer** and **Technical Communicator**, as it is given in Wikipedia.

A Technical Writer writes documents (user guides, administrator guides, online Help, etc.).

A Technical Communicator is a newer term for the same job acknowledging that the writer needs to do more than simply write documents, but also needs to understand the intricacies of systems, processes, business flows, and can develop content that includes more technical reference and information [3].

Technical writing specifically in support of software applications is commonly referred to as User Assistance.

Effective user assistance development requires a wide variety of communication skills. These include writing, editing, task analysis, and subject matter expert (SME) interviewing. Since the user assistance profession is directly involved with software development, the discipline often requires an understanding of user interface design,



usability testing, localization, testing, quality assurance, instructional design, scripting or programming, and accessibility.

Technical communication skills provide the foundation. They are supplemented by skills unique to the software development world. In his survey, Joe Welinske asked the respondents to value the importance of a number of skills commonly employed by user assistance professionals in their daily work.

Here go the top ten skills valued highly with a rating of “4” (Very Valuable) or “5” (Invaluable), the top two ratings on a five-point scale.

Content development skills are highly rated with Writing procedures (80%) in first place. As for other aspects of content development, Task analysis (68%), Writing reference information (68%), and Interviewing (71%) are all highly rated by three quarters of respondents. Information architecture (60%), Copy editing (58%), and Developmental editing (47%) are valued highly by at about half of respondents. Search (48%) remains important to almost half of respondents.

Expertise with authoring tools being a key skill (77%) is up at the top of Other Skills list. The nature of working with a digital medium like software user assistance requires the use of a variety of tools. Project planning is an important skill no matter what your role is. Almost three-quarters (71%) of respondents rank this highly [4].

As we can see, the development of technical communication skills in general and technical writing skills in particular at practically all major participants of software product development and sales process has taken on special significance lately. Although the position of technical writer has not become widespread in the structure of Ukrainian IT companies, the further development of the international outsourcing software business may change the situation. The growing demand for specialists in the field of technical communication and user assistance requires the necessity of changes, like, for example, introduction of Technical Writing theory course, in the educational process and the Applied Linguistics syllabus.

Bibliography

1. *Bernhard Spuida*. Technical Writing made easier. Rev. 1.1, 2002. – 17 p.
2. *Mary K. McCaskill*. Grammar, Punctuation, and Capitalization. – Langley Research Center, Hampton, Virginia, 1998. – 108 p.
3. Wikipedia. Technical communication: http://en.wikipedia.org/wiki/Technical_communication.
4. *Joe Welinske*. Skills and Technologies Survey 2014: <http://www.welinske.com/skills-and-technologies-survey-2014>.



ПРОЯВ ГЕНДЕРА В МОВЛЕННЄВІЙ ДІЯЛЬНОСТІ

Тулюлюк К.В.

*Чернівецький Національний університет,
м. Чернівці, вул. Коцюбинського 2,
e-mail: kt.80@mail.ru*

Соціолінгвістична модель функціонування мови є складною, багатокомпонентною системою багатьох взаємопов'язаних елементів, які обумовлені, в першу чергу психічними та фізіологічними особливостями людини. З іншого боку, соціальними, до яких належать культурні, етнічні чинники, та економічними факторами. Все це впливає на формування індивідуального мовного варіанта – ідіолекту, до якого входять екстралінгвістичні та лінгвістичні аспекти мовленнєвої поведінки [5, с.13].

В центрі досліджень сучасної гендерної лінгвістики знаходиться питання про особливості, та, в першу чергу, відмінності мовної поведінки чоловіків та жінок. Сюди в першу чергу відносяться особливості використання мовних засобів (лексика, синтаксичні конструкції, мовні звороти), тактика мовної поведінки в різних комунікативних ситуаціях представників чоловічої та жіночої статі.

Щодо особливостей прояву гендера безпосередньо в спілкування за допомогою конкретних мовних засобів, існує цілий ряд гендерно-позначених штампів. Дослідження цих штампів дозволяють встановити ряд закономірностей, щодо мовних розбіжностей чоловічого та жіночого мовлення, або особливості мовленнєвої поведінки обох статей. В процесі порівняльного опису комунікативної поведінки чоловіків і жінок встановлені певні гендерно-маркіровані ознаки:

- речення чоловіків є коротшими за жіночі;
- чоловікам властиво робити граматичні помилки;
- чоловіки вживають більше іменників і прикметників, відповідно, жінки вживають більше дієслів і часток [3, с.].

О.В. Дудоладова характеризує відмінності чоловічого та жіночого мовлення відзначає: «...чоловіки частіше перебивають, більш категоричні в своїх судженнях і оцінці, прагнуть керувати діалогом, чоловіки говорять більше, ніж жінки. Чоловіки, в цілому, набагато частіше вживають абстрактні іменники, а жінки – конкретні (в тому числі власні назви)» [4, с.126].

Цікавим є факт використання чоловіками іменників і прикметників протилежної, жіночої, статі і, відповідно, навпаки. Герцовська Н.О. зазначає, що «жінки вживають більше зменшувальних суфіксів і ввічливих форм, частіше називають партнера по комунікації по імені й у цілому використовують більше контактно-налагоджувючих мовних стратегій». [2, с.8].

«My dearest Cecilia» [9].

«a very dear friend» [9].

«Tenderly, dearest, I breathe thy sweet name...» [8].



«Calvert, honey,' Calliope says» [8].

«Oh dear no," said Mary at once» [14].

Щодо емоційної забарвленості чоловічого і жіночого мовлення, слід відзначити, що для чоловіків характерним є більш раціоналістичні оцінки дійсності та їхня мова є менш емоційною в порівнянні з жіночою. Емоційних і сенсорних оцінок у мові чоловіків менше, бо вони виділяли частіше естетичну, ніж етичну сторону предмета чи явища навколишньої дійсності [2, с.8]. Бесонова О.Л. з цього приводу зазначає: «Жіноче мовлення включає в себе велику кількість емоційно оцінювальної лексики, в той час, як чоловіча оцінювальна лексика частіше стилістично нейтральна. Найчастіше жінки схильні до інтенсифікації, насамперед, позитивної оцінки. Чоловіки частіше і більш впевнено використовують негативну оцінку, включаючи стилістично занижену, лайливу лексику і інвективи; вони набагато частіше використовують сленгові слова і вирази, а також не літературну та ненормативну лексику» [1, с.43]:

«Good heavens! - cried I» [10]

«You're *crazy*, Nick,' he said quickly. '*Crazy as hell*» [7]

«a young man *full of zeal and courage*» [9]

Американська дослідниця Робін Лакоф запропонувала ряд ознак, що відрізняють жіночу мову від чоловічої: ухильні фрази (sort of., kind of...), підсилювальні слова і фрази (so, very), «порожні означення» (divine, adorable), надзвичайно ввічливі форми (Would you mind?..), вибачення, евфемізми з метою уникнення грубої лексики, непрямі прохання (I'm so thirsty), модальні конструкції (can, would, should; Should we...), прямі цитати та своєрідне почуття гумору (жінки не вміють розповідати жарти і часто не розуміють анекдотів) [11 с.56]. Також, А.Кириліна запропонувала ще ряд відмінностей чоловічого та жіночого мовлення. На думку дослідниці, маскулініними ознаками мовлення є: вживання термінологічної та професійної лексики; використання абстрактних понять та негативних конструкцій; вживання емоційно-незабарвлених чи емоційно-нейтральних слів [6, с.53].

«Her tone was peculiar, and she spoke in a *sort of* brooding way as though she were weighing something in her own mind» [10, с.10].

«...*would you* object if I make a suggestion? You wear a *very valuable* ring. I think it is entirely safe, but - who can tell?» [13]

Щодо синтаксичних розбіжностей в чоловічому та жіночому мовленні, Е.І. Горошко зазначає, що в реченнях чоловіки частіше використовують підрядний, рідше сурядний зв'язок; в порівнянні з жіночими реченнями рідше зустрічаються окличні, питальні та неповні речення [3, с. 28-41].

«What do you mean by that? That she was educated, lady-like, pretty, or what?» [10]

Наведені факти говорять про те, преференції у використанні тих чи інших мовних засобів чоловіками і жінками в письмовій мові існують і таким чином розбіжності говорять про існування певної системи факторів, що зумовлюють дані розбіжності. Поняття гендерних розбіжностей у мові чоловіків та жінок яскраво розкривається за допомогою аналізу структур мови. Недоречно говорити про існування постійного набору ознак чоловічої і жіночої мови, але можли-



во встановити визначені статистичні закономірності, властиві чоловічому й жіночому стилям письма.

Список літератури

1. Бесонова О.Л. Оцінний тезаурус англійської мови: когнітивно-гендерні аспекти / О.Л.Бесонова. – Донецьк: ДонНУ, 2002. – 362 с.
2. Герцовська Н. О. Проблема гендеру у сучасній лінгвістиці: чоловіча і жіноча мова. Мукачівський державний університет м. Мукачево. Наукові записки. Серія «Філологічна». Випуск 38. Ст.7-9.
3. Горошко Е. И. Особенности мужского и женского стиля письма / Е. И. Горошко // Гендерный фактор в языке и коммуникации. – Иваново, 1999. – С.28-41.
4. Дудолодова О.В. Гендерные аспекты речевого взаимодействия / О.В.Дудолодова // Функциональная лингвистика. Итоги и перспективы: материалы науч.конф. – Ялта: ТНУ, 2002. – 265с.
5. Каменская О.Л.. Гендергетика - наука будущего // Гендер как интрига познания: гендерные исследования в лингвистике, литературоведении и теории коммуникации. Альманах. Пилотный выпуск. – Москва, 2002. – с.13.
6. Кирилина А. В. Развитие гендерных исследований в лингвистике / А.А.Кирилина // Филологические науки. – 2003. - №5. – с. 51 – 56;
7. F.Scott Fitzgerald. The Great Gatsby [Електронний ресурс]. – Режим доступу: https://ebooks.adelaide.edu.au/f/fitzgerald/f_scott/gatsby/
8. Zona Gale. Friendship Village [Електронний ресурс]. – Режим доступу:<http://www.gutenberg.org/ebooks/26644>
9. Anna Katharine Green. The Mill Mystery [Електронний ресурс]. – Режим доступу: https://ebooks.adelaide.edu.au/g/green/anna_katharine/mill_mystery/
10. Anna Katharine Green. A Strange Disappearance [Електронний ресурс]. – http://ebooks.adelaide.edu.au/g/green/anna_katharine/strange/chapter20.html
11. Lakoff G. Women Fire and Dangerous things / G.Lakoff. – Chicago University, 1990 - 123р. Н.Д.
12. Lawrence. Women in Love. [Електронний ресурс]. – Режим доступу: <http://ebooks.adelaide.edu.au/l/lawrence/dh/l41w/chapter31.html>
13. Mary Roberts Rinehart The Amazing Interlude [Електронний ресурс]. – Режим доступу: <https://www.ibiblio.org/gutenberg>
14. Virginia Wolf. Night And Day. [Електронний ресурс]. – Режим доступу: <https://ebooks.adelaide.edu.au/w/woolf/virginia/w91n/>



АНТРОПОЦЕНТРИЧНИЙ ПІДХІД ДО УКЛАДАННЯ ЕЛЕКТРОННОГО ПЕРЕКЛАДНОГО СЛОВНИКА ТЕХНІЧНОЇ ЛЕКСИКИ

Купріянов Є.В.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків, вул. Пушкінська 79/2, тел. 707-63-60,
e-mail: cuprijanow.eugen@yandex.ua*

Сучасна комп'ютерна лексикографія розвивається у двох напрямках. Перший пов'язаний з упровадженням та пристосуванням підходів традиційної лексикографії в електронному словникарстві з використанням інформаційних технологій, а другий – розв'язанням питань щодо розвитку ідей та принципів, які неможливо здійснити у «паперових» словниках. Однією з таких ідей є розроблення принципів теоретичних принципів та практична реалізація антропоцентричного опису галузевої лексики в електронному словнику, зокрема перекладному. З огляду на це, зауважимо, що головним завданням лінгвоцентричних словників є реєстрація, описання та оцінка наявних мовних фактів. Антропоцентричні словники покликані допомогти користувачеві, по-перше, у формування мови як приналежності до свідомості, а, по-друге, в ефективному використанні мови [1, с. 132].

Ураховуючи сказане вище, антропоцентричні перекладні словники мають містити не лише перелік іншомовних термінів разом з їх перекладами, а й також сприяти засвоєнню та зрозумінню галузевої лексики (як іншомовної, так і національної), розвивати здібності користувача розкривати лексичні, граматичні та стилістичні особливості галузевої лексики в тексті оригіналу, враховувати лінгвальні та екстралінгвальні чинники, що можуть впливати на рішення щодо перекладу речення у цілому та окремого терміна зокрема.

Актуальність розвідки зумовлена потребою ефективного інформаційного забезпечення перекладацької діяльності, підготовки перекладачів в умовах виробництва та навчального процесу та вироблення нових підходів до розвитку лінгвістичної та предметної компетенції. Мета нашої розвідки – розглянути окремі базові питання методики укладання електронних технічних словників антропоцентричного типу. Для досягнення цієї мети необхідно: 1) прокоментувати етапи перекладу та когнітивні процеси, пов'язані з ними; 2) визначити основні види лінгвістичної та енциклопедичної інформації, якими оперує перекладач на різних етапах перекладу; 3) схарактеризувати структуру й зміст електронного технічного словника антропоцентричного типу.

Основна ідея когнітивного аспекту дослідження перекладацької діяльності полягає в необхідності переходу від рівня мови до сфери мислення з метою вивчення ментальних операцій, що визначають розуміння, вибір мовних засобів [2, с. 8]. Зважаючи на це, когнітивна модель процесу перекладу поділяється на три основні етапи: 1) аналіз тексту мовою оригіналу; 2) перекодування тексту



засобами мови перекладу; 3) реалізація тексту мовою перекладу. Під час аналізу та розшифрування вихідного тексту перекладач оперує лінгвістичними та енциклопедичними знаннями, представленими в статичному й динамічному фреймах. Статичні фрейми, як зауважує І. М. Ремхе, є тими структурами, які відбивають сформовану систему знань перекладача, що склалася в процесі його когнітивної діяльності. Динамічний фрейм формується під дією вербальних подразників із тексту шляхом запускання особливої пошукової системи, що виокремлює відповідники між інформацією, закладеною в тексті, і когніцією перекладача, які дають змогу зрозуміти інформацію [2, с. 12].

Статичний фрейм містить термінологічну лексику, знання про предметну галузь. Інформаційними елементами динамічного фрейму можуть бути словесне оточення термінів у контексті, особливості функціонування терміну в конкретному значенні у тій чи іншій предметній сфері тощо. На етапі аналізу вихідного тексту відбувається його усвідомлення, шляхом поділення на мінімальні інформаційні одиниці, ідентифікація серед інформаційних одиниць термінів і термінологічних словосполучень, їх оточень та семантики. На етапі розшифрування виробляється стратегія перекладу змісту тексту, що полягає у встановленні синтаксичних та семантичних особливостей термінів мови оригіналу та виборі відповідних еквівалентів мовою перекладу. На останньому етапі перекладач створює текст мовою перекладу.

Отже, електронний словник складають лінгвістичний та енциклопедичний розділи. У лінгвістичному передбачено режими тлумачення, словотвірних характеристик та ілюстративних контекстів. Режим тлумачення призначений для семантизації заголовкового терміну й містить такі лексикографічні параметри: український еквівалент відповідно до окремого значення англійського терміну; англійська дефініція заголовкового терміну з перекладом українською мовою; англійські видові номінації й усталені словосполучення; синоніми до заголовкового терміну. Режим словотвірних характеристик подає для вхідного терміну: англійські регулярні й контекстуальні словосполучення, їх перекладні відповідники. Режим контекстів репрезентує особливості вживання заголовкового терміну в конкретному значенні. Енциклопедичний розділ словникової статті подає інформацію про предмети або поняття, репрезентовані термінами аналізованої галузі у вигляді: вербального опису предметів або технічних процесів; екстралінгвальні засоби опису, що сприяють не лише розширенню фонових знань про предметну галузь, а й точнішому вибору перекладного еквівалента.

Список літератури

1. Морковкин В. В. Антропоцентрический версус лингвоцентрический поход к лексикографированию // Национальная специфика языка и ее отражение в нормативном словаре / В. В. Морковкин – М., 1988. С. 131–136.
2. Ремхе И. Н. Когнитивные особенности перевода научно-технического текста (на материале текстов металлургической промышленности) : автореф. дис. на соискание ученой степени канд. филол. наук : спец. 10.02.20 «Сравнительно-историческое, типологическое и сопоставительное языкознание» / И. Н. Ремхе. – Челябинск, 2007. – 25 с.



БИМАТРИЧНАЯ ИГРА КАК ОСНОВА ФУНКЦИОНИРОВАНИЯ ПРОИЗВОДСТВЕННОЙ МУЛЬТИАГЕНТНОЙ СИСТЕМЫ

Гончар И.А., Лысенко А.И., Шостак Е.И.

*Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ»,
г.Харьков, ул.Чкалова 17, тел. 788-46-02*

Содержание планирования деятельности предприятия как одной из общих функций управления предполагает разработку программы производства с опережением событий будущего.

Так как точно определить, что произойдет в будущем невозможно, то следствием того или иного образа действий является множество возможных исходов, вероятности которых неизвестны.

Таким образом задача планирования состоит в том, чтобы на основе имеющихся данных прогноза в оценках будущего разработать производственную программу, придерживаясь которой предприятие в наименьшей степени будет уязвимо в будущем с точки зрения некоторого выбранного критерия.

Для решения сформулированной задачи в качестве формализованной модели исследования, как правило, используется математический аппарат теории игр двух лиц, одним из которых является рассматриваемое предприятие, а другим – «внешняя среда» его функционирования. При этом интересы «внешней среды», которая представляет собой совокупность таких элементов как: потенциальные потребители, конкурирующие предприятия, фискальные и контролирующие госучреждения, источники трудовых, финансовых, материальных и информационных ресурсов с одной стороны, и интересы исследуемого предприятия с другой стороны, рассматриваются как антагонистические. Это приводит к рассмотрению матричной игры с нулевой суммой [1], решение которой, исходя из принципа гарантированного результата, сводится к нахождению максимальной стратегии производства исследуемого предприятия.

Предлагается для повышения адекватности используемой формализованной модели считать «внешнюю среду» не антагонистическим оппонентом исследуемого предприятия, а индифферентным участником операции, интересы которого в общем случае не совпадают с интересами отдельного производителя, но не являются абсолютно противоположными. Это приводит к необходимости рассмотрения биматричной игры двух лиц с нулевой суммой [2], решение которой сводится к нахождению так называемой ситуации равновесия, в отступлении от которой не заинтересован ни один из участников операции.

При этом стратегия «внешней среды» представляет собой организацию определенного состояния рынков как используемых ресурсов, так и готовой продукции, а стратегией производителя является выбор определенной программы выпуска продукции.

Реализацию представленной выше задачи предполагается осуществить в форме взаимодействия двух интеллектуальных агентов, входящих в состав мультиагентной системы (МАС). Процесс самоорганизации в мультиагентных



системах — внутрішня упорядоченість, согласованность, взаимодействие более или менее дифференцированных и автономных агентов агентной системы, обусловленной ее строением. Таким образом, в МАС агенты смогут общаться, передавать друг другу некоторую информацию, взаимодействовать между собой и решать поставленную задачу. В такой системе подзадачи распределены между агентами, каждый из которых рассматривается как член группы или организации. Распределение задач предполагает назначение ролей каждому из членов группы, определение меры его "ответственности" и требований к "опыту".

Основной формой организации взаимодействия между агентами, характеризующаяся объединением их усилий для достижения совместной цели при одновременном разделении между ними функций, ролей и обязанностей является кооперация.

В отличие от других известных подходов, этот подход реализует формирование общего мира деятельности кооперирующих сторон и миров деятельности каждой из них, путем создания единой комплексной среды. Эта среда служит основой деятельности менеджеров и специалистов и обеспечивает воспроизводство главных компонент процесса деятельности предприятия и взаимопонимание между людьми. Использование интеллектуальных агентов в комбинации с мирами действий и рассуждений позволяет менеджерам моделировать три "кита" кооперации: коллективное мышление, обоснованное поведение и коммуникацию участвующих сторон. За счет этого каждая из сторон возлагает на своего агента миссию согласования большинства возникающих проблем. Это чаще всего реализуется в виде реализации виртуального круглого стола, предлагающего агентам общий мир действия.

Таким образом предлагаемый подход позволяет планировать производственную программу предприятия в условиях рыночной экономики либо с помощью бескоалиционной игры, в которой запрещаются любые соглашения вроде совместного выбора стратегий и побочных платежей [2], либо с помощью кооперативной игры, в которой такое сотрудничество разрешается. При этом методологической основой для программной реализации рассматриваемого подхода является агентная парадигма в форме системы МАС, включающей два интеллектуальных агента.

Список литературы

1. Математичні моделі планування виробництва в аерокосмічній галузі: нав. посіб. / І.В.Чумаченко, О.І.Лисенко, І.А.Скачкова та ін. — Х.: Нац. аерокосм. ун-т «ХАІ», 2012. — 272 с.
2. Воробъёв Н.Н. Основы теории игр. бескоалиционные игры / Н.Н. Воробъёв. — М.: Наука, 1984. — 496.



ОЦІНКА ЕФЕКТИВНОСТІ ВИДОБУВАННЯ ТЕРМІНІВ ПРЕДМЕТНОЇ ОБЛАСТІ З ТЕКСТІВ

Борисова Н.В.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: borisova_nv@mail.ru*

Відповідно до міждержавного стандарту з інформації, бібліотечної та видавничої справи для оцінки ефективності видобування термінів використовуються коефіцієнт точності *Precision*, коефіцієнт повноти *Recall*, коефіцієнт шуму *Fallout*, помилка видобування *Error*.

При проведенні випробувань щодо видобування термінів з текстів предметної області «екологія» інформаційною системою автоматизованого формування лексикографічних ресурсів було проаналізовано близько 400 електронних документів цієї предметної області.

Для проведення розрахунків використовувалися такі параметри:

- загальна кількість термінів предметної області, які містяться в аналізованих документах, відповідно до висновку експерта R ;
- кількість термінів-кандидатів, видобутих системою, A ;
- кількість термінів-кандидатів, видобутих системою, які дійсно є термінами відповідно до висновків експерта, R_a ;
- загальна кількість термінів, які містяться в аналізованих документах, відповідно до висновку експерта, T ;

Також було виділено чотири показника по відношенню до включення у перелік видобутих термінів:

$a = R_a$ – кількість видобутих термінів-кандидатів, які є термінами предметної області;

$b = A - R_a$ – кількість видобутих термінів-кандидатів, які не є термінами предметної області;

$c = R - R_a$ – кількість не видобутих термінів-кандидатів, які є термінами предметної області;

$d = (T - R) - (A - R_a)$ – кількість не видобутих термінів-кандидатів, які не є термінами предметної області.

