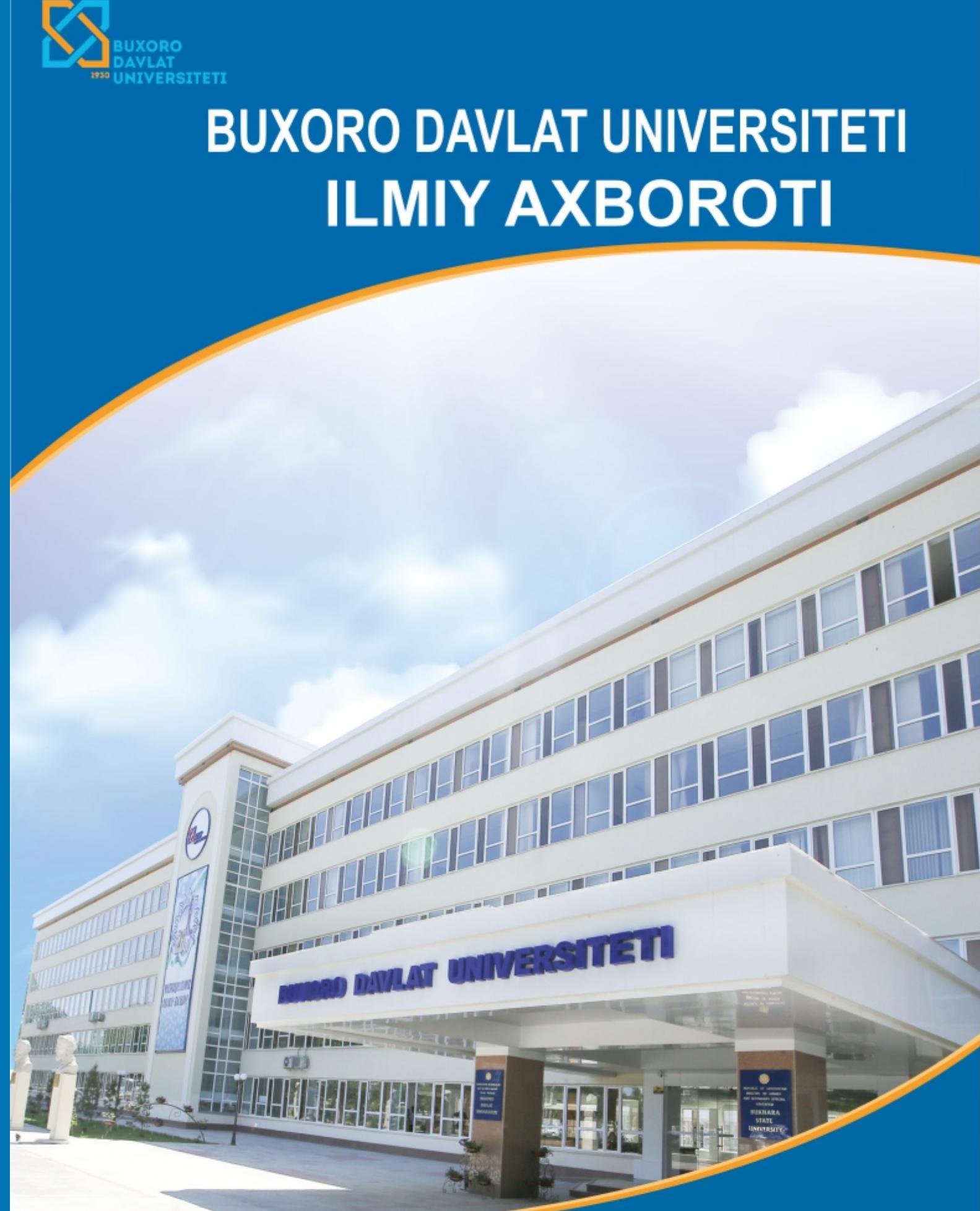


BUXORO DAVLAT UNIVERSITETI ILMIY AXBOROTI

Научный вестник Бухарского государственного университета
Scientific reports of Bukhara State University