Тоді коефіцієнт точності видобування термінів визначається за формулою

$$Precision = \frac{R_a}{A} = \frac{a}{a+b}.$$

Коефіцієнт повноти видобування визначаємо як

$$Recall = \frac{R_a}{R} = \frac{a}{a+c}.$$



Коефіцієнт шуму системи визначаємо як

$$Fallout = \frac{A - R_a}{A} = \frac{b}{a + b}.$$

Помилка видобування розраховується за формулою

$$Error = \frac{(A - R_a) + (R - R_a)}{T} = \frac{b + c}{(a + b + c + d)}.$$

Для оцінки ефективності вирішення задачі видобування термінів з текстів були отримані такі кількісні значення параметрів: $R = 4037$, $A = 3822$, $R_a = 3593$, $T = 22447$. Відповідно коефіцієнти дорівнюють:

$$Precision = \frac{3593}{3822} = 0,94;$$

$$Recall = \frac{3593}{4037} = 0,89;$$

$$Fallout = \frac{3822 - 3593}{3822} = 0,06;$$

$$Error = \frac{(3822 - 3593) + (4037 - 3593)}{22447} = 0,03.$$

Порівняння отриманих результатів роботи розроблено інформаційної системи автоматизованого формування лексикографічних ресурсів з результатами роботи подібних інформаційних систем за коефіцієнтами точності та повноти представлено у таблиці 1. Результати для інших систем отримано з науково-технічних джерел.

Таблиця 1 – Порівняння ефективності роботи розробленої системи з системами, що вирішують аналогічні задачі

Інформаційні системи, що порівнюються	Коефіцієнт точності	Коефіцієнт повноти
Розроблена	0,94	0,89
Заснована на LSPL-шаблонах	0,91	0,85
Заснована на методах, що використовують механістичні критерії	0,75	0,69

Результати порівняння показали ефективність роботи розробленої інформаційної системи щодо вирішення розглянутої задачі.

Таким чином, можна стверджувати, що інформаційна система задовольняє потреби користувача щодо видобування термінів з текстів предметної області.



ВЫСОКОУРОВНЕВЫЙ ЯЗЫК ЗАПРОСОВ К БАЗЕ ДАННЫХ НА ОСНОВЕ «КОННЕКТОМА» КОГНИТИВНОЙ СИСТЕМЫ

Прокопчук Ю.А.

Институт технической механики НАНУ и ГКАУ, Днепропетровский областной центр кардиохирургии, г.Днепропетровск, E-mail: itk3@ukr.net

Вступление. Термин «коннектом» родился в 2005 году. Им обозначили полное описание структуры связей в нервной системе человека [1]. С тех пор коннектомикой называют компьютерный анализ строения естественных нейронных сетей, своего рода картографирование нейронных связей. Коннектом объясняет, почему мозг обладает такой потрясающей автоассоциативностью. Многие когнитивные функции есть производные от его сетевой топологии. Концепция коннектома позволяет выявить ключевые свойства естественных нервных сетей, важные для создания искусственных когнитивных систем (ИКС). Возникает проблема инженерии «ИКС-коннектомов». Задачей исследования является реализация высокоуровневого языка запросов к БД, используя «ИКС-коннектом» [2].

Результаты исследования. Предполагается, что произвольную ситуацию действительности (прецедент, систему, процесс) можно описать с помощью множества элементарных тестов (квалиа), представимых в виде «тест = значение». Результаты любого теста τ могут выбираться из разных доменов T (множеств значений), которые образуют оргграф доменов $G(\tau) = \{T \rightarrow T\}_\tau$. В совокупности оргграфы формируют *Банк тестов*. Для решения той или иной когнитивной задачи (Z -задачи) формируется множество прецедентов с известными исходами $\Omega = \{\alpha(\{z/T\}, z/Z)\}$, где $Z = \{1, \dots, N\}$ – множество заключений (различий, диагнозов, прогнозов, управлений); $\{z/T\}$ – множество значений тестов. Множество Z является одним из доменов оргграфа $G(z)$. Z -задачи вместе с другими тестами формализуют акты различения (дифференциации). *Акт различения* – это системоквант «мысле-действия» когнитивной системы, базовая функция наблюдателя.

Поскольку любой дискретный домен любого теста может образовать Z -задачу различения, для которой имплицитно формируется базис собственных форм (моделей знаний), то оргграф доменов теста τ можно представить в виде:

$$G(\tau) = \{T \{ \{V\}_T \} \rightarrow_e T \{ \{V\}_T \} \}_\tau \quad (1)$$

где $\{ \{V\}_T \}$ – инвариантный базис моделей знаний T -задачи, основанный на всех тестах (квалиа) банка $\{G(\tau)\}$ и базе прецедентов $\Omega(\alpha(\{p\}, T))$. Примеры лингвистических оргграфов доменов:

Уровень гемоглобина $\wedge Hb$ (Пол?) $\{D4 \{норма \wedge^0; отклонение \wedge^1 4\} \{S\}_{D4}\} D3$
 $\{анемия \wedge^1 2 3; норма \wedge^0; повышен \wedge^4 5 6\} \{S\}_{D3}\} D2$ {Резко выраж. анемия \wedge^3
 М [80; 110) Ж [80; 102); Выраж. анемия \wedge^2 М [110; 120) Ж [102; 110);
 Умеренная анемия \wedge^1 М [120; 133) Ж [110; 114); Норм.уровень $Hb \wedge^0$
 М[133;185] Ж [114;168]; Уровень Hb умеренно пов. \wedge^4 М (185; 190] Ж (168;
 170]; Повыш.уровень $Hb \wedge^5$ М (190; 200] Ж (170; 172]; Резко повыш. уровень

Hb \wedge^6 M (200; 250] Ж (172; 250] $\{\{S\}_{D2}\}$ D1 $\{[80; 250]\}$

Степень выраженности воспалительного процесса \wedge^{CBVP} $\{$

D2 {в.п. отсутствует \wedge^1 ; в.п. имеется \wedge^2 3 4 } $\{\{V\}_{D2}\}$

D1 {в.п. отсутствует \wedge^1 ; незначительная \wedge^2 ; средняя \wedge^3 ;
выраженная \wedge^4 } $\{\{S\}_{D1}\}$

Тест «CBVP» определяется на основе анализа крови, включающего тест «Hb» (рис.1, иллюстрирует автоматическую генерацию новых Z-задач-доменов), и в первичных данных не содержится. Инвариантные модели знаний $\{\{V\}_{D2}\}$ и $\{\{S\}_{D1}\}$ формируются автоматически и имплицитно на основе процессов авто/гетеро-ассоциативности в рамках банка тестов $\{G(\tau)\}$.

Фиксация орграфа доменов $G(\tau)$ – это фактически постановки $|G(\tau)|$ задач (кроме непрерывного домена). Таким образом, $\{G(\tau)\}$ формирует транзитивно-индукторное пространство или основу ИКС-коннектома. Полный ИКС-коннектом формируется с учетом орграфов набросков образов $G_S(W) = \{P \rightarrow_e P'\}_W$ [2] или в развернутом виде:

$$G_S(W) = \{P \{\{V\}_P\} \rightarrow_e P' \{\{V\}_{P'}\}\}_W, \quad (2)$$

где P, P' - наброски образа W , $\{\{V\}_P\}$ – инвариантные модели знаний (собственные формы).

Опишем процесс автоматического и имплицитного формирования базиса моделей знаний $\{\{V\}_T\}$ на основе банка тестов $\{G(\tau)\}$. Пусть имеется кадр данных $\{\tau/T_0\}$ и банк тестов $\{G(\tau)\}$. Без потери общности примем, что каждый тест входит в кадр не более одного раза. Поток данных представляет собой текущую совокупность всех кадров $\cup_i \{\tau/T_0\}_i$. Пример потока данных в рамках госпитальной системы (ГИС) показан на рис. 2.

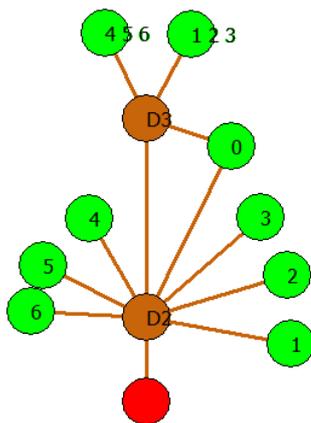


Рис. 1 – G^{++} – Орграф доменов теста «Hb»

Дата	Диагноз	Специалист
10.01.15	УВАН РЕЧЕВНО	Колупа О.В.
10.02.15	СВЕРХ АНАЛИЗ МОЗГ	Ермоленко А.В.
10.03.15	Диагноз	Колупа О.В.
11.01.15	Колупа О.В.	Ермоленко А.В.
11.02.15	УЛЬТРАЗВУКОВЫЙ КОМПЛЕКС	Ермоленко А.В.
11.03.15	СОННОГРАФИЯ КОМПЛЕКС	Ермоленко А.В.
11.04.15	КАРДИОГРАФИЯ КОМПЛЕКС	Ермоленко А.В.
11.05.15	ЭКГ РЕГЕНЕРАЦИЯ И МЕРЦАТЕЛЬНЫЕ	Ермоленко А.В.
12.01.15	УЛЬТРАЗВУК КРОВИ	Ермоленко А.В.
14.01.15	СВЕРХ АНАЛИЗ КРОВИ	Ермоленко А.В.
16.02.15	Диагноз	Артемюк В.В.
17.02.15	ДИАГ	Колупа О.В.
18.02.15	Диагноз	Колупа О.В.
18.03.15	Диагноз дисперсного зрения	Молодцова О.В.
17.03.15	Ультразвуковой сканер	Скороход А.В.
18.01.15	Диагноз	Колупа О.В.
08.03.15	Диагноз	Скороход А.В.
11.04.15	Анализ крови комплексный сканер	Скороход А.В.
10.01.15	Диагноз	Колупа О.В.
07.01.15	Диагноз дисперсного зрения	Молодцова О.В.
08.02.15	Ультразвуковой сканер	Колупа О.В.
04.03.15	Диагноз	Колупа О.В.
04.05.15	Ультразвуковой сканер	Колупа О.В.
12.04.15	Диагноз	Колупа О.В.
08.01.15	Диагноз	Артемюк В.В.
11.02.15	Спектральный анализ крови сканер	Колупа О.В.
11.04.15	Спектральный комплексный сканер	Колупа О.В.

Рис. 2 – База данных ГИС: поток первичных данных

Имплицитное обучение без учителя на основе потока данных с кадром $\{\tau/T_0\}$ означает следующее: 1) фиксируется текущее множество данных; 2) фиксируется произвольный тест $z \in \{\tau\}$; 3) фиксируется произвольный домен Z из $G(z)$; 4) все данные z -теста в потоке $\cup_i \{\tau/T_0\}_i$ преобразуются к домену Z , следовательно, возникает Z -задача с базой прецедентов $\Omega = \cup_i \{\tau/T_0\}_i$; 5) для воз-

никшей Z -задачи формуються базиси моделей знань [2]. Процедура повторюється для всіх $z \in \{\tau\}$ і всіх доменів Z із $G(z)$. Після додавання нового кадра весь процес повторюється заново. В результаті оновлюються всі авто/гетеро-асоціативні моделі знань «ІКС-коннектома» (рис. 3).

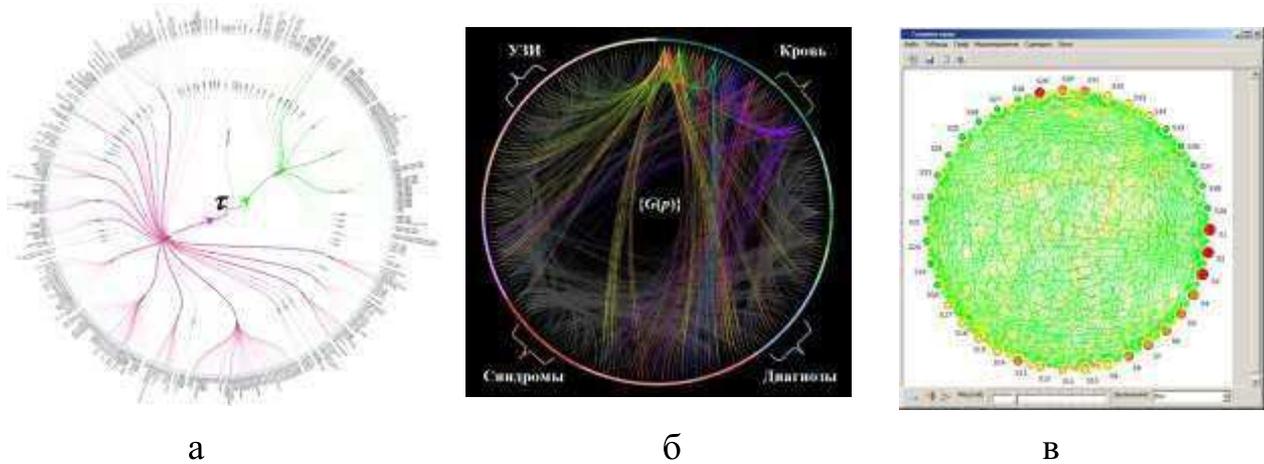


Рис. 3 – Ілюстрації ІКС-коннектома:

- а) конус причинності і конус ефекта произвольного теста τ ; б) ІКС-коннектома госпитальной інформаційної системи; в) Z -коннектома.

Возбуждение части тестов через тисячі зв'язей з іншими тестами ІКС-коннектома спроможно почти моментально вызвать богатейшую сеть ассоциаций. ІКС-коннектома, как и био-коннектома, иллюстрирует авто/гетеро-асоціативность, распределенность, робастность и активность памяти. Задачи «вшиты» в «Коннектома», поэтому решения возникают автоматически (аналог «подсознательной» активности мозга). ІКС-коннектома – это фракталоподобная, постоянно растущая и изменяющаяся структура. Он задает первичную структуризацию ПрО (семантическую сеть предметной области). Можно говорить о глобальном ІКС-коннектома, ПрО-коннектома, Z -коннектома (рис. 1в). В совокупности они образуют масштабируемую систему опыта конкретной ІКС (когнитивного агента).

Запросы. ІКС-коннектома позволяет создавать когнитивные надстройки над СУБД, включая: семантический язык запросов, интерактивный транслятор с естественного языка в SQL/XML (семантический и лексический процессоры [2]); встроенные индуктивные и дедуктивные системы. Подобный инструментарий позволяет преодолеть семантический разрыв между моделями создаваемыми специалистом предметной области и программистом. Примеры запросов: «СВВП/D2?» «Дать послеоперационный прогноз осложнений».

Список литературы

1. Сеунг С. Коннектома. Как мозг делает нас тем, что мы есть. Перевод с англ. – М.: Бином, 2014. - 440 с.
2. Прокопчук Ю.А. Парадигма предельных обобщений: модели когнитивных архитектур и процессов. - Saarbrücken, Deutschland: LAP LAMBERT Academic Publishing, 2014. – 204 с. ISBN 978-3-659-66571-4.



ИНТЕЛЛЕКТУАЛЬНАЯ ОБРАБОТКА ДАННЫХ О СОСТОЯНИИ ПРИРОДНО-ТЕХНОГЕННЫХ ОБЪЕКТОВ

Козуля М.М.

*Национальный технический университет
«Харьковский политехнический институт», Харьков, Украина,
e-mail: mariya_kozulya@mail.ru*

Актуальность задачи исследования определена необходимостью повышения эффективности обработки информации в системе комплексного мониторинга окружающей среды, что отмечено в Концепции экологической политики Украины, рассчитанной до 2020 года.

Целью работы является повышение эффективности применяемых информационных технологий интеллектуальной обработки данных и качества представления таких результатов для принятия решений урегулирования социально-экологических отношений. За основу достижения цели принята комплексная система моделей и методов идентификации знаний.

В качестве практического примера комплексирования методов по оценке состояния объекта исследования рассмотрена обработка данных мониторинга загрязнения почв и состояния здоровья на исследуемых техногенно-нагруженных территориях. Для обработки исходных данных предложено последовательный анализ риск-оценки состояний и компараторной идентификации соответствий требованиям безопасности.

Оценка экологичности систем исследования включает:

- согласование результатов анализа состояния реального объекта на множестве данных экспериментальной выборки и структурно-параметрической идентификации по данным его функциональности;
- процедуру проверки соответствия состояния, процессов в системах объекта требованиям безопасности на сравнении адекватности параметров модели комплексной оценки экологичности систем характеристикам природного гомеостаза – компараторная идентификация экологичности [1–2].

Для идентификации соответствия использован экологический компаратор – измеритель выполнения требований по структуре, функциональности процессов внутреннего регулирования гомеостаза исследуемого объекта [3]. Компараторная идентификация для территориально-объектных систем связана с оценкой оператора структурной идентификации и значением параметрической идентификации. Это позволяет получить отображение реальных физических, физиологических, биохимических процессов, связанных с нарушением экологичности в системах, с целью выявления управляющего действия [1].

Реализация методики компараторной идентификации качества объектов рассмотрена на примере исследования экологического состояния почв территории Змиевского района, которые находятся под влиянием промышленно-энергетического комплекса, характеризуются определенным уровнем загрязнения тяжелыми металлами. Результирующая оценка безопасности и степени

екологічності дослідженої території зв'язана з ризик-аналізом стану здоров'я населення (рис. 1).

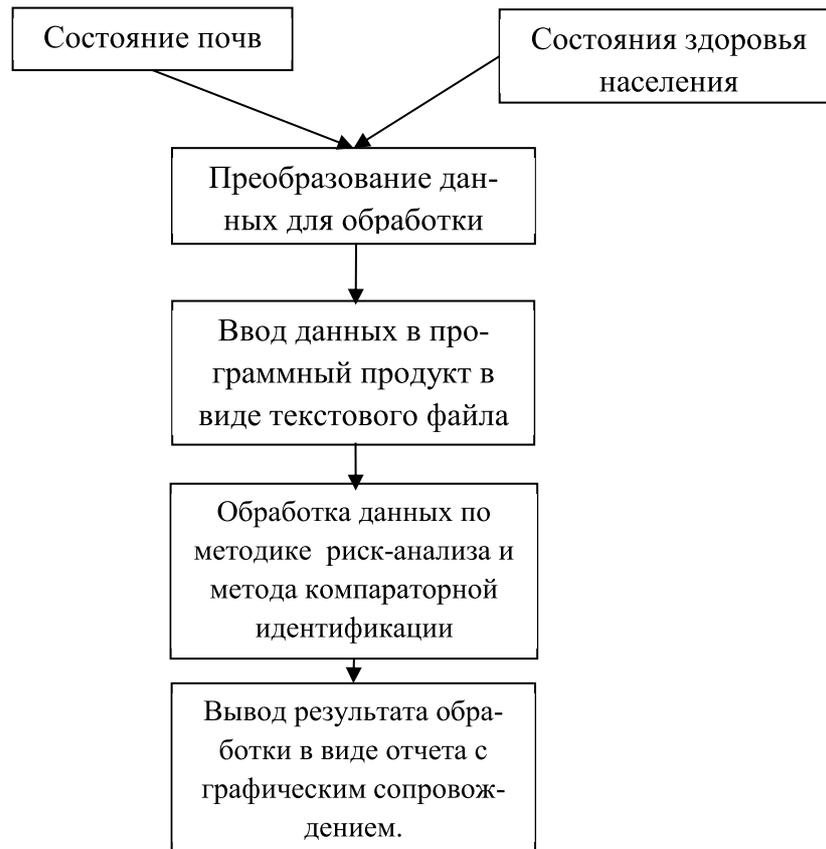


Рисунок 1 – Алгоритм комплексной обработки данных

Предложенный подход интеллектуальной обработки входной информации дает возможность получить комплекс знаний о состоянии системы и ее составляющих, дать оценку соответствия определенным требованиям, определить основной фактор дестабилизации системы, что является основой принятия управленческого решения для нормализации состояния исследуемых систем.

Таким образом, в работе предложена и реализована на конкретной практической задаче система комплексной интеллектуальной обработки данных, определены пути ее усовершенствования для применения в экологической сфере познания окружающей среды.

Литература:

1. Козуля Т. В. Моделирование структуры и идентификация состояния корпоративной экологической системы (КЭС) / Т. В. Козуля, Н. В. Шаронова // Проблеми інформаційних технологій. – 2007. – № 01 (001). – С. 178–187.

2. Козуля Т. В. Определение структуры корпоративной экологической системы и идентификации ее состояния / Т. В. Козуля, Н. В. Шаронова // Системні дослідження та інформаційні технології. – 2009. – № 3. – С. 117–129.

3. Бондаренко М. Ф. Про загальну теорію компараторної ідентифікації / М. Ф. Бондаренко, С. Ю. Шабанов-Кушнарченко, Ю. П. Шабанов-Кушнарченко // Біоніка інтелекту: наук.-техн. журнал. – 2008. – № 2 (69). – С. 13–22.



СУЧАСНІ ТЕНДЕНЦІЇ СТВОРЕННЯ ТЕРМІНОЛОГІЧНИХ ПЕРЕКЛАДНИХ СЛОВНИКІВ (НА ПРИКЛАДІ СТВОРЕННЯ ЕЛЕКТРОННОГО СЛОВНИКУ У ГАЛУЗІ GPS SOFT У ТОПОГРАФІЇ ТА КАРТОГРАФІЇ)

Кудоярова О.В., Жукова К.А.

*Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»
м. Харків, вул. Чкалова 17, тел. 788-48-71
e-mail: olga_kudoyarova@mail.ru, kristya_zhukova@mail.ru*

У зв'язку з науково-технічним прогресом, зростаючим підвищенням рівня знань, загальною інтелектуалізацією мови можна спостерігати стрімкий розвиток термінології. Це супроводжується необхідністю номінації нових термінів. Як наслідок, укладається велика кількість словників. У науковій діяльності вони відіграють важливу роль, відображаючи тезауруси окремих сфер знань. Так, наприклад, сучасний етап розвитку GPS Soft у топографії та картографії характеризується широким впровадженням термінів в процесі створення нової продукції або розробки нових технологій. Відсутність перекладних словників вузькоспеціальної термінології цієї галузі з належним тлумаченням та прикладами вживання зумовлюють актуальність створення словнику такого типу.

Система глобального позиціонування (GPS) представляє собою сукупність радіоелектронних засобів, що дозволяє визначати положення та швидкість руху об'єкта на поверхні Землі або в атмосфері. Сьогодні системи GPS використовуються у багатьох сферах життя.

Слід зазначити, що серед технічної термінологічної лексики цієї галузі можна зустріти не тільки прості, але й складні терміни та термін-словосполучення. В якості окремого засобу категоризації таких термінів можна виділити аббревіатури. В наш час аббревіація стала одним із найбільш розповсюджених засобів створення номінативних лексичних одиниць, оскільки вони допомагають скоротити до мінімуму мовні зусилля, мовні засоби й час, залишаючись при цьому зрозумілими за умови їх належного трактування.

Нашою метою було укладання словника-довідника, для того щоб надати широкому колу користувачів, які працюють з картографічними матеріалами та займаються розробкою GPS технологій, а також студентам, викладачам та перекладачам технічної літератури короткі відомості щодо основних термінів у галузі GPS Soft у картографії та топографії з можливістю їх подальшого вивчення за допомогою додатку у вигляді комплексу наукових матеріалів, створених на основі програми Lingvo Tutor.

Робота по створенню словнику проводилась наступним чином: в першу чергу був визначений матеріал нашого дослідження. В даний словник були включені лексичні одиниці термінологічного характеру, серед яких зустрічаються одно-, дво- і багатокомпонентні терміни, аббревіатури, які належать до різних тематик, таких як інструменти та вимірювальні прилади, картографічні ві-



деоефекти, види карт, небесна механіка, позиційна астрономія та ін.. Джерелом термінологічного матеріалу слугували державні стандарти із спеціалізованою лексикою (ISO), лексикографічні та довідкові видання, глосарії, іноземні статті, в яких висвітлювались проблеми картографії, топографії та GPS технології. Наступним етапом була розробка мікроструктури словника, яка узгоджувалась з основними принципами укладання термінологічного словника-довідника. Беручи за основу праці науковців, таких як, С. В. Гріньова, А. С. Герда, Л. Т. Борисової та ін., основними компонентами словникової статті стали: заголовне слово, формальна характеристика у вигляді транскрипції, семантичний опис слова за допомогою дефініції, контекстуального та ілюстративного визначення. В результаті був отриманий словник, який включає в себе 580 термінів, розміщених в алфавітному порядку. Таким чином, ми створили термінологічний перекладний словник, який представляє собою за способом пояснення слів енциклопедичне видання, за характером описуваних одиниць є термінологічним, за тематичним охопленням — галузевим, за способом упорядкування словника — алфавітним.

Запорукою успішного освоєння будь-якої навчальної дисципліни є оволодіння студентом спеціальною науковою термінологією. Найважливішу роль у цьому процесі відіграє не тільки навчально-методична та навчально-довідкова література, а й комплект навчальних матеріалів. Ми розробили додаток до термінологічного перекладного словника за допомогою програми Lingvo Tutor, яка призначена для навчання іноземної мови. Вона входить в комплект електронного словника АBBYY Lingvo. Величезна перевага цього модуля в тому, що він інтегрується зі словником. Таким чином, надається можливість швидкого створення словників.

Беручи до уваги розроблений нами додаток, можна сказати, що він представляє собою електронні тематичні словники, за допомогою яких можна легко вивчити досліджувану вузькогалузеву лексику. Процес навчання на основі цього модуля відбувається за допомогою виконання 5 видів вправ для запам'ятовування лексики, таких як: «знайомство», «мозаїка», «варіанти», «написання слова» та «самоперевірка». Також, слід зазначити, що в програмі є можливість обирати необхідну кількість слів, кількість повторень для вивчення; задавати час для автоматичного нагадування про урок; налаштовувати спосіб навчання за допомогою прямого, зворотного перекладу або того й іншого разом.

Підсумовуючи сказане, ми вважаємо, що укладений термінологічний словник стане у нагоді багатьом користувачам в процесі пошуку відповідників термінології галузі GPS Soft у картографії та топографії, сприятиме нормалізації української та російської термінології, правильному вживанню термінів як в усному, так і в писемному мовленні, а також поліпшенню знань самих користувачів, які мають можливість читати визначення до терміну та орієнтуватися на приклад його вживання. Підготовлений комплект навчальних матеріалів, в свою чергу, стане корисним помічником в процесі освоєння навчальної дисципліни обраної галузі.



LEARNING GRAMMAR ACTIVITIES TO STUDENTS OF LINGUISTIC DEPARTMENTS

Lutai N.V., Konovalova V.B.

National Technical University

"Kharkiv Politechnic Institute"

Kharkiv city, 79/2, Pushkinskaya str., tel. 707-63-60,

e-mail: lutain@rambler.ru

The learning of grammar should be seen in the long term as one of the means of acquiring a thorough mastery of the language as a whole, not as an end in itself. Thus, although at an early stage we may ask our students to learn a certain structure through exercises that concentrate on virtually meaningless manipulations of language; we should quickly progress to activities that use it meaningfully. And even these activities will be superseded eventually by general fluency practice, where the emphasis is on successful communication, and any learning of grammar takes place only as incidental to this main objective.

Before any classroom exercise you usually make some brief comments to introduce it. The most important thing here is clarity: as a result of the introduction, the students should know exactly what the objectives of the activity are, and how they are expected to achieve them (in the students' native language).

For more mature learners it is a good idea to let them know more precisely what you are going to practise with them: for one thing it helps them and you to feel that there is a sharing of the responsibility for learning; for another, they are more likely to make an effort if they know exactly what it is for. If the class is to do any kind of independent (individual, group or pair) work in the process of the activity, it is vital for the instructions to be clearly given before they start. This is the weak point of many inexperienced teachers: they give instructions that are clear to themselves, and then launch into the activity without checking that the students are sure what they have to do. The result is very often that the teacher has to stop the activity in the middle to reissue instructions, or that there is delay and a constant distracting buzz of talk as students consult each other.

There are various ways of making sure that instructions are clear: by slowing down delivery, repeating, and/or using the students' native language; by doing a 'trial run', or demonstration of an activity with the full class before letting them work independently; by simply asking them, before setting them to work, if there is any unclear point they would like to ask about.

The instructions for an activity based on independent (individual, group or pair) work, incidentally, should usually include some provision for ending: how long the activity is expected to last, what the students should do after they finish, what happens if some finish early or late, what is to be done with any written or recorded results.

The place of a grammar-practice activity is preferably in the middle of a lesson rather than right at the beginning or at the end. Students are freshest and most recep-



tive at the beginning of the lesson: this is therefore the best time to present new language topics or texts, or to re-present difficult material. More extensive fluency practice tends to come later in the period. The end of the lesson we like to leave for 'rounding-off': reviewing what we have done, checking that everyone knows what is to be prepared at home, possibly a brief 'lightweight' activity involving not too much effort of concentration to give the session a pleasant finish.

There are various ways of extending practice activities in order to provide the extra repetition; here are some of them.

i) Repeating in a different mode. An exercise that has been done orally can then be redone in writing, or vice versa.

ii) Repeating selectively. Some bits of the exercise can be selected for review; this lessens the boredom.

iii) Recalling. After an oral practice activity, students can be challenged to recall as much as they can of what was said. If, for example, a conventional discrete-item exercise was used, they can be asked to shut their books and try to remember all the items; if a brainstorm, when a large number of utterances have been heard, they may be asked to try to recall a defined number of sentences: 'Can you remember (at least) ten suggestions that have been made?' The recalling can be done orally or in writing, individually or in pairs or groups.

iv) Editing. After a written activity, students can get together and go through the exercise again in order to help each other correct and improve their work. If the exercise was not open-ended, groups of students can pool their efforts to produce a final joint version as correct as possible. If open-ended - that is to say, if each student has something different, as in essay-writing - then students can read their texts to each other, and exchange comments and advice.

v) Composing. Students can be invited to compose their own exercise as a continuation of a textbook one, using the same kind of texts and task. They can then ask each other, or the entire class, to perform it. You can add further incentive by inviting students to improve on the interest, humour or drama of the original, or to adapt it to refer to their own circumstances or personalities.

vi) Varying. In 'matching' or slot-filling exercises, students are often asked to put together two components in order to form logical propositions. As a follow-up, you might suggest that they match incongruous components to make nonsense or humorous propositions (but still grammatical!) - and then possibly justify them or suggest situations where they might be true. Or they may simply invent their own 'fillers' to make original, personal, or amusing sentences.

Reference:

1. *Carrier M. and the Center for British Teachers. Games and Activities for the Language Learner.* – Harrap, 1983.
2. *Franck C., Rinvolucri M. Grammar in Action.* – Pergamon, 1983.
3. *Jones I. Use of English.* – Cambridge University Press, 1985.



КЛАСИФІКАЦІЯ ТЕКСТІВ ЕКОНОМІЧНОГО НАПРЯМКУ НА ОСНОВІ СЕМАНТИЧНИХ МЕРЕЖ ГІПОНІМІЧНИХ ВІДНОШЕНЬ

Булатнікова Т.С.

*Національний Технічний Університет
«Харківський Політехнічний Інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 0995692641,
e-mail: tbu2641@gmail.com*

Семантична мережа – це інформаційна модель предметної області, що має вигляд орієнтованого графа, вершини якого відповідають об'єктам предметної області, а ребра задають відношення між ними. Об'єктами можуть бути поняття, події, властивості та процеси.

Семантичні мережі відносяться до моделей класичного представлення знань у задачах штучного інтелекту, навчальних системах, системах машинного перекладу та семантичних павутинах. Підхід семантичних мереж базується на трьох основних складових: суб'єкт, відношення, об'єкт. Саме ці три складові є базовими блоками семантичних мереж.

Мережеві моделі формально можна задати у вигляді $H = \langle I, C1, C2, \dots, Cn, R \rangle$, де I – множина інформаційних одиниць; $C1, C2, \dots, Cn$, – множина типів зв'язків між інформаційними одиницями. Відображення R задає між інформаційними одиницями, що входять до I , зв'язки із заданого набору типів зв'язків [1].

Гіпонімія як родо-видове відношення – це сукупність семантично однорідних одиниць, які належать до одного класу. На основі гіпонімії лексичні одиниці об'єднуються в тематичні й лексико-семантичні групи і поля.

Сьогодні гіпонімічні відношення широко використовуються для класифікації компаній за галуззю виробництва. Класифікація за галуззю виробництва впорядковує компанії у виробничі групи, які основані на схожості способів виробництва, продукції або поведінки на фінансовому ринку. Такі угруповання широко використовуються статистичними агенціями у фінансовій сфері послуг для групування схожих інвестиційних компаній при створенні індексів фінансового ринку за секторами [2].

Класифікація текстів – одне із завдань інформаційного пошуку, яке полягає у віднесенні тексту до однієї з кількох категорій.

Класифікація може здійснюватися повністю вручну, автоматизовано за допомогою створеного вручну набору правил або автоматично із застосуванням методів машинного навчання. До автоматичних методів класифікації текстів належать:

- EM-алгоритм;
- наївний байєсівський класифікатор;
- tf-idf;
- латентно-семантичний аналіз;
- штучна нейронна мережа;



- метод k найближчих сусідів;
- дерево прийняття рішень;
- тощо[3].

Класифікація текстів на основі семантичної мережі полягає в тому, що ключові слова, за допомогою яких визначатиметься приналежність тексту до певної теми, та самі теми знаходяться в безпосередньому зв'язку одне з одним. Така класифікація дозволить розширити можливості систем автоматичної обробки англійських текстів економічної тематики, інформаційного пошуку, систем для вилучення інформації, систем аналізу тональності текстів, систем вилучення інформації, систем автоматичного реферування тощо.

Розроблена семантична мережа зберігається у форматі JSON. Такий формат є універсальним ще й тому, що майже всі сучасні мови програмування підтримують роботу з даними, які зберігаються в цьому форматі.

В якості вхідних даних програма бере обраний користувачем файл, в якому збережена потрібна кількість статей новин економічної тематики або вводять їх вручну у відповідне поле. Після того, як програма отримала текстові дані виконується наступний алгоритм:

1. Текст розбивається на статті відповідно до вказаного роздільника та створюється список.

2. Виконується пошук на виявлення економічних понять, якщо таке поняття було знайдено то стаття заноситься в новий список разом із співвіднесеним поняттям, його підгрупою та групою, інакше стаття заноситься в інший список.

3. Виконується перевірка на наявність однієї і тієї ж статті з різними поняттями, якщо такі знайдено, то виконується їх групування.

4. Кожна стаття перевіряється на наявність назви компанії, якщо компанію було знайдено то до статті «прикріплюється» назва компанії та місце її розташування.

5. Якщо компанію не знайдено, то стаття перевіряється на наявність країни, якщо країну знайдено, то статті присвоюється назва країни.

6. Якщо в статті не було знайдено ні країни ні компанії, то стаття заноситься в загальний список без них.

7. В результаті з'являється нове вікно з класифікованими статтями.

Розроблене програмне забезпечення класифікує статті економічного напрямку за тематиками, разом із тематикою зі статті вилучається компанія або країна, назва якої зустрічається в статті, а також для компанії визначається місце розташування її головного офісу. Результат роботи програми користувач може зберегти у файлі, відредагувати або просто подивитися.

Список літератури

1. Roussopoulos N.D. A semantic network model of data bases. – Department of Computer Science, University of Toronto, 1976. – p. 104.
2. Day, A.C.L. The taxonomic approach to the study of economic policies. – The American Economic Review, 2010. – p. 78
3. Hastie, T. The Elements of Statistical Learning. Springer, 2001. – С.758



АЛГОРИТМ СИСТЕМИ АВТОМАТИЗОВАНОГО РЕФЕРУВАННЯ КОРПУСУ ТЕКСТІВ

Дашкевич О.С.

*Національний технічний університет "Харківський політехнічний інститут",
м. Харків, вул. Пушкінська, 79/2, e-mail: esdashkevich@gmail.com*

Метою роботи є дослідження і створення алгоритму системи автоматизованого реферування корпусу англomовних текстів.

Корпус текстів – це вид корпусу даних, одиницями якого є тексти або їх достатньо значні фрагменти, що включають, наприклад, уривки текстів даної проблемної області.

За останнє десятиліття корпусна лінгвістика швидко прийняла той досвід, який був накопичений в інших достатньо розвинених галузях, що об'єднуються терміном АОТ (автоматична обробка тексту), або ж АРТ (автоматичне розуміння тексту), корпусна лінгвістика сьогодні має дві лінії розвитку – лінгвістичний аналіз тексту і інформаційний аналіз тексту.

Інтелектуальний аналіз даних (ІАД, Data Mining), або добування даних – термін, що застосовується для опису здобуття знань у базах даних, дослідження даних, обробки зразків даних, очищення і збору даних. Це процес виявлення кореляції, тенденцій, шаблонів, зв'язків і категорій.

Пропонується декілька методів обробки повнотекстової інформації. На сьогодні існує чітка класифікація лінгвістичних методів та розбиття існуючих підходів на класи еквівалентності для створення єдиної термінології типових лінгвістичних шаблонів. Створення шкали таких методів дозволило чітко відрізнити одні підходи від інших. В неї входять також статистичні методи, як: тематичний аналіз, реферування, авторубрикація та класифікація.

На сучасному рівні розвитку інформаційних технологій з'явилася задача автоматизованого реферування.

Автоматизоване реферування або квазіреферування – це виявлення в тексті первинного документа фрагментів, що містять заздалегідь заявлені змістові аспекти. Найвищого розвитку формалізація методів реферування набула з автоматизацією цього виду аналітико-синтетичної обробки документів. Необхідність реферування щораз більших обсягів документів і при цьому зменшення суб'єктивізму в наданні інформації зумовили впровадження в реферування електронних технологій.

Складання рефератів формалізованим способом не виключає інтелектуальних дій людини. Насамперед, це стосується процедури редагування тексту реферату, яка передбачає такі інтелектуальні дії, як усунення дублювання в тексті, досягнення зв'язності і логічності тексту реферату-екстракту, чіткості й лаконічності викладення.

Заголовок тексту, що входить у корпус текстів, має містити так звані метадані – загальну інформацію про даний текст. Якщо включити у неї відомості про жанр та тематику тексту, можна оптимізувати корпус та здійснювати реферування текстів необхідної спрямованості.



Таким чином, для текстів однієї тематики можна визначити набір ключових слів, які зумовлюватимуть пов'язаність даних текстів та формуватимуть кісток майбутнього реферату.

Основним методом, який використовується для реферування корпусу текстів, є статистичний. Це зумовлено тим, що корпус текстів являє собою набір попередньо розмічених текстів.

Принципом статистичного методу є ствердження, що вся необхідна для реферування інформація розташована навколо ключових слів. В даному методі, ключове слово – це знаменне слово тексту, яке з урахуванням синонімів зустрілося в тексті найбільше число разів. Речення, в яких містяться ключові слова, називаються ключовими. Саме вони використовуються для створення реферату.

Найбільш зв'язаними, а тому такими, що мають бути включені до реферату, вважаються речення, які містять найбільшу кількість однаково значущих слів. Тим не менш, у разі використання статистичного методу реферування обсяг і якість рефератів повністю залежать від статистичних характеристик тексту, тому речення, що містять найважливішу інформацію можуть бути взагалі не виділені та не ввійти до реферату.

Є кілька методів для визначення ключових фраз, такі як метод заголовку, метод позиції, метод ваги запиту та латентний семантичний аналіз. Для розробки системи, що здійснюватиме реферування корпусу текстів англійською мовою, було обрано метод ваги терміна.

Так, задача автоматизованого реферування полягає в тому, щоб створити реферат, максимально наближений за якістю до такого, який отримуємо внаслідок людської когнітивної діяльності. На основі цього ствердження нами був створений алгоритм реферування корпусу текстів англійської мови, який спирається на статистичний метод визначення ключових слів.

Ключове слово в аналізі тексту – слово, що представляє зміст тексту та отримується лінгвістичними і математичними методами (наприклад, аналізуючи частоту появи слова в тексті).

Для знаходження ключових слів у корпусі текстів необхідно:

1. За допомогою інформації про тематику тексту у метаданих визначити колекцію текстів зі спільною тематикою. Подальші дії проводитимуться з текстами однієї колекції.

2. Задля більшої оптимізації визначення ключових слів, на основі розмітки знайти у текстах колекції слова, що є самостійними частинами мови (іменник, прикметник, числівник, займенник, дієслово і прислівник). Інші слова, які не мають самостійного лексичного значення (службові частини мови), цифри, сполучення літер, вигуки і т.д., не оброблятимуться та будуть прийняті за стоп-слова.

3. Вирахувати частоту входження термінів у окремий документ колекції та у всі документи колекції. Частота входження термінів у окремий документ колекції визначається за формулою TF (term frequency – частота слова):

$$TF = \frac{n_i}{\sum_k n_k}$$



де n_i – число входжень t_i у документ, а знаменник – загальна кількість слів документу.

4. Для нормалізації df_i визначимо обратну частоту входження терміну у колекцію. Для цього використовуватимемо формулу:

$$idf = \log \frac{N}{df},$$

де N – кількість документів у колекції, df – кількість документів із t_i .

5. Щоб вирахувати вагу терміну t_i у колекції, використаємо формулу: $TF-idf = TF * idf$. Таким чином отримуємо ключові терміни для всієї колекції документів.

6. Знайдемо речення, в яких зустрічаються ключові слова, та сформуємо на їх основі реферат колекції.

Щільність ключових слів – відношення кількості наведених в тексті ключових слів до загальної кількості слів у даному тексті. Оптимальна щільність ключових слів становить 4-6%. Таким чином, стиснення тексту досягатиме приблизно 50%.

Завдяки виділенню цільних ключових речень в реферат, мінімізується можливість порушення синтаксичних зв'язків та дрібнення смислу та ідеї тексту. Ідея полягає в тому, щоб зберегти оригінальну ідею тексту та використані в ньому семантичні одиниці.

Таким чином, в результаті аналізу методів автоматизованого реферування новинних повідомлень була розроблена математична модель реферування текстів природною мовою та запропонований алгоритм для програмної реалізації автоматизованого реферування.

Список літератури

1. Карпіловська Є. А. Вступ до прикладної лінгвістики : комп'ютерна лінгвістика / Є. А. Карпіловська. – Донецьк : Юго-Восток, 2006. – 188 с.
2. Кутузов А. Б. Корпусная лингвистика. Лекция 2 [Электронный ресурс]. – Режим доступа : http://tc.utmn.ru/files/corpus_2.pdf
3. Mitkov R. Towards automatic annotation of anaphoric links in corpora / R. Mitkov // International Journal of Corpus Linguistics 4th ed. – 1999. – P. 261-280.
4. Nenkova A., McKeown K. Automatic Summarization. / A. Nenkova, K. McKeown. – NY.: Springer US, – 2011. – pp. 216.
5. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. / K. Sparck Jones. – L.: Journal of Documentation, –1972. – pp. 12.
6. Пятецкий-Шапиро Г.С. Data Mining и перегрузка информацией // Вступительная статья к книге: Анализ данных и процессов / А. А. Барсегян, М. С. Куприянов, И. И. Холод и др. 3-е изд. перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.
7. Никоненко А.К. Огляд комп'ютерно-лінгвістичних методів обробки природномовних текстів. / А.К. Никоненко – Київ: «Искусственный интеллект» 3'2011. – 8 с.
8. Шемякин Ю.И. Начала компьютерной лингвистики: Учеб. пособие. / Ю.И. Шемякин. – М.: Росвузнаука, 1992. – 322 с.
9. Mitkov R. Towards automatic annotation of anaphoric links in corpora / R. Mitkov // International Journal of Corpus Linguistics 4th ed. – 1999. – P. 261-280.



ЗАДАЧА КЛАСИФІКАЦІЇ ТЕКСТІВ РОСІЙСЬКОЮ МОВОЮ ЗА ГЕНДЕРНИМИ ОЗНАКАМИ

Борзенкова А.В.

*Національний технічний університет
"Харківський політехнічний інститут",
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: borzenkova-alina@yandex.ru*

На сьогоднішній день можна говорити про існування гендерних досліджень, що вивчають обидві статі, а точніше – процес соціального конструювання розходжень між статями. Гендер вважається соціокультурним конструктом, пов'язаним із приписуванням індивіду певних якостей і норм поведінки на основі його біологічної статі [1].

В останнє десятиліття галузь обробки природних мов і, зокрема, підрозділ «Класифікація текстів» розвивається дуже інтенсивно. Це багато в чому пов'язано з тим, що з кожним роком обсяг інформації, що зберігається на електронних носіях, значно зростає, і потрібні ефективні алгоритми для обробки та аналізу документів, написаних на природних мовах. Удосконалення алгоритмів, у свою чергу, можливо завдяки збільшенню потужності і продуктивності сучасних комп'ютерів.

Соціальні науки вступили в епоху науки даних, використовуючи безпрецедентні джерела писемності [2-4]. Через засоби масової інформації, такі як Facebook і Twitter [5], регулярно користуються більше 1/7-й населення світу, простежуються відмінності між жіночим та чоловічим мовленням. Щоб розібратися у масивних даних, необхідні багатопрофільні співробітництва між такими областями, як комп'ютерна лінгвістика та соціальні науки. У даній роботі демонструється інструмент, який описує схожості та відмінності між групами людей з точки зору їх використання мови.

У самому загальному плані дослідження гендера у мовознавстві стосується двох груп проблем.

1. Мова і відображення в ньому статі. Мета такого підходу полягає в описі і поясненні того, як маніфестується у мові наявність людей різної статі (досліджуються в першу чергу номінативна система, лексикон, синтаксис, категорія роду та ін.), які оцінки приписуються чоловікам і жінкам і в яких семантичних областях вони найбільш помітно та чітко виражені.

2. Мовну, і в цілому комунікативну, поведінку чоловіків і жінок, де виділяються типові стратегії і тактики, гендерно специфічний вибір одиниць лексикону, способи досягнення успіху у комунікації, переваги у виборі лексики, синтаксичних конструкцій та ін. – тобто специфіка чоловічого і жіночого мовлення [6].