4/2024



4/2024

<https://buxdu.uz>



MUNDARIJA * СОДЕРЖАНИЕ *** CONTENTS**

TILSHUNOSLIK * LINGUISTICS *** ЯЗЫКОЗНАНИЕ**

Safarov F.S.	O'zbek adabiy tilida ko'plik ma'nosining ifodalanishi va tejamkorlik	4
Akhmedova R.A.	Antropocentric aspect of phraseological units with the component "red"	12
Alieva N.X., Mukhtoraliyeva F.A.	The role of translation in the enrichment of languages: analysis of some neologisms presented in «Steve Jobs» by Walter Isaacson	17
Barnoyeva Z.S.	O'zbek leksikografiyasida boshlang'ich sinf o'quv lug'atlarini tuzishning ahamiyati	22
Barotova N.Sh.	O'zbek kinodiskursida "milliylik" obrazini yaratishning tamoyillari	27
Ganieva O.Kh., Khakimova D.M.	Concept of term definition and its characteristics	32
Kurbanova G.S., Pulatova U.B.	Frazeologik konsepsiylar talqini	37
Jumayev E.B., Nurullayeva G.X.	O'zbek va nemis tilida shaxs ifodalovchi vositalar	45
Raxmonova Y.X.	Matn tili va tarjima muammolari	50
Ro'ziyev Y.B., Halimova Sh.Sh.	O'zbek va nemis tillarida o'tgan zamonning voqelanishi	55
Turayeva D.M.	Muallif punktuatsiyasining yuzaga kelishi va punktuatsion tizimdag'i yangi an'analar	61
Xodjayeva D.I.	Lingvodidaktik terminologiya tadqiqot obyekti sifatida	65
Yaqubova M.M.	Tarjima jarayonida asl matnning semantik va pragmatik jihatlarini saqlash masalalari	70
Ахмеджанова С.Дж.	Сбор и анализ корпусных данных для узбекского и русского языков	76
Жўраева И.А.	Инглиз ва ўзбек антропонимик лугатларида атоқли отларнинг берилиш усуллари	81
Шермуҳамедова Н.А.	Двуязычный и многоязычный устный корпус	87
Qobilova N.S.	Til tejamkorligi nazariyasining tadqiq qilish tamoyillari	92
Qodirova H.X.	Badiiy diskurs tarjimasida kommunikativ-funksional yondashuv	98
Yuldasheva F.E.	Expression of the modesty maxim in communication	102
Рабиева М.Ф.	Эвфемизмларнинг дискурсив турлари ва уларнинг таржима стратегияларини белгилашдаги роли	106
Ataeva G.B.	The power of idioms in business English: communicating with impact	117
Karimova D.H.	Tarjimada variantlilik hodisalarining asosiy belgilari xususida (Aka-uka Grimm ertaklarining o'zbekcha tarjimalari misolida)	122

СБОР И АНАЛИЗ КОРПУСНЫХ ДАННЫХ ДЛЯ УЗБЕКСКОГО И РУССКОГО ЯЗЫКОВ

Ахмеджанова Ситора Джураходжаевна,
преподаватель кафедры русского литературоведения
Бухарского государственного университета
iamhappyandglad@gmail.com

Аннотация. Данное исследование углубляется в практику сбора и анализа корпусных данных для узбекского и русского языков: на материале гипонимических отношений. Делается важный шаг для создания интерактивной онлайн-платформы для визуализации гипонимических отношений в русском и узбекском языках. Выявляется цель сбора корпусных данных, определяются методы, критерии и аспекты сбора данных, описываются важные источники для обогащения корпуса данных, представляются программные инструменты, необходимые для сбора данных. Также разрабатываются стратегии решений некоторых специфических проблем, которые возникли в ходе нашего исследования.

Ключевые слова: корпус данных, сбор и обработка данных, гипонимические отношения, узбекский и русский языки, веб-скрейперы, текстовые анализаторы, базы данных, менеджеры корпуса.