Актуальність цього дослідження обумовлюється важливістю визначення, вивчення і опису способів формування гендерних стереотипів, які несвідомо і/або усвідомлено закладаються у світосприйнятті людиною за допомогою сприйняття їм текстів повідомлення, впливу з боку засобів масової інформації



та безпосередньої комунікації в соціумі. Важливість дослідження обумовлена збільшеною потребою лінгвістів, культурологів, психологів, педагогів в освоєнні механізмів, що надаються інтернет-простором для самопрезентації особистості, а також в адекватному і детальному визначенні тієї ролі, яку сьогодні відіграє віртуальний світ у житті людини, надаючи йому варіант альтернативної реальності. Самопрезентація у соціальній мережі є для сучасної людини одним з найбільш важливих атрибутів мовного оформлення та підтвердження самобутності власного Я.

Існуючі алгоритми класифікації можна використовувати не тільки безпосередньо для класифікації текстів, а й, наприклад, для вилучення з них додаткової інформації. У даній роботі буде розглядатися одне з таких напрямків, а саме, автоматична класифікація російськомовних текстів за гендерними ознаками, тобто профілювання автора анонімного тексту. Профілювання автора – це встановлення деяких значущих характеристик людини на основі написаного ним тексту.

Автоматичне профілювання має безліч застосувань. Одне з них – це судова авторознавча експертиза. Отримання будь-яких даних про особу злочинця дозволяє значно звузити область пошуку і заощадити час. Актуальною зараз є і проблема пошуку інтернет-зловмисників, коли у слідства немає ніяких інших доказів, окрім декількох повідомлень з погрозами або звинуваченнями

Метою роботи є вирішення задачі класифікації текстових повідомлень за гендерними ознаками.

Задача класифікації – формалізована задача, в якій є множини об'єктів (ситуацій), розділених деяким чином на класи. Задана кінцева множина об'єктів, для яких відомо, до яких класів вони відносяться. Ця множина називається вибіркою. Класова приналежність інших об'єктів невідома. Необхідно побудувати алгоритм, здатний класифікувати довільний об'єкт з вихідної множини.

У процесі роботи над роботою зроблено огляд основних напрямків гендерної лінгвістики, методів класифікації текстових документів та існуючих систем, що реалізують класифікацію текстових документів. Автором були проаналізовані особливості гендерних ознак у соціальних мережах, розроблено алгоритм класифікації текстів з використанням лінійного онлайн класифікатора та коефіцієнтів гендерного гепу [7].

Ідея лінійного класифікатора полягає у тому, що кожній категорії сі відповідає вектор $\vec{c}_i = \{c_{i1}, \dots, c_{in}\}$, де n – розмірність простору документів. В якості правила класифікатора використовується наступна формула:

$$CSVi(d) = \vec{d} \cdot \vec{c}_i = \sum_j c_{ij} d_j. \quad (1)$$

Звичайно проводиться нормалізація так, що підсумкова формула для $CSVi(d)$ – це косинус кута між вектором категорії \vec{c}_i і вектором документа \vec{d} .

$$CSVi(d) = \frac{\vec{c}_i \cdot \vec{d}}{|\vec{c}_i| |\vec{d}|}. \quad (2)$$

Аналіз гендерного гепу у реалізації мовних одиниць у письмових текстах був проведений через вимірювання параметрів загальної закономірності побу-



дови структури тексту: коефіцієнти предметності, якості, динамізму, активності і зв'язності тексту.

Предметність (Pr) вимірювалася співвідношенням числа іменників і займенників до числа прикметників і дієслів:

$$Pr = \frac{\text{іменники} + \text{займенники}}{\text{прикметники} + \text{дієслова}} \quad (3)$$

Якісність (Qu) визначалася співвідношенням числа прикметників і прислівників з числом іменників і дієслів:

$$Qu = \frac{\text{прикметники} + \text{прислівники}}{\text{іменники} + \text{дієслова}} \quad (4)$$

Активність (Ac) тексту виводилася як співвідношення дієслів і дієслівних форм із загальною кількістю слів у тексті (N):

$$Ac = \frac{\text{дієслова} + \text{дієслівні форми}}{N} \quad (5)$$

Динамізм (Din) був визначений як співвідношення числа дієслів і дієслівних форм з іменниками, прикметниками і займенниками:

$$Din = \frac{\text{дієслова} + \text{дієслівні форми}}{\text{іменники} + \text{прикметники} + \text{займенники}} \quad (6)$$

Коефіцієнт зв'язності тексту (Con) виміряли як співвідношення числа прийменників і союзів з числом самостійних речень (PN):

$$Con = \frac{\text{прийменники} + \text{сполучники}}{PN} \quad (7)$$

Виходячи з результатів лінгвістичних досліджень, можна зробити висновок про те, що у висловлюваннях чоловіків і жінок дійсно існують помітні відмінності. При цьому вони можуть бути представлені у формі, допустимою для комп'ютерної обробки, а значить, їх можна використовувати і в методах машинного навчання для класифікації текстів.

Інші характеристики особистості автора, за даними лінгвістів, також впливають на мову людини. Тому, використовуючи цю інформацію при класифікації, можна сподіватися на отримання більш точних даних про автора анонімого тексту.

Список літератури

1. Coats J. Women, men and language. A sociolinguistic account of sex differences in language. – New York, 1986. – 389 p.
2. Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, et al. (2009) Computational social science. – Science 323: 721-723.
3. Weinberger S (2011) Web of war: Can computational social science help to prevent or win wars? the pentagon is betting millions of dollars on the hope that it will. Nature 471: 566–568. doi: 10.1038/471566a
4. Miller G (2011) Social scientists wade into the tweet stream. Science 333: 1814–1815. doi: 10.1126/science.333.6051.1814
5. Facebook (2012) Facebook company info: Fact sheet website. Available: <http://newsroom.fb.com>. Accessed 2012 Dec.
6. Preisler B. Linguistic Sex Roles in Conversation. – Paris: Mouton, 1986. – 287 p.
7. Горошко Е. Особенности мужского и женского стиля письма // Преображение / Е. Горошко . – 1998. – № 6. – С. 48-64.



О РЕФЕРЕНЦИИ И КОРЕФЕРЕНЦИИ В ЛИНГВИСТИКЕ

Кузьмина М.А., Бабкова Н.В.

*Национальный технический университет
«Харьковский политехнический институт»,*

г. Харьков, ул. Пушкинская, 79/2, (057) 707-63-60, nadjenna@gmail.com

Еще десятилетие назад термин «референция» был редкостью и практически не встречался в научной литературе. Однако настоящее время вопросы референции обрели актуальность и занимают ведущее место в исследованиях теории языка. Референция – это отношение включенного в речь имени или выражения (именной группы) к объекту действительности [1].

Теория референции переживает период бурного развития: возникают новые подходы в изучении этого понятия. Если на начальном этапе развития данная теория основывалась на проблеме соотнесения лингвистического объекта и его референта – объекта реальной действительности, то с развитием казуальных теорий (С. Крипке, Д. Каплан и др.), основные понятия референции получили другую трактовку.

Началом современных теорий референции стали наблюдения за значениями и употреблением имен нарицательных, и, в первую очередь, конкретной лексики. Имена нарицательные наделены определённым понятийным содержанием (концептом), однако способны к денотации (обозначению) предметов действительности. Они указывают на какой-либо предмет и одновременно с этим сообщают о нём новую информацию. Референция осуществляется теми, кто пользуется языком, а не теми выражениями, которые употребляет для своих целей говорящий [2].

Связь теории с говорящим субъектом дала положительные результаты, а понимание референции как субъективного акта, который определяется намерением говорящего, вызвало интерес к косвенной референции – уловкам, намекам. То есть к тем средствам, к которым говорящий прибегает, преследуя разные цели: при желании высказаться о себе, не делая в то же время никаких признаний, или при нежелании нести ответственность за свои слова. В данном случае референция определяется нормами ведения разговора и уровнем владения искусством хитросплетения.

Однако преувеличение роли намерений говорящего в установлении связи между языковым выражением и предметом или явлением действительности вызвало протест у некоторых ученых и в дальнейшем С. Крипке предложил асемантическую теорию референции [3]. В рамках этой теории различают референцию говорящего и семантическую референцию. Различия между этими теориями определяются такими факторами: различием в понимании механизмов установления связи между именем и предметом; различием в понимании роли именных выражений в формировании суждения; различием в степени прагматизации теорий.



Явление кореференции – это не просто различное обозначение объекта, проявляющееся при его повторной номинации или переименование одного и того же текстового референта. В большинстве лингвистических исследований при использовании понятия «кореференция», в него вкладывается неодинаковый смысл. Помимо повторной номинации (переименования) «упоминание» референта сопровождается приписыванием некоторых признаков, которые обогащает наше представление о референте новой дополнительной информацией. Основная текстообразующая функция таких кореферентных выражений (связность, целостность и т. д.) остается неизменной. В дискурсе цепочка кореферентных выражений формирует понятие (концепт) - так называемую макроцепочку. Структура макроцепочки представлена микроцепочками кореферентных выражений, которые определяют признаки концепта (понятия). Совокупность микроцепочек позволяет сформировать единое понятие на основе его признаков. Более того, кореферентные выражения позволяют удерживать в памяти информацию об объекте на протяжении всего дискурса, постоянно подкрепляя ее терминальной номинации - номинации, вызывающей репрезентацию концепта.

Таким образом, самым точным определением кореферентности можно считать следующее. Кореферентность - это отношение между компонентами высказывания (обычно именными группами), которые обозначают один и тот же объект или ситуацию, т. е. имеют один и тот же референт [1]. Кореферентность лежит в основе связности текста и чаще всего является одним из видов анафоричных отношений, которые выражаются местоимениями или значением определенности в составе одного из кореферентных выражений.

Кореферентные выражения формируют тематическое ядро текста с помощью лексических средств - однозначных лексем, из которых слагается смысловой каркас текста и которые могут служить ключевыми словами текста. Кореференцию в основном рассматривают как отношение между именными группами. Кореференция определяется как «различное обозначение одного и того же объекта, денотата» [2] и рассматривается в тесной связи с повторной номинацией, которая может осуществляться не только на уровне слова, но и на уровне предложения.

В исследовании кореференции (в когнитивном аспекте), нельзя ограничиться исключительно лингвистическим подходом, поскольку текст рассматривается как лингвистическая и коммуникативная единица. Не все виды лингвистического анализа могут служить прагматическим целям, ибо имеют в виду текст как единицу языка (элементам которого приписываются значения в рамках лингвистической семантики), а не как единицу общения, отражающую актуальную знаковую деятельность в знаковой ситуации (по определенному поводу, для определенной цели и определенным адресатом) [3].

Теория референции во многом стала общим местом слияния вопросов лингвистики, языкознания и логики, поставив своей задачей изучение соотношения языка и реальности. На сегодняшний день большинство исследований и публикаций связано с автоматизацией определения кореферентных связей в естественно-языковом тексте. Использование предметно-ориентированных се-



мантических моделей позволяет эффективно выполнять семантическое сравнение посимвольно разных имен в базах данных для различных прикладных целей.

Список литературы.

1. Милевская Т. В. Связность как категория дискурса и текста: когнитивно-функциональный и коммуникативно-прагматический аспекты. дис... д. филол. наук. – Ростов-на-Дону. –2003. –390 с.
2. Островский О. Л. Когнитивная кореференция и ее структурно-организующая роль в тексте: на материале французских газетно-информационных текстов: дисс ... канд. филол. наук. –М. –2002. –162 с.
3. Попова О. В. Функциональные особенности английских личных местоимений в когнитивном аспекте: дис. ... канд. филол. Наук. –Бийск. –2006. –165 с.



АНАЛІЗ ІСНУЮЧИХ СИНОНІМАЙЗЕРІВ ТА ОГЛЯД ПРОБЛЕМ ЩОДО ЇХ СТВОРЕННЯ

Юр'єва К.О.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків, вул. Пушкінська 79/2, (057) 707-63-60*

Лексика мови постійно збагачується. Одне й те ж саме поняття характеризується з різних сторін, виділяються нові його ознаки. В лексемах може відбиватися ставлення мовців до предметів реального світу, їхні емоції та волевиявлення. Між словами, що позначають одне розгалужене поняття чи його компоненти або різні відтінки й оцінки, встановлюються близькі зв'язки. Їх називають синонімічними, а слова — синонімами [1].

Синоніми – це слова, які мають близьке або однакове лексичне значення, але різні за написанням і звучанням. Вони допомагають уникати одноманітності мови, не виправданих повторень однакових слів. Синоніми також дають змогу більш точно сформулювати думку, роблять мову виразною. Такі слова виникають у мові постійно. Це обумовлено рядом причин. Однією з основних є прагнення людини знайти у вже відомих предметах, явищах навколишнього світу якісь нові риси й відтінки, тобто поглибити і розширити існуючі поняття навколишньої дійсності. Нова додаткова ознака поняття називається новим словом, подібним або тотожним за значенням з уже наявними найменуваннями.

Нерідко виникнення синонімів зумовлено проникненням і освоєнням запозичених слів, близьких або тотожних за значенням українським, наприклад, літак - аероплан. Спостерігалось і зворотнє явище, коли предмет, названий спочатку запозиченим словом, потім отримував синонімічне українське найменування, наприклад: гелікоптер - вертоліт.

Таким чином, можна сказати, що в українській мові багато синонімів, але процес появи нових синонімів не переривається. Синонімічні ряди постійно поповнюються [1].

Словник — упорядкований в алфавітному чи тематичному порядку список заголовних слів, лексикографічно опрацьованих. Таке довідкове видання містить впорядкований перелік мовних одиниць (слів, словосполучень тощо) з короткими їх характеристиками або характеристиками позначених ним понять або з перекладом на іншу мову. Словник є одним із засобів накопичення результатів пізнавальної діяльності людства, показником культури народу. Вони виконують інформативну та нормативну функції: являють собою універсальні інформаційні джерела для розуміння того чи іншого явища та найточніша консультація щодо мовних норм [2].

Словники синонімів мають досить цікаву функцію, яка здатна полегшити написання того чи іншого тексту з можливістю уникання одноманітності словникових форм. Гарним прикладом роботи з словниками синонімів є також синонімайзери.



Синонімайзер – це програма, яка дозволяє зробити з одного вихідного тексту велику кількість однакових за змістом текстів, які написані різними словами. Його зазвичай використовують для розмноження статей при просуванні сайту за допомогою бірж статей або каталогів статей, а також для розмноження опису сайту при реєстрації у великій кількості каталогів.

Попередньо, перед генерацією статті, кожній фразі тексту, де це тільки можливо, задається група синонімів і допустимих перестановок. Те ж саме робиться і зі словами усередині фрази.

Якщо провести хорошу роботу з синонімами в тексті статті і потім включити випадкову генерацію заданої кількості статей, то можна отримати значну кількість текстів, які з певною ймовірністю будуть унікальними для пошукових систем.

Для того, щоб запобігти появі занадто схожих статей внаслідок випадкового генерування, в синонімайзерах зазвичай передбачена функція видалення схожих статей за методом шинглів. Метод шинглів використовується пошуковими системами для визначення унікальності тексту сторінок сайтів. Шингл – це набір з певної кількості слів, що стоять один за одним у тексті [3].

Основні проблеми й складності використання синонімайзера укладається в наступному:

- 1) здійснюваний добір синонімів повинен проводитися в рамках контексту словосполучення;
- 2) в українській мові відсутній порядок слів у реченні, що суттєво ускладнює синтаксичний аналіз тексту;
- 3) для полегшення читання тексту в ньому не повинно бути використана безліч однакових слів;
- 4) повинна виконуватися закономірність Зіфра.

Для подальшого аналізу роботи з текстом, отриманого за допомогою синонімайзера необхідно здійснювати порівняння текстів на унікальність частинами й вибирати найбільш унікальний текст, який бажане вичитувати для виправлення помилок у закінченнях слів.

Отже, актуальними залишаються проблеми розробки сучасних комп'ютерних технологій у лексикографії. Особливістю електронних словників є те, що вони неможливі без технічної підтримки. Зокрема, завдання які виконують такі видання формують їх переваги, через що вони відповідають читацьким потребам.

Список літератури:

1. *Королева Н.В.* Проблема синонимии как одна из важных проблем лексикологии / Н.В. Королева, Е.В. Артамонова // Вестник КАСУ. – 2009. – №2. – с. 91-97
2. *Реформатский А.А.* Введение в языковедение: Учебник для вузов/ А.А.Реформатский; Под ред. В.А. Виноградова - 5-е изд., испр. – М.: Аспект Пресс. – 2006. – 536с.
3. *Кронгауз М.А.* Семантика. Учебник для студ. лингв. фак. высш. учеб. Заведений / М. А. Кронгауз. - 2-е изд., испр. и доп. – М.: Академия. – 2005. –352 с.



ВИРІШЕННЯ ЗАВДАННЯ SENTIMENT ANALYSIS ЗА ДОПОМОГОЮ СЛОВНИКОВОГО ПІДХОДУ (НА ПРИКЛАДІ МАТЕРІАЛУ ФОРУМІВ СМАРТФОНІВ)

Медведська А.В.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: nyuta3871@inbox.ru*

Вивчення думок і відгуків інших людей стало важливою частиною процесу прийняття рішень. Коли певний вибір пов'язаний з фінансовими витратами, то люди так чи інакше спираються в цьому питанні на досвід інших людей. Сьогодні «social web» дає нові можливості для обговорення та поширення інформації серед людей на форумах, у блогах, соціальних мережах, де люди мають можливість обмінюватися думками щодо різних товарів та послуг. Аналіз думок ще називають тональністю, або opinion mining – це клас методів контент-аналізу в комп'ютерній лінгвістиці, призначений для автоматизованого виявлення в текстах емоційно забарвленої лексики та емоційної оцінки думок по відношенню до об'єктів [1].

В аналізі тональності тексту вважається, що текстова інформація в мережі Інтернет ділиться на два класи: факти і думки. Ключовим поняттям є визначення думок.

Думки діляться на два типи:

1. Проста думка.
2. Порівняння.

Проста думка містить висловлювання автора про один об'єкт. Вона може бути висловлена прямо, або неявно. В обох випадках проста думка зазвичай має емоційне забарвлення – позитивне чи негативне.

Другий тип думок – порівняння – можна розділити на три види [2]:

1. Порівняння аспектів об'єктів на користь одного (Non-equal Gradable).
2. Прирівнювання аспектів різних об'єктів (Equative).
3. Перевага одного об'єкту над іншими (Superlative).

Загальний підхід у вирішенні задачі Opinion Mining виглядає наступним чином: збір колекції документів, попередня обробка текстів, виділення ознак емоційності, визначення полярності тексту.

Визначення полярності тексту зазвичай розглядається на двох рівнях:

1. На рівні документа.
2. На рівні речення.

Завдання на рівні документа полягає у визначенні полярності документа в цілому. Причому, текст документа може одночасно містити речення, як з негативним, так і з позитивним емоційним забарвленням.

Завдання визначення полярності речень вирішується як в якості підзадачі аналізу тональності документа, так і в якості самостійного завдання, наприклад,



при аналізі коротких повідомлень і коментарів у соціальних мережах. Цей метод призначений для більш точного аналізу думки в тексті [3].

При вирішенні задачі аналізу тональності тексту популярними є методи, засновані на знаннях, Machine Learning та прихована Марковська модель.

Етапи вирішення задачі аналізу тональності текстів:

1) Збір текстів для перевірки роботи програми.

Для перевірки роботи програми було зібрано 200 повідомлень позитивної на негативної забарвленості за тематикою смартфони. Для збору повідомлень було обрано англomовний сайт <http://www.pcadvisor.co.uk>. На цьому сайті розміщені огляди різної техніки, у тому числі мобільної, де користувачі мають можливість обмінюватися думками щодо певного товару. Повідомлення мають наступний вигляд: «I love Samsung products! I had an iPhone prior to this GalaxyS5. I would never go back! Samsung products are so user friendly. I have to Samsung tablets that are wonderful. Thus phone has excellent photo quality! To much to name.»

2) Складання словників позитивно на негативно забарвленої лексики.

Були складені словники емоційно забарвленої лексики, що виражають думки користувачів щодо мобільної техніки. Мова повідомлень англійська. Кількість позитивних слів – 103, кількість негативних – 86, позитивних смайликів – 26, негативних – 22. Наприклад, словник позитивно забарвленої лексики містить такі слова: strong, excited, gorgeous, delightful та інші. А негативно забарвленої – такі, як: disgruntled, terrible, dreadful, ugly та інші.

3) Розбір повідомлення на слова. У такому вигляді їх можливо буде використовувати для подальшої обробки.

4) Пошук емоційних елементів за допомогою системи Analysis of emotion. Якщо повідомлення містить емоційні елементи – сильно емоційно забарвлене, якщо не містить – не сильно.

5) Підключення адаптованого словнику емоційно забарвлених слів та перевірка повідомлень на наявність графічних усмішок (смайликів).

6) Визначення емоційної забарвленості повідомлення з використанням математичних методів для підрахунку ваг слів та смайликів. Для подальшого аналізу були обрані методи [4] та [5].

Список літератури

1. Stefano Baccianella Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining (англ.) // Proceedings of LREC : конференція. – 2010. – С. 2200-2204.
2. Bing Liu. Sentiment Analysis Tutorial // AAAI-2011, San Francisco, USA.
3. Pang B. & Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval, v.2 n.1-2, January, 2008 - pp.1-135.
4. Усталов Д. А. Вилучення термінів з російськомовних текстів за допомогою графових моделей // Теорія графів і додатків. Єкатеринбург: Вид-во Урал. – 2012. – с. 62-69.
5. Клековкіна М.В., Котельников Є.В. Метод автоматичної класифікації текстів за тональністю, заснований на словнику емоційної лексики // XIV Всерос. наук. конф. Переславль-Залеський: Вид-во «Університет міста Переславль». – 2012. – с. 118-123.



АВТОМАТИЧНЕ ВИЛУЧЕННЯ ДІЕСЛІВНИХ КОЛОКАЦІЙ АНГЛІЙСЬКОЇ МОВИ

Чалюк Г.Є.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: anuntus@mail.ru*

Проблема вивчення синтагматичною сполучуваності та стійкості сполучень слів є однією з ключових в лінгвістиці. Існуюча література та словники не завжди повно і послідовно відображають інформацію про сполучувані переваги лексем та стійкі словосполучення. Актуальність теми обумовлена тим, що отримання нових даних про сполучуваність, розробку нових методів її вивчення повинні сприяти розвитку лексикографії, синтаксису, семантики. Робота спрямована на опис та експериментальну верифікацію лінгвістичних і статистичних прийомів виявлення колокацій в корпусах текстів на матеріалі англійської мови. Статистичний метод вилучення колокацій отримав широке поширення в корпусній лінгвістиці. Найпростішим способом виявити колокації у тексті є зіставлення переліку частотності слів, який з'являється зліва або справа ключового слова всередині даного проміжку. Розмір такого проміжку зазвичай складає 5 слів зліва або справа від ключового слова. Статистичні показники асоціацій отримали вельми широке розповсюдження у сучасних лінгвістичних дослідженнях. Ці показники базуються на частотній сполучуваності пар слів й сполучуваності кожного елемента, який можна вираховувати у межах певного проміжку.

У нашому дослідженні в якості міри асоціацій ми використовуємо коефіцієнт взаємної інформації (МІ), який можна розуміти як коефіцієнт синтагматичної сили між елементами колокації.

МІ порівнює залежні контекстно-пов'язані частоти з незалежними, якщо б слова у тексті з'являлись абсолютно випадково.

Коефіцієнт взаємної інформації для біграми може бути обчислений за такою формулою:

$$MI(n, c) = \log_2 \frac{f(n, c) * N}{f(n) * f(c)}$$

де МІ – коефіцієнт взаємної інформації;

n – ключове слово;

$f(n, c)$ – частота зустрічальності ключового слова n у парі з колокатом c;

$f(n), f(c)$ – абсолютні (незалежні) частоти ключового слова n і колоката c в корпусі;

N — загальне число словоформ в корпусі.



Проте, необхідно зауважити врахувати той факт, що слова синтаксично пов'язані, вони не зустрічаються у тексті випадково. Отже, вилучення колокації потребує не тільки статистичного методу, а й підходу, що базується на синтаксисі, який бере до уваги морфологічні й синтаксичні властивості слів у корпусі.

Застосування описуваних методів для отримання інформації про лексичну і синтаксичну сполучуваність на базі великих корпусів текстів вже сьогодні служить основою для створення словників і граматик нового типу. В даний час в сучасній лінгвістиці незамінним інструментом і одночасно матеріалом для лінгвістичних досліджень і вирішення прикладних завдань стали корпуси текстів. Об'єктом нашого дослідження виступає явище синтагматичною сполучуваності в англійській мові. Предмет дослідження - статистично стійкі поєднання (колокації), що відповідають певним лексико-синтаксичним моделям.

Планується провести ряд експериментів, метою яких є порівняння різних підходів до автоматичного вилучення колокацій дієслово-іменник. Основними темами розглядання є вплив розміру проміжку й POS-tagging, у той час як розширення розміру проміжку має неоднозначний вплив. З одного боку, це дозволяє вилучати віддалені словосполучення, але, з іншого, це призводить до помилкових сполучень, що в свою чергу призводить до розгляду використання підходу, що базується на синтаксисі для вилучення колокації дієслова та іменника.

В якості інструмента для нашого дослідження обрано IntelliText – систему, розроблену Центром для Дослідження Перекладу Університету Лідса. IntelliText надає достатню можливість для лінгвістичних досліджень і містить типові корпуси, включаючи морфологічно анотовані корпуси англійської мови.

Список литературы

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. - М.: МИЭМ, 2011. - 272 с.
2. Lewis, M. Teaching Collocation: Further Development in the Lexical Approach, Hove: Language Teaching Publications, 2000.
3. Biber D., Conrad S., Reppen R. Corpus Linguistics. Investigating language structure and use. Cambridge University Press, 1998
4. Герд А.С. РНК и академическая лексикография // Труды международной конференции «Корпусная лингвистика – 2006». – СПб.: Изд-во С.-Петербур. ун-та; Изд-во РХГА, 2006. – С. 88-91.
5. Cowie, A. The Treatment of Collocations and Idioms in Learner's Dictionaries. Applied Linguistics, 1981. – 2(3). – P. 223–235.



СОЗДАНИЕ ЭЛЕКТРОННОГО СЛОВАРЯ АНТОНИМОВ

Пилипенко А.А.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,
e-mail: nastya7523@yandex.ru*

Прикладная лексикография является одним из направлений прикладной лингвистики. В ее задачу входит создание различных проблемно- и предметно-ориентированных автоматизированных и автоматических словарей и баз данных.

Электронный словарь представляет из себя: оболочку, которая предоставляет пользователю различные функции (стандартный перевод, перевод слов в браузере, перевод слов с нечетким запросом, перевод слов с последующим озвучиванием и т.д.), а также, дополнительно подключаемые словари [1].

В языкознании нет достаточно строгой дифференциации словарей. К лингвистическим словарям относятся: толковые, переводные, частотные, грамматические и др. К специальным лингвистическим словарям относятся: фразеологические словари, словари "крылатых слов", словари пословиц и поговорок, словари синонимов, антонимов, омонимов и др. [2].

Словарь антонимов занимает важное место среди других словарей: он дает описание лексики с точки зрения смысловой противоположности ее единиц. Электронный словарь антонимов является разновидностью электронных словарей, именно поэтому функции словарей распространяются на все виды словарей антонимов.

Следовательно, в рамках электронного словаря антонимов должны быть реализованы следующие функции:

- поиск слова непосредственно по введенной словоформе;
- вывод найденных данных (толковой словарной статьи или списка лексем) или сообщение об их отсутствии;
- ввод слова и антонима с проверкой на уникальность и с возможностями редактирования вводимых данных;
- вывод списка "похожих" слов на входное слово с ошибкой или на ввод отсутствующего в словнике слова (режим "Варианты");
- перекрестные ссылки (ссылки в статьях на другие статьи);
- установка закладок (для быстрого открытия словарных статей при дальнейшей работе);
- возможность пополнения словаря.
- поиск по слову, встречающемуся в словарной статье;
- просмотр вводимых в сеансе работы слов (режим "Предыстория");
- поддержка различных операционных систем;
- озвучивание словарной статьи;



Для понимания строения электронного словаря на структурном уровне следует рассмотреть их макроструктуру (под которой в лексикографии понимается общая структура словаря, содержание и связь его частей) и микроструктуру (формат отдельных словарных статей и характер их заполнения).

При ознакомлении с микроструктурой электронного словаря важно отметить, что отношения между частями словарной статьи не являются линейными и структурирование происходит на основе гипертекста. Словарная статья имеет четкую логическую структуру с иерархическими связями между элементами. Условно словарную статью можно разделить на левую и правую часть. В случае разработанного словаря антонимов левая часть состоит из лексем. Правая часть словаря — та, в которой объясняется заглавная единица. В словаре антонимов правая часть представляет собой антоним заглавного слова [3].

Макроструктура словаря определяет отбор лексики, объем и характер словаря, принципы расположения материала. Макроструктура электронного словаря в линейной последовательности во многом повторяет макроструктуру традиционного словаря, и к ее составляющим относятся: предисловие, правила пользования, перечень сокращений, корпус, грамматический очерк, различные списки (например, список источников). В электронном словаре, благодаря гипертекстовой структуре каждый из них может быть активизирован в любой последовательности по желанию пользователя [4].

Процесс создания электронных словарей включает в себя три этапа.

Первый этап создания словаря включает в себя следующее:

1) критический анализ современного состояния лексикографии (анализ существующих словарей; исследование потребностей адресата — конечного пользователя словаря);

2) определение требований к словарям на основе полученных данных.

3) разработка макроструктуры (и медиоструктуры) словаря.

4) разработка микроструктуры словаря.

Второй этап

1) Составление словника (перечень заглавных языковых единиц (слов, морфем, словосочетаний), которые описываются в лексикографическом труде).

2) Распределение слов по семантическим полям (для словарей с элементами тезауруса).

3) Подбор переводных эквивалентов, синонимических, антонимических пар, для слов словника или же тезаурусов.

4) Исследование словарных единиц (переводных эквивалентов) с точки зрения их денотативно-сигнификативных, экспрессивно-прагматических и синтагматических характеристик и занесение результатов исследования в лексикографическую базу данных.

5) Оформление словарных статей и словаря в целом в соответствии с разработанной ранее структурой.

Третий этап заключается в экспериментальной проверке соответствия полученного продукта поставленным целям [5] [6] [7].

При выполнении данной работы была поставлена цель комплексного изучения существующих этапов составления электронных словарей в свете разви-



тия электронной лексикографии. Нами был сделан вывод, что электронные словари обладают рядом очевидных и существенных преимуществ по сравнению со словарями традиционными. Одним из самых главных преимуществ электронных словарей является одновременный поиск не только по названию словарной статьи, но и по всему объему словарей, что невозможно в бумажном варианте.

Очевидно, что в ближайшие годы электронные словари будут играть все более и более важную роль в нашей жизни, что приведет к повышению интереса к изучению проблемы возможностей, которые предлагают электронные словари и ожиданий пользователей от словарей на информационных носителях. Электронный словарь прочно занимает свое место в практике пользователей.

Список литературы

1. Савина А.С., Типикина Т.А. Что внутри электронного словаря? // Наука и жизнь. – №10, 1999 г. – 63 с.
2. Маслов Ю.С. Введение в языкознание. М.: Высшая Школа, 1997, с. 120-124.
3. Дубічинський В.В. Лексикографія [Текст] : навч.-метод. посіб. для тсуд. ден. та заоч. навчання зі спец. "Переклад" і "Прикладна лінгвістика" / В. В. Дубічинський ; Нац. техн. ун-т "Харк. політехн. ін-т". - Х. : НТУ "ХПІ", 2012. - 68 с.
4. Берков В.П. Двухязычная лексикография / В. П. Берков. – СПб.: СПУ, 1996. – 248 стр.
5. Городецкий Б.Ю. Новое в зарубежной лингвистике. Вып. 14. Проблемы и методы лексикографии. - М.: Прогресс, 1983. - 399 с.
6. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – М.: Едиториал УРСС, 2003.
7. Дубичинский В.В. Теоретическая и практическая лексикография. – Wien Charkov: Wiener Slawistischer Almanach, 1998. – 160с.



РОЗРОБКА СЛОВНИКА СИНОНІМІВ ДЛЯ СИНОНІМАЙЗЕРА НОВИНИХ ПОВІДОМЛЕНЬ

Тупікова Н.С.

*Національний технічний університет
"Харківський політехнічний інститут",
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: elpys@rambler.ru*

В останні роки спостерігається динамічне зростання Інтернет-ресурсів, найважливішим видом яких є веб-сайти. Оптимізуючи сайт під пошукові системи великої уваги слід приділяти унікальності контенту. Унікальність контенту, в першу чергу, потрібна для позитивної оцінки сайту пошуковими машинами.

Одним із способів отримати унікальний контент для сайту новин є рерайтинг готових статей. Рерайтинг полягає в переписуванні вихідного тексту таким чином, щоб у результаті вийшов унікальний текст без втрати сенсу. Рерайтинг тексту проводиться, як правило, підбором синонімів до слів, або заміною фраз і цілих речень на смислові аналоги, в результаті чого пошукові системи сприймуть новий текст як унікальний [1].

У ході роботи були проаналізовані різні програми й он-лайн сервіси, які вже існують для реалізації рерайтингу, це програми підбору синонімів [2].

Синонімайзер – програма або сервіс для заміни слів у тексті на синоніми, що знаходяться в базі даних з метою видозміни тексту. Розроблялися синонімайзери виключно для допомоги тим, хто в тій чи іншій мірі пов'язаний з розробкою, управлінням та наповненням інтернет – ресурсів.

Синонімайзери можна розділити на автоматичні і ручні. Ручні пропонують користувачеві самому вибирати зі списку синоніми, які підходять. Автоматичні проробляють всю роботу без участі користувача.

Синонімайзери поділяються на серверні або он-лайн, тобто ті, що працюють виключно в Інтернеті, та десктопні, ті, що встановлюються на комп'ютер користувача [3].

Дія синонімайзера заснована на механізації роботи з текстами. Якість роботи синонімайзера безпосередньо залежить від словника синонімів, який використовує програма. Таким чином, словник синонімів для синонімайзера повинен бути створений коректно.

Словник синонімів для синонімайзера має вид $X | Y$, де X і Y можуть бути як будь-якою частиною мови, так і словосполученням або цілим реченням [4].

Аналіз електронних словників для синонімайзерів показав, що більшість слів мають не один синонім, а декілька. Це позначається в електронних синонімічних словниках наступним чином:

автомобиль|авто
автомобиль|автомашина
автомобиль|машина



Крім того, словники для синонімайзерів містять у собі усі форми слова. Так, наприклад, дієслово «краснеть» у словнику має вид:

краснев|багровев
краснел|багровел
краснела|багровела
краснеть|багроветь
краснея|багровея

Таким чином, можна помітити, що одне слово вживається у словнику декілька разів. Через це словник виходить об'ємним.

В наш час все прагне до спрощення і зменшення. Тому необхідно удосконалювати електронні словники для синонімайзерів, а саме зменшувати їх об'єм.

Якщо слово має декілька синонімів, ми пропонуємо писати їх через кому (,) після заголовного слова та вертикальної лінії (|). Це допоможе зменшити об'єм тому, що ми не будемо писати одне і те ж заголовне слово декілька разів, а також надасть користувачеві синонімайзера вибір синонімів. Це буде мати наступний вид: компанія | холдинг, підприємство

Деякі частини мови, наприклад, дієслова та прикметники, мають різні закінчення під час відмінювання. Надання всіх можливих форм слова робить словник об'ємним і громіздким. Для уникнення подібної ситуації, ми пропонуємо надавати в словнику лише слово у називному відмінку або інфінітив слова. Для того, щоб синонімайзер знаходив словоформи і замінював їх на синоніми, ми пропонуємо невраховувати закінчення слів. Це підвищить рівень релевантності.

Після користування синонімайзером, текст обов'язково повинен перевірятися рерайтером. Він повинен обрати потрібний синонім, якщо їх декілька, змінити закінчення обраного синоніма, а всі інші варіанти видалити.

При створенні електронного синонімічного словника було відібрано багато текстів з новинних сайтів з тематики саме інформаційних технологій. Крім слів, які стосуються тематики інформаційних технологій, словник містить слова загальної вживаності. Це такі слова як: «бесплатный» – «неоплачиваемый» тощо [5]. Об'єм створеного електронного синонімічного словника для синонімайзера складає близько 400 слів.

Список літератури

1 Рерайтинг: [Електронний ресурс]. – Режим доступу: <http://igroup.com.ua/seo-articles/rerajtynh/>

2 Програми для рерайтера: [Електронний ресурс]. – Режим доступу: http://mastersloga.ru/news/programmy_dlja_rerajtera/2013-04-14-47

3 Синонімайзер: [Електронний ресурс]. – Режим доступу: <http://animatika.ru/info/gloss/synonymizer.html>

4 Особливості побудови словника: [Електронний ресурс]. – Режим доступу: http://ua-referat.com/особливості_побудови_словника

5 Караванський С.Й. Практичний словник синонімів української мови: близько 15 000 синонімічних рядів / С.Й. Караванський. – К. : Кобза, 1995. – 472 с.



РОЗРОБКА ЕЛЕКТРОННОГО ТЛУМАЧНОГО СЛОВНИКА ТЕРМІНІВ З КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ НА ОСНОВІ КОРПУСУ ТЕКСТІВ

Скуменко С.І.

*Національний технічний університет
«Харківський політехнічний інститут»
г. Харків, ул. Пушкінська, 79/2, тел. 707–63–60
e-mail: s.relaxed@yandex.ua*

Дане дослідження присвячене вивченню однієї з наук, що динамічно розвивається – комп'ютерна лінгвістика, яка є новою гілкою прикладного мовознавства.

Спеціалізований тлумачний словник термінології комп'ютерної лінгвістики належить до словників комп'ютерного типу, який має базу даних і в якому передбачено різні режими його використання: пошук і систематизація інформації, яка використана у дефініціях, а також можливість перетворення словника на дослідну, навчальну або фактографічну базу даних. Комп'ютерний спеціалізований тлумачний словник термінів з комп'ютерної лінгвістики репрезентує терміносистему цієї галузі у її сучасному стані, на рівні передових лексикографічних й інформаційних технологій і є першим досвідом її систематизації та стандартизації.

Під терміном "комп'ютерна лінгвістика" (computational linguistics) зазвичай розуміється широка область використання комп'ютерних інструментів - програм, комп'ютерних технологій організації та обробки даних - для моделювання функціонування мови в тих чи інших умовах, ситуаціях, проблемних областях, а також сфера застосування комп'ютерних моделей мови не тільки в лінгвістиці, а й у суміжних з нею дисциплінах [1].

"Термін" комп'ютерна лінгвістика "задає загальну орієнтацію на використання комп'ютерів для вирішення різноманітних наукових і практичних завдань, пов'язаних з мовою, ніяк не обмежуючи способи вирішення цих завдань" [2].

Словотвір - постійний шлях поповнення лексичними засобами будь-якої мови і підмови, що дає найбільш значущий результат у кількісному відношенні. Способи словотворення, які традиційно виділяють в системі мови це: афіксація, конверсія, словоскладання, аббревіація.

Афіксація в українській та англійській термінології активно використовує приставки латинського походження, характерні для загальнонаукової мови (табл. 1).

Корпус текстів - це зроблена за певними правилами вибірка з проблемної області, тобто під корпусом текстів розуміється великий, структурований і оброблений спеціальним чином масив мовних даних кінцевого розміру, призначений для вирішення різних лінгвістичних завдань. Всі тексти, що входять в масив об'єднані деяким логічним задумом, логічної ідеєю [3].



Таблиця 1 - Приклади приставок латинського походження, як одного із способів словотворення

Афікс	Приклад
inter-	interactive (інтерактивний)
super-	superuser (привілегований користувач)
mini-	minidriver (мінідрайвер)
macro-	macrocommand (макрокоманда)
micro-	microfile (мікрофайл)
auto-	autodump (авторазгрузка)
multi-	multisystem (мультисистема)
mega-	megaword (мегаслово)
re-	гесору (повторно копіювати); recreate (відновлювати дані); reformat (переформатувати);
e-	e-mail (електронна пошта); e-book (електронна книга); e-cash (електронні гроші); e-form (електронна форма, електронний бланк).

Для вирішення різних лінгвістичних завдань мало лише наявності масиву текстів. Потрібно також, щоб тексти містили в собі різного роду додаткову лінгвістичну та екстралінгвістичну інформацію.

Розмітка полягає в приписуванні текстам і їхнім компонентам спеціальних міток: зовнішніх, екстралінгвістичних (відомості про автора і відомості про текст: автор, назва, рік і місце видання, жанр, тематика; відомості про автора можуть включати не тільки його ім'я, але також вік, стать, роки життя та багато іншого; це кодування інформації має назву метарозмітка), структурних (глава, абзац, речення, словоформа) і власне лінгвістичних, що описують лексичні, граматичні та інші характеристики елементів тексту.

Серед лінгвістичних типів розмітки виділяються [4]:

а) морфологічна розмітка. В іноземній термінології вживається термін *part-of-speech tagging (POS-tagging)*, дослівно - частиномовна розмітка. Насправді морфологічні мітки включають не тільки ознаку частини мови, але й ознаки граматичних категорій, властивих даній частині мови.

б) синтаксична розмітка, є результатом синтаксичного аналізу, або парсинга (англ. *parsing*), виконаного на основі даних морфологічного аналізу. Цей вид розмітки описує синтаксичні зв'язки між лексичними одиницями і різні синтаксичні конструкції (наприклад, підрядне речення, дієслівне словосполучення тощо).

в) семантична розмітка. Хоча для семантики немає єдиної семантичної теорії, найчастіше семантичні теги позначають семантичні категорії, до яких належить дане слово чи словосполучення, і вузчі підкатегорії, які специфікують його значення.



Розмітка корпусів - це трудомістка операція, особливо враховуючи розміри сучасних корпусів. Якщо деякі види розмітки, здійснюються вручну, то морфологічний та синтаксичний аналіз можливо здійснити автоматично [5].

Підсумковим завданням даної роботи є розробка алгоритму автоматичної розмітки тексту та вибору з тексту можливих термінів англійської комп'ютерної лінгвістики на основі корпусу текстів.

На першому етапі вручну виділяємо найпоширеніші префікси, суфікси та основи, за допомогою яких утворюються терміни комп'ютерної лінгвістики, з якими у подальшому будуть порівнюватися словоформи.

Нерозмічений лінгвістичний текст автоматично розбивається на окремі словоформи, які порівнюються з морфемами та основами, які найчастіше використовуються для словотворення термінів комп'ютерної лінгвістики.

Якщо збіг не знайдено, програма виводить на екран «Нажаль у цьому тексті немає термінів комп'ютерної лінгвістики.»

Якщо програма знайшла збіг, вона автоматично помічає його тегом `<ap></ap>`, який вказує, що це слово є терміном комп'ютерної лінгвістики.

На виході ми отримаємо розмічений текст, який можна занести до корпусу текстів, та надати можливість користувачеві переглянути імовірні терміни англійської термінології з комп'ютерної лінгвістики та дізнатися їх дефініцію.

Створення корпусів текстів значно полегшило збір і зберігання інформації. Це дуже цінується при створенні словників, глосаріїв, лексикографічних робіт. Також спосіб зберігання корпусів текстів дозволяє надійніше і довше зберігати будь-який мовний матеріал, що є важливою знахідкою та інструментом в лексикографії.

Список літератури

1. Баранов А.Н. Введение в прикладную лингвистику. - Эдиториал УРСС, 2001.
2. Довідково-інформаційний портал: http://www.gramota.ru/slovari/types/17_2
3. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики. - Донецьк, 2003
4. Апресян, Ю.Д., Иомдин, Л.Л., Санников А.В., Сизов, В.Г. Семантическая разметка в глубоко аннотированном корпусе русского языка. - Издательство Санкт-Петербургского университета, 2004.
5. Баранов, А. Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема. - Просвещение, 1998.
6. Труды Международного семинара по компьютерной лингвистике и ее приложениям «Диалог-2000», «Диалог-2001», «Диалог-2002», «Диалог-2003», «Диалог-2004», «Диалог-2005».



ЛІНГВІСТИЧНИЙ АСПЕКТ УКЛАДАННЯ ЕЛЕКТРОННОГО ПАРАЛЕЛЬНОГО СЛОВНИКА ФРАЗЕОЛОГІЗМІВ

Гончарова С.О.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: softabarvinok@gmail.com*

Комп'ютерні технології відіграють важливу роль у розв'язанні цілої низки проблем комп'ютерної лінгвістики, зокрема лексикографічних: розроблення комп'ютерних алгоритмів, програм, систем та технологій укладання та використання словників. Лексикографічні системи дають змогу формувати словникові статті, зберігати текстову, візуальну та звукову інформацію, здійснювати обробку словникової інформації (аналіз, пошук, фільтрування, відворення тощо).

У цьому зв'язку стає можливим розв'язання проблеми укладання електронних словників національно-культурного напрямку. До них належать фразеологічні словники, що відображають специфічні риси картин світу, репрезентованих у мовах різних народів. З огляду на це актуальним постає питання розроблення електронного паралельного російсько-українського та українсько-російського словника, що відображає спільні, відмінні та відсутні риси картин світу, які відбиваються у фразеологізмах аналізованих мов.

Метою нашої розвідки є створення тематичного паралельного двомовного словника фразеологізмів. Для досягнення мети необхідно: 1) дослідити лексичні, семантико-граматичні, експресивно-стилістичні риси фразеологізмів української та російської мов; 2) визначити випадки паралельностей між фразеологічними одиницями (ФО) досліджуваних мов; 3) схарактеризувати структуру та параметри репрезентації ФО у словниковій статті.

Джерельну базу дослідження складають 358 фразеологізмів: 274 – з «Фразеологічного словника української мови в 2-х томах» та 84 – з «Фразеологического словаря русского языка». Словник буде виконувати дві функції: тлумачну та перекладну.