O'ZBEK VA RUS TILLARI UCHUN KORPUS MA'LUMOTLARINI TO'PLASH VA TAHLILI

Annotation. Ushbu tadqiqot o'zbek va rus tillari uchun korpus ma'lumotlarini yig'ish va tahlil qilish amaliyotini o'rGANADI: giponimik munosabatlarga asoslangan. Rus va o'zbek tillarida giponimik munosabatlarni vizuallashтирish uchun interaktiv onlayn platforma yaratish bo'yicha muhim qadam qo'yilmoqda. Korpus ma'lumotlarini to'plashdan maqsad aniqlanadi, ma'lumotlarni yig'ish usullari, mezonlari va jihatlari aniqlanadi, ma'lumotlar korpusini boyitishning muhim manbalari tavsiflanadi, ma'lumotlarni yig'ish uchun zarur bo'lgan dasturiy vositalar taqdim etiladi. Tadqiqotlarimiz davomida yuzaga kelgan ayrim muayyan muammolarni hal qilish uchun strategiyalar ham ishlab chiqilmoqda.

Kalit so'zlar: ma'lumotlar korpusi, ma'lumotlarni yig'ish va qayta ishslash, giponimik munosabatlar, o'zbek va rus tillari, veb-skreperlar, matn analizatorlari, ma'lumotlar bazalari, korpus menejerlari.

COLLECTION AND ANALYSIS OF CORPS DATA FOR UZBEK AND RUSSIAN LANGUAGES

Abstract. This study delves into the practice of collecting and analyzing corpus data for the Uzbek and Russian languages: based on hyponymic relations. An important step is being taken to create an interactive online platform for visualizing hyponymic relations in the Russian and Uzbek languages. The purpose of collecting corpus data is identified, methods, criteria and aspects of data collection are defined, important sources for enriching the data corpus are described, and software tools necessary for data collection are presented. Strategies are also being developed to address some of the specific problems that arose during our research.

Keywords: data corpus, data collection and processing, hyponymic relations, Uzbek and Russian languages, web scrapers, text analyzers, databases, corpus managers.

Введение. В эпоху цифровизации и глобализации языковых процессов, актуальность сбора и обработки корпусных данных в лингвистических исследованиях невозможно переоценить. Современные лингвистические исследования опираются на массивные корпусы текстовых данных, предоставляя беспрецедентные возможности для глубокого понимания языковых структур и динамики. Особое внимание в данной работе уделяется корпусам русского и узбекского языков, каждый из которых обладает уникальными особенностями и представляет интерес для компаративного анализа.

Целью сбора корпусных данных является создание обширной и репрезентативной базы текстовых материалов, которая позволяет проводить качественные и количественные исследования языковых явлений. Важно подчеркнуть, что сбор данных не ограничивается простым накоплением текстов: он включает в себя тщательный отбор материалов, их аннотацию и стандартизацию, что обеспечивает их последующую обработку и анализ.

В контексте узбекского и русского языков, где исторические, культурные и социолингвистические факторы играют ключевую роль, методы сбора данных должны учитывать специфику каждого языка [1]. Например, для узбекского языка, который испытал значительные изменения в связи с переходом на латиницу, важно собирать данные, как в старом, так и в новом орографических стандартах. Русский язык, с его богатой литературной традицией и обширным диалектным разнообразием, требует внимательного подхода к выбору репрезентативных текстов, охватывающих различные жанры и стили [2].

Сбор корпусных данных включает в себя не только литературные тексты, но и устные речевые данные, интернет-коммуникации, техническую и научную литературу, что позволяет обеспечить максимально полное покрытие языкового материала. Такой подход способствует более глубокому пониманию языковых процессов, их изменений и развития [3].

Следовательно, начальный этап сбора и обработки корпусных данных не только заложит основу для всего исследования, но и определит глубину и качество возможного анализа. Сбор корпусных данных является критически важным шагом, требующим тщательного планирования и реализации, учитывая особенности исследуемых языков.