В російській та українській мовах, як у мовах споріднених, дуже часто трапляються повні еквіваленти, але синонімічний ряд при цьому теж має місце. За нашими спостереженнями, між ФО української та російської мов відбуваються такі випадки паралельностей:

1) абсолютна, коли дві ФО збігаються повністю за своїм сигніфікативно-денотативним значенням та можлива наявність синонімічного ряду в обох мовах;

2) часткова, якщо семантика ФО частково відрізняється та можлива наявність синонімічного ряду в російській або українській мовах;

3) відсутня паралельність, що полягає у відсутності як еквівалента, так і синонімічного ряду.



Українські та російські ФО класифіковано за такими характеристиками: семантико-граматичною, структурною, експресивно-стилістичною. Вибір саме таких критеріїв класифікації зумовлений тим, що саме вони репрезентують ФО як із графічної, так і зі смислової сторони. Семантико-граматичний критерій показує яке синтаксичне місце займає обраний фразеологізм, яким членом речення виступає, яке синтаксичне наповнення несе. Експресивно-стилістичний показує до якого часу, стилю відноситься ФО, в якій ситуації доречно буде використати тій чи інший фразеологізм. Структурний графічно відображає фразеологічні одиниці. Функціональний демонструє особливості уживання ФО у живому мовленні, для чого використано цитатний матеріал з художньої літератури.

Стаття розроблюваного електронного словника складається з двох частин, кожна з яких репрезентує характеристики української та російської ФО (та навпаки). Українська та російська частина словникової статті відображає інформацію про ФО у такому порядку: 1) українська/російська ФО, 2) показник повної, часткової або відсутньої паралельності (відображається символами «=», «≈» та «∅»); 3) тлумачення, 4) семантико-граматична характеристика, 5) структурна модель; 6) експресивно-стилістична риса та 7) контекстне уживання. Перекладна функція представлена таким чином: до обраної ФО буде подаватися еквівалент(и) з паралельної мови; у разі відсутності повного еквівалента подається описовий переклад, тобто синонім або синонімічний ряд.

Наш словник буде призначений для широкого кола користувачів: наукових працівників, письменників, журналістів, редакторів видавництв, викладачів вузів, учителів, студентів, учнів та просто шанувальників українського слова.

Список літератури

1. ФСУМ 1999: Фразеологічний словник української мови : В 2-х кн. / Уклад. В. М. Білоноженко та ін. – К. : Наук. думка, 1999. – Кн.1: А-Н. – 528 с.; Кн.2: Н-Я. – С. 529-980.
2. Фразеологический словарь русского языка / Под ред. Л. А. Войновой, В. П. Жукова и др. – М.: Советская энциклопедия, 1967. - 543 с.
3. *Селегей В. П.* Электронные словари и компьютерная лексикография [Електронний ресурс]. – Режим доступу: http://www.lingvoda.ru/transforum/articles/selegey_a1.asp
4. *Широков В. А.* Інформаційна теорія лексикографічних систем. – К. : Довіра, 1998. – 331 с.; Широков В. А. Інформаційно-енергетичні трансформації та інформаційне суспільство / В. А. Широков // Українсько-польський науково-практичний журнал «Наука, інновація, інформація». – К. , 1996. – С. 48–66.
5. *Дубічинський В. В.* Українська лексикографія : історія, сучасність та комп'ютерні технології: навч. посібник / В. В. Дубічинський. – Х. , 2004.



MOODLE ВІРТУАЛЬНЕ НАВЧАЛЬНЕ СЕРЕДОВИЩЕ ДЛЯ КЕРУВАННЯ ДИСТАНЦІЙНИМ НАВЧАННЯМ

Мусієнко О.П.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail:msnp1@rambler.ru*

Дистанційне навчання – взаємодія вчителя та учнів між собою на відстані, що відображає всі притаманні навчальному процесу компоненти і реалізоване засобами інформаційних технологій, що передбачають інтерактивність. Дистанційний курс в дистанційному навчанні є основним засобом навчання.

Дистанційний курс – це комплекс навчально-методичних матеріалів та освітніх послуг, створених у віртуальному навчальному середовищі для організації дистанційного навчання на основі інформаційних і комунікаційних технологій за моделлю дистанційного навчання.

Дистанційний курс, розміщений у віртуальному навчальному середовищі, забезпечує:

- управління навчальним процесом та адміністрування;
- надання знань шляхом вивчення теоретичного матеріалу;
- самоконтроль;
- закріплення матеріалу;
- формування навичок і вмінь на основі отриманих знань;
- сумісну діяльність студентів у малих групах;
- синхронне та асинхронне спілкування;
- контроль за засвоєнням теоретичного матеріалу; виконання практичних завдань та їх контролю.

Moodle – безкоштовне, відкрите віртуальне навчальне середовище. Moodle дозволяє створювати курси і web-сайти, які базуються у мережі Internet.

Moodle перекладена на десятки мов, є і переклад на українську. Система використовується у 175 країнах світу.

Головним розробником системи є Martin Dougiamas з Австралії. Цей проєкт є відкритим, у ньому бере участь велика кількість інших розробників. Система знаходиться в постійному розвитку.

Moodle реалізує філософію «педагогіки соціального конструктивізму» та орієнтована насамперед на організацію взаємодії між викладачем та членами навчальної групи, хоча підходить і для організації традиційних дистанційних курсів, а також підтримки очного навчання.

«Педагогіка соціального конструктивізму» в даному разі означає що викладач менше заціклюється на звичайній публікації матеріалів, якими слухачі повинні оволодіти, та подальшим виставленням оцінок. Викладач з простого "джерела знань" перетворюється на натхненника, рольову модель культури навчальної групи, зв'язуючись зі слухачами в індивідуальному порядку і працюю-



чи за їх особистими потребами, одночасно з цим направляючи дискусії і діяльність всієї групи для досягнення загальних навчальних цілей

В умовах НТУ "ХПИ" режим використання Moodle визначається доступом до інтернету. В разі необмеженості доступу нормальним вважається в умовах щотижневих лекцій та лабораторних занять також щотижнєве тестування для закріплення відповідної теми. Імовірно, це дозволить кардинально підвищити рівень засвоєння знань. В разі наявності обмежень варто провести в комп'ютерному класі хоча б тест по першому і другому семестрових модулів. У будь-якому разі викладач дає доступ студентам до однієї спроби тесту, а для запобігання списуванню визначається достатньо жорстке обмеження тесту в часі.

Moodle дозволяє оперативно надавати студентам лекційні матеріали, інструкції до лабораторних робіт, поточні рейтинги, новини і оголошення; дозволяє оперативно надавати відповіді на студентські питання, причому відразу для всієї «віртуальної спільноти», пов'язаної з даним курсом; дозволяє оперативно інформувати студентів про результати перевірки завдань (зокрема, лабораторних робіт).

Студенти	Навчання
Викладач	Наповнення дисциплін
Адміністратори факультетів (кафедр)	Створення користувачів, дисциплін, методична
Адміністратор НТУ	Оновлення версій, українізація

Рис.1 Організація процесу дистанційного навчання у НТУ "ХПИ".

Враховуючи досвід інших учбових закладів, в разі всезагального використання Moodle в НТУ "ХПИ" видається доцільною така організація процесу дистанційного навчання (рис.1).

Список літератури

1. Теория и практика дистанционного обучения / Под ред. Е.С. Полат. – М., "Академия", 2004.
2. Зайченко Т.П. Основы дистанционного обучения: Теоретико-практический базис: Учебное пособие. - СПб.: Изд-во РГПУ им. А.И. Герцена, 2004.
3. Кухаренко В.М. Дистанційне навчання у схемах. Посібник. – Харків: НТУ «ХПИ», 2001.



РОЗРОБКА СПЕЦІАЛІЗОВАНОГО ПЕРЕКЛАДНОГО УКРАЇНСЬКО-АНГЛІЙСЬКОГО СЛОВНИКА З ТЕКСТИЛЬНИХ МАТЕРІАЛІВ

Стребкова О.О.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: strebkova@ro.ru*

Комп'ютерна лексикографія як галузь прикладної лінгвістики є особливим напрямком у практичній лексикографії зі своїми власними підходами не тільки відображення, але і до змісту словника. Отже, можна сказати, що «електронний словник - це особливий лексикографічний об'єкт, в якому можуть бути реалізовані і введені в обіг продуктивні ідеї, не затребувані з різних причин в паперових словниках». Електронні словники мають ряд очевидних і істотних переваг у порівнянні зі словниками традиційними. Сучасні електронні словники не тільки значно перевершують за обсягом книжкові, але і допомагають заощадити час. З технічної точки зору, електронний словник – комп'ютерна база даних, яка містить особливим чином закодовані словникові статті, які дозволяють здійснювати швидкий пошук потрібних слів (словосполучень, фраз).

Словники виконують інформативну та нормативну функції: вони універсальні інформаційні джерела для розуміння того чи іншого явища та найточніша консультація щодо мовних норм.

Основними критеріями у процесі створення словника є великий обсяг, раціональний, продуманий спосіб подачі словникового матеріалу, швидкий та зручний пошук потрібного слова. Створення кожного словника, незалежно від його типу і призначення, розпочинається ретельним відбором реєстрових слів (лексичних одиниць), які або витлумачуються, або перекладаються.

Базовою одиницею словника вважається словникова стаття – кожен окремий об'єкт опису словника і співвіднесені йому словникові характеристики. Сукупність словникових статей і становить основний текст словника (інакше базу даних словника).

Сучасні комп'ютерні технології істотно спростили процес збору і зберігання лексикографічної інформації, використовуючи базу даних, яка є аналогом каталожної картки. На відміну від традиційних методів, бази даних дають можливість автоматично сортувати інформацію, відбирати потрібні приклади і об'єднувати їх у різні групи. Після підготовки словникової картотеки – етап складання словникової статті. Словникова стаття вводиться в базу даних, яка стає основою для створення словника. Редагування статей також здійснюється в базі даних, що значно скорочує час розробки словника.

Структура словникової статті спеціалізованого перекладного українсько-англійського словника з текстильних матеріалів виглядає наступним чином:

Вхідне слово (українська лексема)



Заголовна форма слова: *Бавовна*;

Граматична інформація: [*баво́вна*], жін.;

Дефініція, містить словесний опис поняття, яке номінує термін (у разі потреби з елементами енциклопедичного пояснення).

1) поняття: *Волокно бавовнику, з якого виробляють пряжу і т. ін.*;

2) вид: *нитка, зерно, канат/мотузка, шнурок.*

Зона перекладних еквівалентів (англійська лексема)

Перекладні еквіваленти англійською мовою (передають однаковий обсяг інформації та є функціонально рівнозначними).

Заголовна форма слова: *Cotton*;

Граматична інформація: [*'kɒtn*];

Дефініція: 1) *A cloth or thread made from cotton fibres*;

2) *thread, grain, rope, string.*

Для створення бази даних слів електронного словника з текстильних матеріалів, була обрана реляційна СУБД MS Access. При розробці бази даних словника з текстильних матеріалів, була побудована таблиця яка налічує більше 100 слів.

Словник призначено для:

– перекладачів-філологів, які працюють із вузькоспеціальною технічною термінологією, пов'язаною з виробництвом та експлуатацією текстильних матеріалів;

– лексикографів, котрі створюють вузькоспеціалізовані термінологічні словники;

– студентів-філологів, що вивчають курс практичної лексикографії.

Список літератури

1. *Дорошенко С.І.* Загальне мовознавство. Підручник. – Вид. Київ, 2006 – 289 с.
2. Словотворення: [Електронний ресурс]. – Режим доступу: <http://ukr-mova.at.ua/index/slovotvorennja/0-7>
3. Лексикографія: [Електронний ресурс]. – Режим доступу: <http://lib.lntu.info/book/fof/ippy/2010/10-166/page8.html>
4. Предмет лексикографії. Типи словників: [Електронний ресурс]. – Режим доступу: <http://westudents.com.ua/glavy/10882--13-predmet-leksikograf-tipi-slovnkv.html>
5. Лексикографическая концепция: [Електронний ресурс]. – Режим доступу: <http://lab314.brsu.by/kmp-lite/kmp2/OTT/tLecture/tLexGraphCon.htm>
6. Структура словарной статьи: [Електронний ресурс]. – Режим доступу: <http://slovari.ru/default.aspx?s=0&p=2644>
7. Системы управления базами данных: [Електронний ресурс]. – Режим доступу: <http://www.lessons-tva.info/edu/e-inf2/m2t4.html>



РОЗРОБКА СИСТЕМИ ПЕРЕВІРКИ І ОЦІНЮВАННЯ РІВНЯ ЗНАНЬ СТУДЕНТІВ СПЕЦІАЛЬНОСТІ "ПРИКЛАДНА ЛІНГВІСТИКА"

Федосєєв П.І.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 096-123-31-88
e-mail: fedoseev_pavel@rambler.ru*

У сфері освіти комп'ютеру належить виняткова роль. Комп'ютерні системи оцінювання знань є важливим компонентом будь-яких систем дистанційного навчання [1].

У нинішній теорії ще немає усталеного підходу до визначення понять «оцінка», «контроль», «перевірка», «облік» та інших, з ними пов'язаних. Нерідко вони змішуються, взаємозміщуючись, вживаються то в однаковому, то в різному значенні. Загальним родовим поняттям виступає «контроль». *Контроль* – виявлення, вимірювання і оцінювання знань, умінь учнів. Виявлення і вимірювання називають перевіркою. Тому перевірка – складовий компонент контролю, основною дидактичною функцією якого є забезпечення зворотного зв'язку між учителем і учнями, отримання педагогом об'єктивної інформації про ступінь засвоєння навчального матеріалу, своєчасне виявлення недоліків та прогалин у знаннях. Перевірка має на меті визначення не лише рівня і якості навченості учня, а й обсягу навчання праці. Крім перевірки контроль містить у собі оцінювання (як процес) та оцінку (як результат) перевірки [2].

Системи автоматизованого тестування – система тестування стану знань і умінь того, кого навчають, у конкретній предметній області з засобами можливого удосконалення знань і умінь. У такій системі повинна відбуватися автоматична генерація завдань та автоматичний вибір тестових завдань, виходячи з різних цілей тестування, що вибирає викладач чи студент. Результатом тестування служить числова оцінка і модель поточних знань. Можуть видаватися також рекомендації з розділів, необхідних для повторного вивчення. Система тестування знань може використовуватися як у самостійній роботі студента так і в роботі студента під керівництвом викладача [3].

Системи комп'ютерного контролю знань – це програмні системи (системи тестування), що дозволяють проводити аналіз знань студентів за допомогою сучасних інформаційних технологій. Одна з переваг автоматизованих систем контролю знань у тому, що вони можуть використовувати складні методики подання завдань студентам, які називають стратегіями тестування [4].

Комп'ютерні системи автоматизованого тестування дають можливість студенту самостійно оцінити рівень своїх знань, а викладачу – систематизувати і спростити процес оцінки якості знань студентів.

Таким чином, застосування автоматизованого тестування для перевірки знань студентів не тільки полегшує роботу викладача, але і підвищує мотива-



цію навчальної діяльності студентів, одночасно знижуючи їхню емоційну напруженість у процесі контролю[5].

Доречним прикладом є система тестування OpenTEST, яка дозволяє:

- створення тестів закритого типу, їх редагування, експорт/імпорт у систему;
- проведення тестування через Інтернет;
- експертне оцінювання окремих питань або тесту в цілому.

Система OpenTEST відповідає багатьом вимогам сучасного web-сервісу тестування знань, але в ній наявна велика кількість недоліків, головні з яких:

- слабка, нерозвинена система оцінювання складності й фільтрування питань в тесті, яка базується на однобальній оцінці всіх питань;
- особиста відповідальність авторів тестів за коректність і балансування складності питань у тесті[6].

У даній роботі для створення системи тестування було обрано адаптивну модель подання запитань, бо вона є найбільш прогресуючою на сьогоднішній день. Адаптивні тести – запитання не просто ідуть у чітко визначеному порядку, але й залежать від того, яку відповідь дасть студент на попереднє запитання. Так можуть бути збудовані соціологічні опитування, психологічні тести Наприклад, у першому запитанні питають, який напій він вживає частіше, чай чи каву. Відповідно, для тих, хто вибере каву, наступне запитання буде: “мелену чи розчинну?”, а для тих, хто вибере чай – “чорний чи зелений?”. В тестах такого типу запитання теж показуються по одному. Тільки після того, як студент дасть відповідь на запитання, він зможе побачити наступне[7].

Процедура зміни складності завдання:

Складність завдань зменшується, якщо:

- складність вища за «дуже легко» (тобто її можна зменшити);
- відповідь на питання є невірною;

Аналогічно зі збільшенням складності:

- складність нижче рівня «дуже складно» (тобто її можна збільшити);
- відповідь на питання є вірною.

Список літератури

1. Комп'ютерні технології в тестуванні – [Електронний ресурс]: http://moodle.ndu.edu.ua/file.php/1/Fetisov_komp_tehnol_v-testuvanni.pdf
2. Гершунский Б.С. «Компьютеризация в сфере образования» – М.: Педагогика, 1987. – 264 с.
3. Пишємо систему автоматизованих тестів – [Електронний ресурс]: http://www.protesting.ru/automation/practice/automation_from_scratch.html
4. Розробка комп'ютерної системи тестування – [Електронний ресурс]: http://masters.donntu.edu.ua/2011/fknt/semenova/diss/index_ua.htm
5. Клейман Г.М. «Школы будущего: компьютеры в процессе обучения» – М.: Радио и связь. – 1987. – 176 с.
6. Система тестування OpenTEST – [Електронний ресурс]: <http://radnuk.info/statti/563-putanya/14753-2011-01-19-02-03-34.html>
7. Майоров А.Н., Теория и практика создания тестов для системы образования: Учеб. пособие - М.: Интеллект-центр, 2001.- 296 с.



ДИНАМІЧНЕ ФОРМУВАННЯ СТОРІНКИ WEB-САЙТУ НА ОСНОВІ ВИКОРИСТАННЯ ОНТОЛОГІЇ ПРЕДМЕТНОЇ ОБЛАСТІ

Піскова Н.О.

*Національний технічний університет
«Харківський політехнічний інститут»
Харків, вул. Пушкінська, 79/2, тел. 096-123-31-88
e-mail: piskova.n@mail.ru*

Сьогодні стрімко розвиваються веб-орієнтовані інформаційні системи, зокрема інформаційні портали, системи електронної комерції і т.д. Розвиваються засоби і технології для програмування веб-додатків як розподілених клієнт-серверних (JSP, .NET і т.п.).

Всі сайти в мережі можна розділити на дві великі групи: статичні і динамічні сайти. Статичної прийнято називати сторінку, яка цілком зберігається на сервері і показується відвідувачеві у своєму незмінному вигляді (статична сторінка може містити деякі змінювані елементи, наприклад банери, проте вона все одно залишається статичною) [1].

Динамічна сторінка – веб-сторінка, яку згенеровано програмно на відміну від статичної сторінки, які є файлами, що лежать на сервері. Сервер генерує HTML код для обробки браузером або іншим агентом користувача. Динамічні сторінки зазвичай обробляють і виводять інформацію з бази даних. Вміст і зовнішній вигляд таких сторінок змінюється залежно від інформації, яка міститься в базі даних.

Для розробки динамічних сайтів використовуються такі технології:

– PHP (докладніше: www.php.net) - ця технологія, що має свою мову програмування, використовується практично на всіх операційних системах [2];

– ASP (Active Server Pages) – технологія від лідера світового програмування компанії Microsoft, використовує мови програмування VBScript, JScript, застосовується тільки на серверах, що працюють під операційною системою ОС Windows;

– бібліотека ARC була написана в 2007 році для використання в мові PHP. Вона була створена для написання динамічних сторінок, а точніше для вставки в них динамічного контенту. Дана бібліотека була створена для обробки онтологій, як у вигляді RDF, так і у форматі OWL. Бібліотека працює із запитамі SPARQL, і виводить отриманий результат в динамічну частину сайту.

У бібліотеці ARC розробники намагаються тримати речі простими і гнучкими. Всі функції з підтримки практичних випадків використання, не реалізують специфікації заради повного дотримання. Насправді, один з мотивів ARC навіть не створення більш семантичних додатків, а перенесення RDF в повсякденні web-розробки [3].

Термін "онтологія" вперше з'явився в роботі Томаса Грубера в 1993 році, в якій розглядалися різні аспекти взаємодії інтелектуальних систем між собою і с людиною. Інтелектуальними системами називаються програми, які моделю-



ють деякі аспекти інтелектуальної діяльності людини. Онтологія являє собою опис декларативних знань, призначене для читання людиною і виконане у вигляді класів з відношенням ієрархії між ними. Складання опису декларативних знань вимагає великої роботи і певних навичок. Для позначення цієї роботи, а також її результату, Грубер ввів спеціальний термін – «концептуалізація». Опис він назвав «специфікацією». Таким чином, онтологія по Грубер визначається як «специфікація концептуалізації».

Для машинного представлення різних предметних областей в Internet, використовуються онтології. Онтологія – специфікація концептуалізації, або явне, формальний опис предметної області. Як і в об'єктно-орієнтованому описі галузі знань, онтологія складається з класів та їх примірників, у класів і екземплярів виділяються властивості, на властивості можуть накладатися логічні обмеження.

Онтології вмісту Web-сторінок необхідні пошуковим програмам для поліпшення якості пошуку по Web. Ідея побудови специфікацій концептуалізацій вмісту Web-сторінок знаходиться в основі концепції так званого Розумного Web або Semantic Web. Semantic Web являє собою наступне покоління World Wide Web, в якому окрім гіпертекстових документів містяться описи семантики цих документів, а також опису семантики різних сервісів, що надають ці документи кінцевим користувачам [4].

Формальна специфікація вмісту Web-документа дає можливість пошуковій програмі робити висновки про відповідність пошукового запиту даному Web-документу не тільки на основі синтаксичної інформації, одержуваної з тексту цього документа, але і ґрунтуючись на семантиці вмісту даного документа. Це може кардинально поліпшити якість Web-пошуку, так як опис світу Web-сторінки, ясна пошуковій програмі, дає останньої набагато більше інформації, ніж вона може отримати з неструктурованого тексту.

Ідеї розумного Web давно були сприйняті спільнотою W3, в результаті чого вже протягом більше десяти років ведуться роботи по втіленню цих ідей у життя. Першим завданням, яке необхідно вирішити для цього, є розробка стандартного мови, який був би зрозумілий усім пошуковим програмам.

Онтології використовуються науковими спільнотами – для опису термінології, в електронній комерції – для опису товарів і послуг, і в інших додатках Internet. Через своєї популярності онтології стали використовуватися і в якості баз знань локальних інтелектуальних систем.

В Internet використовується безліч мов представлення даних, засновані на XML. В рамках проекту Semantic Web консорціумом W3 розроблено ряд специфікацій для роботи з машинним представленням знань. Зокрема, для представлення даних, що мають графову структуру, на основі XML, розроблена мова RDF (Resource Definition Framework – середа Описи Ресурсу). RDF надає засоби для запису триплетів, трійок даних – суб'єкт – предикат – об'єкт.

Для опису онтологій, доступних через Веб, створені мови RDFS (RDF Schema – RDF Схема) і OWL (Ontology Web Language – Мова Мережевих Онтологій). В якості своїх базових елементів дані мови використовують RDF ресурси. RDFS використовується для запису словників, а OWL – онтологій. Ме-



режеві онтології надають більш виразні можливості в порівнянні з RDF словниками, наприклад логічні операції над класами і логічні обмеження властивостей.