В основе сбора корпусных данных лежит принципиально важный процесс определения и выбора источников данных, который требует учёта множества параметров и особенностей, присущих узбекскому и русскому языкам. Корпус данных в лингвистике представляет собой собрание текстовых материалов, организованных и структурированных таким образом, чтобы обеспечить максимальную репрезентативность и охват различных языковых явлений и стилей.

При определении источников для корпуса данных, особое внимание уделяется многообразию языковых материалов. Это включает в себя не только литературные произведения, но и устные народные тексты, научные публикации, СМИ, интернет-ресурсы, включая блоги, форумы и социальные сети. Особенно это актуально для узбекского языка, который обладает богатым фольклорным наследием, и русского языка, имеющего обширное литературное и научное пространство [4].

Критерии отбора данных для корпуса включают в себя несколько ключевых аспектов. Во-первых, это степень репрезентативности источника, обеспечивающая широкое покрытие языковых явлений. Во-вторых, качество и достоверность данных, что особенно важно при работе с устными источниками и интернет-ресурсами. Также необходимо учитывать хронологическое и географическое разнообразие источников, что позволяет охватить исторические изменения в языке и диалектные особенности.

В контексте узбекского языка важным аспектом является выбор текстов, написанных как на кириллице, так и на латинском алфавите, что отражает переходный период в развитии языка [5]. Для русского языка акцент делается на включении текстов из различных регионов России и стран бывшего Советского Союза, чтобы обеспечить диалектное разнообразие.

Таким образом, этап определения и выбора источников данных требует комплексного подхода, учитывая языковую специфику, стилевое и жанровое многообразие, а также динамические процессы в развитии языка. Эффективное выполнение этого этапа лежит в основе формирования качественного и насыщенного корпуса данных, который послужит надёжной основой для последующего анализа и визуализации гипонимических отношений в узбекском и русском языках.

В процессе сбора корпусных данных для анализа гипонимических отношений в русском языке, основное внимание было удалено выбору источников, способных предоставить обширный и многоаспектный материал [6]. Среди ключевых ресурсов, использованных в исследовании, стоит выделить несколько значимых и авторитетных источников.

Первым и весьма важным источником стал онлайн-словарь прилагательных, доступный по адресу <http://nskhuman.ru/unislov/morfolog.php>. Этот ресурс предоставляет широкий спектр морфологически разнообразных прилагательных, что позволяет изучить их гипонимические и гиперонимические отношения, а также их вариативность в различных контекстах.

Далее, значительный вклад в корпус данных внесли труды в области специализированной лексики, в частности, «Словарь биологических терминов» (январь 2013 года, авторы Белякова Г.А. и

LINGUISTICS

Владимир Зданович, издательство Московского университета) [7]. Этот источник обогатил корпус терминологией, что, несомненно, важно для анализа научного дискурса и специализированной лексики в контексте гипонимических отношений.

Также важным элементом корпуса стали классические словари русского языка, включая работу В.И. Даля «Толковый словарь русского языка», издание Эксмо 2005 года, и «Толковый словарь русского языка» Ожегова С.И. и Шведовой Н.Ю., 1999 года. Эти словари обеспечили необходимую глубину исторического и культурного контекста, предоставив обширный материал для анализа эволюции и разнообразия языковых единиц [8].

Кроме того, использование «Толкового словаря русского языка» Ушаковой Д.Н. (Москва, 1935 год) позволило дополнить корпус уникальными историческими данными, что является ключевым для понимания динамических изменений в языке.

Не менее значимым для нашего исследования оказался Национальный корпус русского языка (НКРЯ), доступный онлайн по адресу <https://ruscorpora.ru/>. Этот ресурс представляет собой ценное собрание текстов современного русского языка, включающее в себя, как письменные, так и устные источники, обеспечивая тем самым обширное покрытие различных жанров и стилей.

Эти источники, каждый из которых отличается своим уникальным вкладом в корпус данных, были тщательно отобраны с учётом их репрезентативности, достоверности и способности обеспечить всестороннее покрытие языковых явлений [9]. Их совокупное использование позволило создать многофункциональный и детализированный корпус, который лежит в основе нашего исследования гипонимических отношений в русском языке.