На сьогоднішній день знайти цікаву інформацію, що не переглянувши величезну кількість непотрібних сторінок сайтів, практично неможливо. Для вирішення цієї проблеми і були створені методи динамічного формування сторінок web-сайтів. При верстці будь-якого сайту виникає одна проста проблема, яка, незважаючи на свою простоту, потребує вирішення. Це – проблема статистичного контенту.

Список літератури

1. Стандарти W3C консорциума World Wide Web Consortium - Web Standards [Електронний ресурс] - Режим доступу: [www/ URL http://www.w3.org](http://www.w3.org) - 20.05.2007 - Загл. с екрана.
2. Semantic Web на PHP при помощи библиотеки ARC [Электронный ресурс] - Режим доступа:<http://sheremetov.com/php/semantic-web-php-arc/>- Загл. с екрана.
3. Library ARC [Электронный ресурс] – Режим доступа: <http://www.w3.org/2001/sw/wiki/ARC> - Загл. с екрана.
4. Глибовець М.М., Олецкий О.В. Штучний інтелект. – К. :Видавничий дім “КМ Академія”, 2002. – 366 с.



ПРИНЦИПЫ РАБОТЫ СИСТЕМ МАШИННОГО ПЕРЕВОДА

Колесник А.С.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707-63-60,
e-mail: kolesnik_nastya23@mail.ru*

Машинный перевод – это выполняемое на компьютере действие по преобразованию текста на одном естественном языке в эквивалентный по содержанию текст на другом языке, а также результат такого действия.

Современный машинный, или автоматический перевод осуществляется с помощью человека: пред-редактора, который тем или иным образом предварительно обрабатывает подлежащий переводу текст, интер-редактора, который участвует в процессе перевода, или пост-редактора, который исправляет ошибки и недочеты в переведенном машиной тексте.

Для осуществления машинного перевода в компьютер вводится специальная программа, которая реализует алгоритм перевода, под которым понимается последовательность однозначно и строго определенных действий над текстом для нахождения переводных соответствий в данной паре языков при заданном направлении перевода (из одного конкретного языка в другой).

Извлечение информации из текста производится на основании набора атрибутов: морфологических, синтаксических, лексических, семантических и т.п. Атрибуты не указаны в тексте в явном виде, их нужно предварительно получить. Для этого производятся различные виды анализа текста с целью выделения атрибутов, используемых алгоритмом извлечения информации. Анализ, как правило, носит многоуровневый характер и выполняется модулем лингвистического процессора. Обычно выделяют следующие составляющие анализа текста:

- графемный анализ (выделение слов и предложений);
- морфологический анализ;
- синтаксический анализ;
- семантический анализ;
- построение модели предметной области (сценария или ситуации).

На каждом уровне фрагментам текста сопоставляются новые атрибуты. На основании таких наборов атрибутов алгоритм извлечения информации выполняет поиск фрагментов текста, релевантных цели. Естественно, не всегда нужно использовать все уровни текста в полном объеме. Все зависит от предметной области, информации, которую нужно извлечь, источников информации, а также точности и полноты, с которой эту информацию нужно извлекать.

Графематический анализ - это программа начального анализа естественного текста, вырабатывающая информацию, необходимую для дальнейшей морфологической и синтаксической обработки. В задачу графемного анализа входят:

- разделение входного текста на слова, разделители и т.д.



- сборка слов, написанных в разрядку;
- выделение устойчивых оборотов, не имеющих словоизменительных вариантов;
- выделение фамилии, имени и отчества, когда имя и отчество написаны инициалами;
- выделение электронных адресов и имен файлов;
- выделение предложений из входного текста;
- выделение абзацев, заголовков, примечаний.

Алгоритмы **морфологического анализа** делятся на две группы: словарные и бессловарные. Бессловарные алгоритмы более компактны и производительны, но не обладают высокой скоростью, поэтому их применение целесообразно лишь для выявления простых морфологических атрибутов и только в том случае, если нет требования к высокой точности. Если же предполагается использовать синтаксический анализ, то высокая точность является необходимым требованием, и применяется словарный метод.

Словарный метод предполагает наличие словаря основ и флексий. По словарю отыскиваются допустимые наборы атрибутов для каждой графемы. В случае отсутствия слова в словаре, выполняется предсказание парадигмы (аналогично бессловарным методам). Одной графеме может соответствовать несколько наборов атрибутов. Такие случаи - морфологическая омонимия - довольно часто встречаются в русском языке. Существуют алгоритмы для решения этой проблемы с высокой вероятностью успеха.

Целью **синтаксического анализа** является построение синтаксических групп на одном морфологическом варианте одной клаузы, т.е. одного простого предложения в составе сложного.

Целью компьютерной лингвистики в области синтаксиса является построение автоматизированного анализатора отдельного языка. Этот анализатор должен уметь выделять простые предложения в составе сложного, устанавливать связи между словами и по возможности строить полное синтаксическое дерево предложения.

Чтобы разрешить проблему с анализом синтаксически омонимичных конструкций, необходимо построение дерева синтаксических зависимостей между словами во фразе. В случае удачного разбора предложение сворачивается в полносвязное дерево с единственной корневой вершиной.

Поскольку одна словоформа может соответствовать нескольким грамматическим формам слова, в том числе формам различных слов, в ходе анализа необходимо производить свертку предложения для всех возможных вариантов грамматических форм. Те грамматические формы, которые обеспечивают максимальную свертку дерева (минимальное число висячих вершин), следует считать наиболее достоверными.

Семантический анализ строит семантическую структуру одного предложения. Семантическая структура состоит из семантических узлов и семантических отношений. Семантический узел - это такой объект текстовой семантики, у которого заполнены все валентности, как эксплицитно выраженные в тек-



сте, так и имплицитные - те, которые получаются из экстралингвистических источников. Из определения следует, что семантический узел может быть построен только в самом конце семантического анализа. Собственно говоря, главная цель семантического анализа - построение семантических узлов, которое подразумевает заполнение всех валентностей.

Семантический анализ представляет собой выявление в тексте смысловых связей и групп. Этот тип анализа представляется в виде набора составляющих, направленных на выявление различных семантических связей. Во-первых, это выделение именованных сущностей, объектов, которые имеют различную форму записи в тексте и могут принимать различные значения. Второй полезной составляющей является механизм выявления семантических классов. К семантическому классу относится группа понятий, связанных с одной предметной областью и являющихся одной и той же частью речи. Третий момент связан с расширением кореферентности в тексте. Под кореферентностью понимается ссылка разными словами на один и тот же объект действительности. Четвертым элементом семантического анализа является разрешение анафоры. Анафора - это использование языковых выражений, которые могут быть интерпретированы лишь с учетом другого, как правило, предшествующего фрагмента текста. Разрешение анафоры сводится к установлению связи между анафорическим выражением и его интерпретацией (антецедентом).

Построение модели предметной области

Наиболее сложным, но и приносящим наиболее точные результаты этапом является построение модели ситуации или предметной области, которая описывается в тексте. Этот этап реализует представление в структурном виде, отражающем все значимые смысловые связи, всего текста или набора текстов. Но так как задача построения модели очень сложна, в прикладных системах редко прибегают к ее использованию.

Перевод был важен всегда. Научный прогресс дошел до изобретения машинного перевода, который во многом облегчил жизнь переводчикам. Конечно, и сейчас, существует огромное количество недостатков и в таком, казалось бы, совершенном изобретении. Но мы должны приложить все усилия, чтобы развивать машинный перевод.

Список литературы

1. Сокирко А. Будущее машинного перевода. // Компьютерра. - 2002, №21.
2. Кузнецов П. С., Ляпунов А. А., Реформатский А. А. Основные проблемы машинного перевода. Вопросы языкознания, 1956, № 5.
3. Машинный перевод [Электронная статья]. - [Режим доступа]: <http://study-english.info/article065.php>



ЭЛЕКТРОННАЯ И ТРАДИЦИОННАЯ ЛЕКСИКОГРАФИЯ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ

Гайденко Т.В.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,
e-mail: gaudenko.ua@mail.ru*

Электронная лексикография – наиболее динамично развивающееся направление современной прикладной лингвистики, целью которой является создание электронных словарей. Появление электронных лексикографических продуктов – это требование нового времени: в эпоху глобальной компьютеризации необходима принципиально новая форма подачи информации, учитывающая особенности компьютера. В связи с развитием компьютерных технологий возрастает потребность в словарях активного типа (которые содержат активную лексику). В них нуждаются переводчики и люди, изучающие иностранный язык. Словари на печатной основе не имеют возможности снабдить каждое значение слова примером употребления. Существенным преимуществом электронных переводных словарей является использование аудиовизуальных средств обучения, т.е. иллюстраций, видеороликов, аудиофрагментов. Термин «электронный словарь» стал уже привычным. Широкое использование электронных словарей связано с существующими преимуществами этих словарей над традиционными. Электронный словарь удобнее, практичнее и быстрее в использовании, чем его бумажная версия. Появилась целая научная дисциплина, занимающаяся теорией и практикой составления таких словарей, – компьютерная лексикография.

Традиционная практическая лексикография страдает от двух фундаментальных противоречий, характерных для этой области деятельности.

1. Чем больше объем словаря, чем полнее и доказательнее описание лексических значений, тем сложнее им пользоваться. Характерный пример – 20-томный Оксфордский словарь.

2. Чем полнее и глубже описание лексических значений, тем в меньшей степени словарь соответствует текущей языковой и культурной ситуации. Чрезвычайно долгий цикл создания и модификации фундаментальных бумажных словарей приводит к тому, что образ мира, который они фиксируют в системе своих значений, примеров и переводов, уже заметно отличается от действительности. Многие словари, основной корпус статей которых сформировался в языковой атмосфере середины века, представляют собой лексикографические музеи.

Компьютерная реализация бумажного словаря сама по себе позволяет преодолеть часть указанных проблем. Можно рассмотреть следующие преимущества электронных словарей.

1. Существенно более «изошренные» возможности показа содержания словарной статьи, включая возможность частичного показа по разным критериям



(различные "проекции" словаря), разнообразные графические средства, которые не используются в обычных словарях.

2. Использование для доступа к содержанию различных лингвистических технологий, таких как морфологический и синтаксический анализ, полнотекстовый поиск, распознавание и синтез звука и т.п.

С точки зрения пользователя смысл реализации в электронном словаре всех этих технологий состоит в том, что становится возможным быстро получить информацию, которая содержится где-то «в недрах» словаря и непосредственно отвечает тому запросу, который сформулирован пользователем в удобной для него форме.

Также преимуществами электронных словарей являются возможность нетрадиционного оформления словарной статьи, использование для доступа к содержанию различных лингвистических технологий, актуальность, новизна, экономичность, включение пользовательских словарей, переключение направления перевода.

К наиболее известным электронным переводным словарям относят *Abbyu Lingvo*, *Мультилекс*, *Контекст*, *Multitran*, *VER-Dict*, *Polyglossum*, *Slovoed*. Далеко не по всем научно-техническим направлениям существуют переводные русско-английские словари.

Учитывая недостатки существующих словарей необходима новая структура учебного электронного двуязычного словаря, основной целью которого должен стать быстрый контекстный перевод. На основе этого был разработан прототип переводного русско-английского словаря, который обеспечивает перевод, добавление, удаление и поиск слов.

Список литературы

1. *Кочерган М.П.* Вступ до мовознавства: Підручник для студентів філологічних спеціальностей вищих навчальних закладів. – Київ: Видавничий центр «Академія», 2001. – 368 с. (Альма-матер)
2. *Карпова О.М.* Электронные словари и кибернетическая лексикография: Методические рекомендации. – Иваново: Издательство «Ивановский государственный университет», 2002. – 24 с.
3. *Потапова Р. К.* Новые информационные технологии и лингвистика: Научная литература. — М.:МГЛУ, 2002. — 575 с.
4. *Нелюбин Л. Л.* Перевод и прикладная лингвистика: Научная литература. — М.: Высшая школа, 1983. — 208 с.
5. *Баранов А.Н.* Введение в прикладную лингвистику: Научная литература. — Эдиториал УРСС, 2001. — Р. 360.



ПОСТРОЕНИЕ ЧАСТОТНОГО СЛОВАРЯ СЛОВОФОРМ НЕСКОЛЬКИХ ТЕКСТОВ

Клименкова Е.Г.

*Национальный технический университет
"Харьковский политехнический институт",*

г. Харьков, ул. Пушкинская, 79/2, тел. 707-63-60, e-mail: klim789@bk.ru

В современном мире, как научная сфера, так и повседневная жизнь людей невообразима без автоматизированных информационных технологий. На протяжении последних десятилетий их значение стремительно растет. Появление ПК и быстрое развитие кибернетических идей укрепило надежды исследователей-лингвистов в том, что современные точные науки (и прежде всего математика) помогут лингвистике обрести недостающую ей точность. Появилась возможность автоматизировать многие трудоемкие процессы, например, статистическую обработку текстов, ведение разнообразных словарных и лексических картотек. Но с появлением компьютеров почти сразу возникла проблема общения с ними неподготовленных пользователей. Наилучшей формой для таких пользователей мог быть привычный естественный язык. Но для организации такого взаимодействия надо прежде понять законы и особенности использования естественного языка в процессе общения людей между собой [1].

В настоящее время поиск решения проблем автоматической обработки текстовой информации на естественном языке представляет особый интерес. И это объясняется не только тем, что естественный язык является инструментом мышления и общения между людьми, но и тем, что естественный язык — это универсальное средство накопления, хранения, обработки и передачи информации.

Словари играют большую роль в современной культуре, в них отражаются знания, накопленные обществом на протяжении веков. Они служат целям описания и нормализации языка, содействуют повышению правильности и выразительности речи его носителей. Словарь — справочная книга, содержащая собрание слов (или морфем, словосочетаний, идиом и т. д.), расположенных по определенному принципу, и дающая сведения об их значениях, употреблении, происхождении, переводе на др. язык и т. п. (лингвистические словари) или информацию о понятиях и предметах, ими обозначаемых, о деятелях в каких-либо областях науки, культуры и др [2].

В большинстве словарей описывается семантическая структура слов, т. е. словам сопоставляются объяснения их значений и употребления. Термин "словарь" используют для обозначения всей совокупности слов того или иного языка (другими словами, его лексику) и противопоставляют термину "грамматика", который обозначает совокупность правил построения из слов осмысленных речевых отрезков. Лингвистическая наука, которая занимается разработкой методов составления словарей и их изучением, называется лексикографией. В последние десятилетия в рамках лексикографии складывается новое направление — лексикографическая статистика. Лексикографическая статистика занима-



ется созданием частотных словарей и решает связанные с этой задачей вопросы теории и методики создания такого словаря [2].

Частотный словарь — вид словаря, в котором лексические единицы характеризуются с точки зрения степени их употребительности в совокупности текстов, представительных либо для языка в целом, либо для отдельного функционального стиля, либо для одного автора. Словарь может быть отсортирован по частоте, по алфавиту (тогда для каждого слова будет указана его частота), по группам слов и т. д. Частотные словари могут строиться на основе словоформ, лемм (нормальных форм слова) или словосочетаний [3].

Обычно частотные словари строятся не для одного текста, а для корпусов текстов. То есть, берется набор текстов из определенной предметной области или представительный для языка в целом, для конкретного функционального стиля речи, для творчества конкретного автора, и из него извлекаются словоформы, части речи, словосочетания или основы слов. По данным частотных словарей выделяются слова с высокой частотностью и низкочастотные слова. Это позволяет выявить ядро и периферию лексики, разграничить активный и пассивный запас, определить стилистическую принадлежность и жанровую приуроченность лексики, ее социально-возрастное расслоение. Данные о частотности употребления необходимы, например, при установлении авторства текста. Важны данные о наиболее частотных словах и при разработке компьютерных программ проверки орфографии. Частотный словарь также используется для создания эффективных методик обучения языку [3].

Проблемы при создании частотных словарей следующие: воспроизводимость (будут ли результаты идентичны на другом аналогичном корпусе); всплеск частоты отдельных слов (частота слова в одном тексте может повлиять на его позицию в частотном списке); сложность определения позиции менее частотных слов, что не дает возможности ранжировать их рационально; естественно, что практически в любом наборе текстов на первых местах по встречаемости будут служебные слова — союзы, предлоги и т. д.

При разработке прототипа системы, предназначенной для создания частотного словаря словоформ нескольких текстов, использовался метод подсчета слов. На его основе подсчитывается количество использования каждого слова и составляется словарь. Такую программу может использовать преподаватель иностранного языка для составления списка слов, которые необходимо знать его ученикам для изучения какой-либо темы. Также, такая программа может помочь человеку, который самостоятельно изучает иностранный язык. Можно составить словарь до начала чтения книги на другом языке, затем изучить наиболее встречаемые слова, что поможет лучшему пониманию текста.

Список литературы:

1. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. / Г.Г.Белоногов, Ю.П. Калинин, А.А. Хорошилов. - М.: Русский мир, 2004. -248 с.
2. Баранов А. Н. Введение в прикладную лингвистику - М.: УРСС Эдиториал. - 2001. -360 с.
3. Марчук Ю.Н. Компьютерная лингвистика. -АСТ.: Восток-Запад. -2007. -320с.



О МЕТОДАХ СОКРЫТИЯ ИНФОРМАЦИИ

Чухненко М.В.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60*

Под сокрытием информации (Information Hiding) обычно понимают методы, позволяющие скрывать некоторую дополнительную информацию, являющуюся или не являющуюся секретной, в некоторой, не привлекающей внимания информации.

Необходимость скрывать содержание важных сообщений существует уже тысячи лет. Со временем люди находили все более и более сложные способы кодирования сообщений, поскольку простые способы кодировки декодируются с большей легкостью. Коды и шифры не являются синонимами, как многие думают. Код – это совокупность знаков (символов) и система определенных правил, при помощи которых информация может быть представлена (закодирована) в виде набора из таких символов для передачи, обработки и хранения. Шифр – это совокупность условных знаков (условная азбука из цифр или букв) для секретной переписки или для передачи текста секретных данных по техническим средствам связи. Таким образом, при кодировании каждое слово в сообщении заменяется кодовым словом или символом, в то время как при шифровании каждая буква в сообщении заменяется буквой или символом шифра.

Защита информации необходима для уменьшения вероятности утечки (разглашения), модификации (умышленного искажения) или утраты (уничтожения) информации, представляющей определенную ценность для ее владельца. Проблема защиты информации от несанкционированного доступа возникла еще в древние времена, и с тех пор выделилось два основных направления решения этой проблемы, которые существуют и сегодня: криптография и стеганография.

Стеганография является более древней, чем коды и шифры, и обычно её называют тайнописью, искусством скрытой записи. Лучший способ тайнописи – это использование обычных предметов для сокрытия сообщения. В Англии был популярен метод тайнописи, для которого использовали обычную газету с крошечными точками под буквами на первой странице, которые указывали, какие буквы следует читать, чтобы получить сообщение. Некоторые люди могли составить сообщение, используя первые буквы каждого слова в каком-либо тексте или используя невидимые чернила. Стеганографию лучше всего использовать в сочетании с кодом или шифром, так как существует риск, что тайное послание может быть обнаружено.

Среди классических методов стенографии можно выделить следующие: манипуляции с носителем информации (контейнером); симпатические чернила; литературные приемы; семаграммы. Компьютерные технологии придали новый



импульс розвитку и совершенствованию стеганографии, появилось новое направление в области защиты информации - компьютерная стеганография.

Современный прогресс в области глобальных компьютерных сетей и средств мультимедиа привел к разработке новых методов, предназначенных для обеспечения безопасности передачи данных по каналам телекоммуникаций и использования их в необъявленных целях. Эти методы, учитывая естественные неточности устройств оцифровки и избыточность аналогового видео или аудио сигнала, позволяют скрывать сообщения в компьютерных файлах (контейнерах).

Криптография - наука о методах обеспечения конфиденциальности (невозможности прочтения информации посторонним) и аутентичности (целостности и подлинности авторства, а также невозможности отказа от авторства) информации. Традиционная криптография образует раздел симметричных криптосистем, в которых шифрование и расшифровка проводится с использованием одного и того же секретного ключа. Помимо этого раздела современная криптография включает в себя асимметричные криптосистемы, системы электронной цифровой подписи, хеш-функции, управление ключами, получение скрытой информации, квантовую криптографию[1].

Шифрование — способ преобразования открытой информации в закрытую и обратно. Применяется для хранения важной информации в ненадёжных источниках или передачи её по незащищённым каналам связи. Существует множество разных шифров. Самые известные шифры это: шифр Атбаш, шифр Френсиса Бекона, шифр Блеза Виженера, шифр Лестера Хилла, шифр Тритемиуса, шифр Гронсфельдаза, шифр Цезаря[2].

Современный уровень развития компьютерной техники позволяет обрабатывать информационные потоки в реальном времени. На основе криптографических методов и средств реализуются процедуры шифрования и дешифрования; системы идентификации и аутентификации, реализуемые на основе асимметричных криптографических стандартов (цифровая подпись).

Использование криптографии особенно важно для разработки протоколов передачи информации в глобальных компьютерных сетях для надежной защиты передаваемых данных от незаконного перехвата. Протоколы, разработанные с учетом криптографических методов и средств защиты, позволяют обеспечить требуемую секретность передаваемой информации, т.к. задача расшифровки перехваченных данных потребует от противника высоких вычислительных и материальных затрат, как правило, несоизмеримых с ценностью этих данных.

Список литературы

1. Яценко В.В. Введение в криптографию. [Электронный ресурс] –СПб.: Питер. –2001. ISBN: 5-900916-40-5
2. Аграновский А. В., Хади Р. А. Практическая криптография: алгоритмы и их программирование. [Электронный ресурс] –М.: Солон-Пресс. –2009.



ДО ПРОБЛЕМИ ІНШОМОВНИХ ЗАПОЗИЧЕНЬ ТЕРМІНІВ

Верховець М.В.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: verkhovetsmargo96@mail.ru*

Запозичена лексика є невід'ємним складником української літературної мови. Так само, і в науково-технічній термінології простежуємо функціонування значної кількості термінів іншомовного походження. Адже термінологія – це відкрита система, яка відображає міжнародний характер науки і техніки. Тому в будь-якій терміносистемі неможливо обійтись без іншомовних назв [1].

В українському термінознавстві існує дві тенденції щодо доцільності вживання термінів: одні термінологи вважають, що запозичення та інтернаціоналізми – необхідне й корисне надбання термінології, інші – що термінологія має бути витворена на національному ґрунті, тобто генетично українською.

Дослідники інтернаціоналізації терміносистем української мови аргументують свою позицію тим, що наука це міжмовна сфера, тому основа наукового стилю – термінологія – за своєю сутністю має бути інтернаціональна. Терміни відображають систему незалежних від конкретної мови наукових понять, дефініції термінів устанавлюються й суворо детермінуються терміносистемами, безвідносними до світогляду того чи іншого народу. Зростання кількості інтернаціоналізмів у сучасній мові науки і техніки відображає тенденції міжнародної співпраці, об'єднання зусиль учених і спеціалістів із метою високоефективного обміну досвідом [2].

Інші термінологи вважають, що термінологія хоч і є специфічним різновидом мови, проте вона є складовою частиною лексичного складу національної мови, з потужними внутрішніми словотворчими ресурсами, а кількість запозичень не дає взагалі підстав говорити про існування окремої метамови [2].

Запозичення може бути повним чи частковим. При повному запозичується як внутрішня, так і зовнішня форма терміна. При частковому зберігаються деякі фонетичні і граматичні особливості мови-продуцента.

Існують два шляхи термінологічних запозичень: прямий (безпосередньо з мови-продуцента) та непрямий (через посередництво інших мов) [3].

Важливим аргументом, що пояснює активне вживання термінів іншомовного походження в українській мові є те, що «базова термінологія будь-якої галузі науки за походженням давньогрецька або латинська. А свідоме уникнення запозичень у термінології неодмінно призведе до труднощів перекладу наукових праць іншими мовами й поступово спричинить замкнутість науки на себе. Але позичати варто лише те слово, яке вносить у певну галузь науки і техніки нове поняття і за умови, що цим терміном послуговуються науковці різних країн [4].

Таким чином, розглядаючи питання проблеми запозичення нових термінів з інших мов, слід з'ясувати причину цього явища.



По-перше, запозичуємо термін з іншої мови тоді, коли в національній мові немає (не було) відповідного денотату. Прикладами таких запозичених слів є такі слова, як «футбол» і «комп'ютер». Ці терміни були запозичені через відсутність таких предметів дійсності. Проте з часом вони увійшли в нашу мову й засвоїлись нею.

По-друге, в національних та іншомовних відповідниках є певні розрізнення у значенні, тому при запозиченні нових термінів слід уникати надлишковості, аби не обтяжувати мову іншомовними словами, синоніми яких вже існують. Наприклад, слова «оплески» і «аплодисменти» є синонімами, але розрізняються за сферою вживання: перше слово є більш характерним для побуту, а друге – для документів.

По-третє, запозичення відбувається тоді, коли слово в національній мові обростає багатьма конотаціями, тоді ми запозичуємо «чистий» термін для полегшення користування ним. Цей новий термін не переобтяжений національно-культурними особливостями його розуміння. Ось, слова «пристосування» й «адаптація» схожі за значенням, але перше має більш глибоку семантичну структуру, тому широко вживається в різних галузях і стилях мовлення; на відміну від другого, яке несе конкретний сигніфікат – пристосування до мінливих умов навколишнього середовища, і вживається переважно в біології.

Отже, уникнути запозичень у науково-технічній термінології, як і в літературній мові загалом неможливо, бо частина з них традиційно функціонує у різних терміносистемах і зрозумілі для більшості користувачів. Зважаючи на велику кількість українських відповідників, які пропонували у різні часи українські термінознавці, не варто їх одразу відкидати, бо вже існує термін іншомовного походження для номінації науково-технічного поняття. У тому разі, коли питомий термін має те саме значення, що й запозичений, його варто закріпити як єдиний нормативний варіант або як синонім поряд із терміном іншомовного походження. Якщо запропоновані паралельні назви, насправді, мають значні відмінності у значенні, їх варто вважати різними термінами. Національна технічна термінологія повинна увібрати в себе надбання певної науки (терміносистему), залишаючись національно самобутньою [5].

Список літератури

1. Мартиняк О. А. До проблеми запозичень в українській науково-технічній термінології / О. А. Мартиняк: [Електронний ресурс]. – Режим доступу: <http://lib.chdu.edu.ua/pdf/naukrpraci/movoznavtvo/2009/105-92-14.pdf>
2. Волошина Т. М. Проблеми аналізу кібернетичної термінології в контексті історії термінознавства / Т. М. Волошина // Актуальні проблеми української лінгвістики: теорія і практика / відп. ред.: Л. І. Шевченко . – Київ : Київський нац. ун-т ім. Т. Шевченка, 2003. – Вип. 8. – С. 83-89.
3. Васенко Л. А. Фахова українська мова : навч. посіб. / Л. А. Власенко, В. В. Дубічинський, О. М. Кримець. – К. : Центр навчальної літератури, 2008. – 270 с.
4. Рибак Т. Джерела збагачення національної наукової термінології / Т. Рибак, Н. Рибак, Я. Рибак // Проблеми української науково-технічної термінології : Тези 4-ої Міжнародної наукової конференції. – Львів : ДУ «Львівська політехніка», 1996. – С. 14-15.
5. Коваль М., Багдасар'ян Г. До проблеми формування української технічної термінології // Науково-технічне слово. – № 1 (93). – С. 75-78.



ЛОЖНЫЕ ДРУЗЬЯ ПЕРЕВОДЧИКА

Олифенко И.В.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707-63-60,
e-mail: InnaOlifenko@gmail.com*

Ложные друзья переводчика (ЛДП) – лексические единицы, близкие по написанию или звучанию в языке оригинала и в языке перевода, но отличающиеся своим смысловым наполнением и направляющие переводчика по ложному пути. Избитое, но в то же время достаточно условное понятие, так как семантический состав даже тех слов в двух языках, которые используются как переводческие эквиваленты, редко полностью совпадает. К категории ЛДП можно почти автоматически отнести подавляющую часть иностранных слов. В то же время интернационализмы, как правило, не являются ложными друзьями переводчика.

Принципиально следует различать „ложные друзья переводчика" в устной и письменной формах речи. Это требование обязательно в случае сопоставления языков с совершенно различными письменностями или, напротив, в случае языков с общей письменностью, но фонематически несходной лексикой. Для русского и английского языков с их сходными видами письма, находящимися в закономерных соответствиях, данное разграничение фактически может не проводиться, хотя степень, до которой сопоставляемые лексемы отождествляются двуязычными лицами, и здесь оказывается несколько различной в каждой из форм речи, а в определенных случаях отождествление разноязычных слов вообще имеет место только в одной из форм речи (например, русск. гейзер и англ. geuser ['gi:za] „газовая колонка для ванны" сходны только в написании).

Исторически „ложные друзья переводчика" являются результатом взаимовлияний языков, в ограниченном числе случаев могут возникать в результате случайных совпадений, а в родственных, особенно близкородственных, языках основываются на родственных словах, восходящих к общим прототипам в языке-основе. Их общее количество и роль каждого из возможных источников в их образовании оказываются различными для каждой конкретной пары языков, определяясь генетическими и историческими связями языков.

В английском и русском языках слова этого рода в подавляющем большинстве случаев представляют собой прямые или опосредствованные заимствования из общего третьего источника (часто это интернациональная или псевдоинтернациональная лексика или параллельные производные от таких заимствований). Значительно меньше представлены результаты собственно англо-русских языковых контактов: слова английского происхождения в русском языке (имидж, риэлтор, постер, провайдер, лидер, митинг, департамент) и русского происхождения в английском (intelligentsia - интеллегенция, kolkhoz-колхоз, sagene - ‘сажень’, rouble - ‘рубль’ izba - ‘изба’, telega - ‘телега’), хотя сре-



ди заимствованных слов этой группы иногда наблюдаются существенные расхождения со словами-образцами, затрудняющие носителям языка-источника понимание, казалось бы, „своего" слова в другом языке.

Количество, степень расхождения и распределение по частям речи „ложных друзей переводчика" различны для различных пар языков.

В английском и русском языках „ложные друзья переводчика", насчитывающие несколько тысяч слов, встречаются в пределах четырех частей речи: существительных, прилагательных, наречий и глаголов. В большом числе случаев в данной роли выступают не единичные слова, а все представители соответствующих словообразовательных гнезд. Естественно, для лиц, владеющих основами второго из языков, ложные отождествления имеют место лишь в сфере одинаковых частей речи: так, существительные ассоциируются с существительными и т. д.. С семантической точки зрения вводящими в заблуждение оказываются слова, принадлежащие к аналогичным или смежным семантическим сферам или, во всяком случае, могущие оказаться в сходных контекстах; явно случайно совпадающие лексемы, по сути не встречающиеся в одинаковых контекстах (типа англ. rock „скала" — русск. рок), не вызывают ложных ассоциаций. Расхождения в парах „ложных друзей переводчика" могут намечаться в понятийном содержании, реалиях, стилистических характеристиках и лексической сочетаемости; на практике все эти типы расхождений нередко переплетаются.

Приведем несколько примеров англо-русских ЛДП:

faggot	(оскорбительное) гомосексуалист (а не фагот - bassoon)
figure	чертёж, иллюстрация; цифра; число (не только фигура)
fiction	художественная литература, беллетристика, а не только фикция
football	(в американском английском) — американский футбол, а не футбол (soccer)
formula	состав; рецепт; закон; формулировка, а не только химическая формула
fruit	плод (в самом широком смысле), не только фрукт
gallant	храбрый (не только галантный; тж. gallant sail — брамсель)
most	превосх. ст. от much — более всего, большинство или наибольший (а не мост — bridge)
net	сеть
novel	роман (чаще, чем новелла)
velvet	бархат (а не вельвет — corduroy)
virtual	действительный, фактический, а не только виртуальный

Список литературы

1. Борисова Л.И. Ложные друзья переводчика (общенаучная лексика). Английский язык / Л.И. Борисова. – М. : Наука, 2002.
2. Краснов К.В. Англо-русский словарь «ложных друзей переводчика» / К.В. Краснов. – М. : ЭРА, 2004.
3. Дубичинский В.В. Лексические параллели / В.В. Дубичинский. – Харьков, 1993. – 155с.



ГЕНДЕРНИЙ АСПЕКТ СЕМАНТИКО-СТИЛІСТИЧНИХ ОСОБЛИВОСТЕЙ РЕКЛАМНОГО ІНТЕРНЕТ-ДИСКУРСУ

Золотько Д.В.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707-63-60,
e-mail: darthviolince@gmail.com*

Як інструмент економічного процесу, частина масової культури, різновид мистецтва та вид словесності реклама є об'єктом дослідження багатьох наук, зокрема й лінгвістики.

Реклама як інструмент лексико-семантичних досліджень. Реклама у всесвітній мережі дає можливість перевіряти ефективність мовних елементів та структур, які набули популярності в рекламному дискурсі в доелектронний період комунікації, і, як наслідок, здійснювати відбір найефективніших елементів, у тому числі й гендерно маркованих.

Абсолютна більшість досліджень у галузі рекламного дискурсу використовує у якості матеріалу англомовні зразки, у той час, як в рамках користування російсько- та україномовними інтернет-ресурсами доцільнішим є звернення уваги саме на українські та російські рекламні проекти.

Базовою складовою формування рекламного простору є статистична вибірка, що дає можливість оцінити, в першу чергу, потреби споживача та категоризувати ринок. В рамках таких досліджень одну з найбільш значущих ролей відіграє гендерна диференціація та її соціальні прояви.

Для досягнення базових цілей сфери маркетингу, у зв'язку з орієнтацією на абсолютну статистичну більшість, відбувається навмисна гіперболізація пов'язаних із жіночим та чоловічим архетипами стереотипічних узагальнених властивостей.

Фемінність у рекламному дискурсі. У випадку з націленими на жіночу аудиторію рекламними слоганами і текстами у лінгвістичному плані це частіше за все виражається у використанні зменшувально-пестливих суфіксів, «м'якої» лексики; особливо активно подібні прийоми маніпуляції використовуються у сфері товарів для дітей або вагітних жінок.

*«Одежда для дома и спорта для беременных будет незаменимой женщи-
нам, которые привыкли вести активный образ жизни. В этой одежде будущие
мамочки всегда смогут заниматься активным отдыхом, определенными физи-
ческими нагрузками и при этом чувствовать себя комфортно.»*

«Детские коляски. Комфортные прогулки для мамы и малыша.»

*«Цены уже снижены! Часы-браслет Pandora. О таких мечтает каждая
девушка»* - прийом узагальнення, спричинює так званий «ефект натовпу», прив'язаний у даному випадку саме до гендерної ідентичності.



Наступні зразки базуються на суспільно-прийнятому типі емоційної поведінки за жіночим гендерним типом. Використано також розповсюджений у рекламі з універсальною цільовою аудиторією прийом розмовної лексики.

«Девченки, налетайте! Офіціальна распродажа Nike со скидкой 55%»;

«Девчонки, пора пицать от радости! Часы-браслет PANDORA со скидкой!»

«Часы Royal London: Роскошь времени в драгоценном металле!» - специфіка жіночої орієнтованості реалізована шляхом використання нарочито художніх епітетів.

Лукізм і комерція. У рекламі, націленій на жіночу аудиторію, використовуваний виробниками товарів для догляду за зовнішністю, наявні використання лексико-семантичних прийомів, притаманних стереотипічно-жіночій лексиці. Підкреслюється пріоритетність зовнішніх характеристик у зв'язку з вимогами, пред'явлюваними до жіночої поведінки соціумом.

«Женщина в свои 48 лет выглядит, как девочка! "Вот моя хитрость:..."»

«Bust Salon Spa - крем нового поколения для УВЕЛИЧЕНИЯ груди!»

«Ты жінка, а не посудомийка! Hotpoint Ariston.» - яскравий приклад використання стереотипічної ролі жінки у рамках родинної моделі співіснування методом протиставлення.

У рекламі, орієнтованій на чоловічу аудиторію є дуже частим звертання до образу маскулінності і пов'язаних із нею характеристик, що супроводжуються відповідною лексикою, вживанням вільних, нетривіальних, іноді грубих висловлювань, що пояснюється тим, що мовленнєва поведінка чоловіків у одноставовому колективі тяжіє до некодифікованих форм.

«МУЖИКИ! Нас давно ОБМАНЬВАЮТ на заправках, этого можно ИЗБЕЖАТЬ!»

Однією з характерних рис є тяжіння до глобалізму, масштабності, очевидного перебільшення значимості речей і понять; часто використовуються лексичні структури на зразок «найкращий», «попереду», «світовий бренд», «найбільший у світі» і т.п.

«Книги скачать бесплатно. Книги меняющие мир к лучшему. Сенсационная книга тысячелетия!»

Одним з найуспішніших маніпулятивних прийомів у рекламі, орієнтованій на чоловічу аудиторію, є тяжіння до мілітаризму, його романтизація. Особливо розповсюдженим цей прийом є у сфері продажу відеоігор.

«Управляй громадным космическим крейсером! Радуй неприятелей лазерными вспышками и накрывай вражеские корабли огнем!»

«Даже взрослых МУЖИКОВ за уши не оттащишь от этой ИГРЫ!»

«Война Престолов - теперь и MMORPG! ИГРАЙ скорее! ГРАФА на 1000% удовольствия! ПРОЧУВСТВУЙ!» - використано специфічний сленг.

Глобальним аспектом рекламування у цій та інших сферах є сексуалізація та об'єктивізація образу жінки, що являє собою чи не найефективніший засіб привернення уваги споживача і запоруку комерційного успіху.

«Лига ангелов – соблазни богиню!»

«МУЖИКИ сходят с ума по этой ИГРЕ! Здесь ТАКОЕ творится!»



Ты напрочь забудешь про СОН и ЕДУ!»

«Начнешь играть в нее и забудешь даже о жене! Проверено!»

«Героини в этой культовой игре просто загляденье! Один сюжет чего стоит...»

«Няшние героини ждут своего героя.»

«Ты потеряешь голову от этой взрывной игры! Даже ладони начнут потеть!»

Яскравий приклад прояву беззаперечного андроцентризму, як важеля маніпуляції споживачем:

«Хлібний дар. Чоловікам: тим, хто обертає світ.»

Протиставлення акцентуації за гендерною диференціацією. Якщо порівняти приклади рекламних слоганів одного рекламодавця, направлених на продаж тієї ж самої продукції, очевидно є різниця в акцентуації у зв'язку зі стереотипними моделями розстановки пріоритетів: у випадку з жінками ставка робиться на сприйняття оточуючими, зовнішній вигляд, трендовість, у той час як «чоловіча» реклама акцентує увагу в першу чергу на якість і статусність. Це є дуже розповсюдженою особливістю, яка широко використовується багатьма рекламодавцями.

«Сумки известных брендов!»

«Женская обувь от 235 грн. Весенняя распродажа для модниц!»

Актуальність. Рекламний дискурс грає значну роль в визначенні потреб сучасного соціуму, відповідно до цього гендер є маркером вибору вербальних і невербальних методів регуляторного та маніпулятивного впливу на споживача. Незважаючи на велику кількість робіт, присвячених лінгвістичній галузі рекламного дискурсу, гендерний аспект залишається недостатньо дослідженим, що також підтверджує актуальність питання.

Літературні джерела:

1. Крутько Т. В. Мовні особливості рекламних текстів у гендерному аспекті (на матеріалі віртуальної реклами) / Т. В. Крутько // Лінгвістика XXI століття: нові дослідження і перспективи. – К. : Логос, – 2008. – С. 226-231.
2. Грошев И.В. Образ женщины в рекламе /Женщина. Гендер. Культура. – М. : МЦГИ, 1999. – С. 331-343.
3. Дубовская И.Н. Женский портрет на фоне рекламы // Вестник Московского университета. Сер.19, Лингвистика и межкультурная коммуникация. – 04/2002. – №2. – С.98-106.
4. Полубиченко Л.В. “Мужской” и “женский” язык рекламы // Вестник МГУ. Сер. 19. – Лингвистика и межкультурная коммуникация. – М., 2000. – № 2. – С. 48-97.



ЛЕКСИЧНІ ТРУДНОЩІ ПРИ ПЕРЕКЛАДІ (НА МАТЕРІАЛІ ПОВІСТІ Г.Ф. ЛАВКРАФТА «ПОШУК НЕЗВІДАНОГО КАДАТУ»)

Детинич К.Г.

Національний технічний університет
"Харківський політехнічний інститут",
м. Харків, вул. Пушкінська, 79/2, тел. 707-63-60,
e-mail: faramantthefirst1@gmail.com

Лексичні труднощі перекладу — тема, яку потрібно вивчати та мати на увазі при перекладі що художніх, що наукових текстів, адже для різних ситуацій важливо використовувати різні прийоми передачі авторської думки.

Актуальність даного аналізу. Аналізувати переклади текстів майстра фантастики минулого сторіччя Говарда Лавкрафта важливо й зараз хоча б тому, що у своїх творах він нерідко використовував слова-архаїзми, не був скупим на описи з одного боку, і міг виразити вельми цікаві думки за допомогою усього пари слів, тому влучний переклад його текстів — це не аби яке випробування навіть для досвідченого перекладача.

З чим пов'язані труднощі перекладу? Лексичні труднощі перекладу художніх текстів головним чином пов'язані з передачею безеквівалентної лексики, назв, багатозначних слів, аббревіатур, неологізмів, термінів, образної фразеології. Окремо треба виділити багатозначні англійські слова, які складають значну частину англійської лексики. Останню групу складають слова, які не мають готових відповідників у цільовій мові.

У чому ж різниця? Треба зазначити важливість структурних і лексико-семантичних розходжень між англійською та українською мовами, які вимагають під час перекладу перебудови синтаксичної структури речення або лексичних змін, і називаються лексико-граматичними трансформаціями.

Види трансформації. Які вони і для чого існують. Ці трансформації застосовують тоді, коли словникові відповідники того чи іншого слова мови оригіналу не можуть бути використані у перекладі з причин невідповідності з точки зору значення і контексту.

Цими трансформаціями є: транскодування, калькування, антонімічний переклад, генералізація, конкретизація, модуляція, ампліфікація.

Розглянемо із якими лексичними труднощами зіткнувся перекладач даної повісті та розглянемо приклади використаних прийомів трансформації.

Транскодування: Randolph Carter — Рендольф Картер, Nyarlathotep — Ньярлатотеп.

Загалом, у даному тексті транскодування використовується для назв міст або географічних об'єктів та імен персонажів. В тому ж випадку, коли для слів оригіналу можна віднайти більш влучні аналоги, перекладач використовує саме їх (напр. ancient — не античний, а древній).

Калькування: the tall black towers of Dylath-Leen, which is built mostly of *basalt* — чорні високі *базальтові* башти Дайлат-Ліну; zebra — зебра.



Калькування у данному перекладі використовується у тому випадку, коли у власне українській мові потрібний аналог відсутній (тобто, для певних видів тварин, будівних матеріалів, певних видів споруд, і такого іншого).

Антонімічний переклад: one of the *uncomfortable* merchants spoke to him — один з тих *химерних* купців звернувся до нього; dream-quest — сомнамбулічний пошук.

Антонімічний переклад дозволяє вмістити думку висловлювання у більш звичні рамки, використовуючи слова або словосполучення для поширення або звужування значення. Іноді для влучного перекладу необхідно змінювати не лише одне слово, але й усю будову речення.

Генералізація:

1. no man had ever suspected *in what part of space it may lie*; — ніхто й гадки не мав, *у якій частині всесвіту він знаходиться*.

2. Kuran was *the one soul* — Куранес *був єдиним, хто...*

Генералізацію перекладач застосовував із метою донести думку автора, використовуючи слова, більш ширші за значенням, ніж слова, які використовувались у мові оригіналу.

Конкретизація: and having no power *or habitation* — не маючи ані влади, ані *власного притулку*; but these *lawless spirits* were soon restrained by their elders — але *нахаб* швидко поставили на місце старші.

Широкі значення слів у мові-носії при перекладі краще передавати більш конкретним значенням із метою зберегти авторську думку.

Модуляція:

1. Seven Cryptical Books of Hsan — Семикнижжя Хсана.

2. It was hard to get this information from the traders and sailors — Важко було дізнатись бодай щось від купців чи моряків.

Дослівний переклад іноді знищує авторську думку, і у поданих ситуаціях перекладач спирався на загальний контекст, а не на словники.

Компенсація:

1. You came not as one curious, but as one seeking his due — ти прийшов не через пусті цікавинки, але через пошук власної великої мети

2. The earth has no longer any gods that are gods — не лишилось на землі богів, що гідні зватись богами.

Компенсація використовується для того, щоб передати потрібне емоційне забарвлення, яке дослівний переклад може втратити.

Амліфікація: Mt. Natheg-Kla — гора Хатег-Кла.

Список літератури

1. Мацько Л.І. Стилїстика української мови: Підручник / Л.І. Мацько, О.М. Сидоренко, О.М. Мацько; За ред. Л.І. Мацько. — К.: Вища школа, 2003.- 462 с.
2. Рецкер Я.И. Теория перевода и переводческая практика. — М., «Международные отношения», 1974.- 216с.
3. H.P. Lovecraft. Dream-quest of the Unknown Kadath. — New York: Ballantine Books, 1970, 1–141.

Матеріали

IV Всеукраїнської науково-практичної конференції

**"ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ТА
ПРИКЛАДНА ЛІНГВІСТИКА"**

(Українською, російською та англійською мовами)

Відповідальний за випуск *Н.В. Шаронова*

Технічна редакція та комп'ютерна верстка: *С.В. Петрасова*

Формат 60x90/16. Ум. друк. аркушів. 7,25
Наклад 50 прим. Зам. № 066876

Надруковано у ТОВ «ПЛАНЕТА-ПРІНТ»
61002, м. Харків, вул. Фрунзе, 16
ЄДРПОУ 3125131 від 19.12.2000 р