В рамках исследования, посвящённого визуализации гипонимических отношений в узбекском языке, был проведён тщательный отбор источников для формирования корпусных данных, при этом особое внимание было удалено комплексности и многообразию лексических материалов [10].

Одним из первых важных источников, использованных в исследовании, стал Национальный корпус узбекского языка (НКУЯ), доступный по адресу <https://uzbekcorpus.uz/newIndex>. Этот ресурс представляет собой обширную базу текстовых данных, включая литературные произведения, публицистические и научные тексты, что обеспечивает глубокое понимание лексического разнообразия и структурных особенностей узбекского языка.

Далее, значительное место в корпусе занимает Образовательный корпус узбекского языка (ОКУЯ), доступный на <https://uzschoolcorpora.uz/>. Этот ресурс включает в себя образовательные и учебные тексты, что позволяет исследователям анализировать языковые структуры, используемые в образовательной сфере, и их влияние на формирование гипонимических отношений.

Важным источником для анализа стал словарь «O‘zbek tilining izohli lug‘ati» (новая редакция) под редакцией А. Мадвалиева, Н. Махкамова и других, изданный издательством «O‘zbekiston nashriyoti» в 2021 году. Этот словарь обеспечивает обширное описание узбекской лексики, включая современные и исторические термины, что критически важно для глубокого анализа языковых отношений.

Дополнительно, корпус обогащён словарём «Botanikadan qisqacha izohli lug‘at» авторства А.И. Косимова и М.М. Набиева, изданным в Ташкенте издательством «O‘qituvchi» в 1990 году. Этот ботанический словарь внес вклад в корпус уникальной специализированной лексикой, расширяющей понимание гипонимических структур в узкоспециализированных областях знаний.

Наконец, «Ўзбек тилининг изоҳли лӯғати» под редакцией Э. Бегматова, А. Мадвалиева и других, изданный издательством «Ўзбекистон миллӣ энциклопедияси» Давлат илмий нашриёти в 2006 году, представляет собой фундаментальный словарь узбекского языка. Он включает в себя богатый массив данных о лексических единицах и их употреблении, что, несомненно, способствует детальному исследованию гипонимических отношений.

Комбинирование данных из этих источников позволило создать многоуровневый и глубоко аналитический корпус, который служит основой для изучения гипонимических отношений в узбекском языке. Такой подход гарантирует комплексное освещение языковых явлений, отражая их в контексте современного узбекского языкового пространства и исторических традиций.

В рамках нашего исследования, нацеленного на создание интерактивной онлайн-платформы для визуализации гипонимических отношений в русском и узбекском языках, ключевое значение имеет использование программных инструментов для автоматизированного сбора данных. Эти инструменты обеспечивают эффективность, скорость и объём сбора данных, что критически важно для обеспечения достаточной репрезентативности и объективности корпуса.

Для сбора данных были использованы следующие программные инструменты:

Веб-скрейперы: Инструменты веб-скрейпинга, такие как Beautiful Soup и Scrapy, предназначены для извлечения данных из веб-страниц. Они автоматизируют процесс сбора текстов из различных онлайн-источников, включая новостные порталы, форумы и блоги. Эти инструменты особенно цепны для сбора актуальной и динамичной информации, что позволяет отслеживать современные языковые тенденции. Применение Beautiful Soup оказывается особенно эффективным при работе с HTML и XML документами, тогда как Scrapy предпочтительнее для более крупномасштабных проектов сбора данных.

Текстовые анализаторы: Программы анализа текста, такие как NLTK (Natural Language Toolkit) и Stanford NLP, предоставляют функциональность для обработки и анализа естественного языка. Эти мощные инструменты обработки естественного языка обеспечивают глубокий лингвистический анализ, включая токенизацию, морфологический разбор, определение частей речи и синтаксический анализ. Это позволяет извлекать из текстов ключевую информацию, необходимую для последующего анализа и визуализации.

Базы данных и менеджеры корпуса: Программы управления корпусами, такие как AntConc и Sketch Engine, позволяют организовать собранные тексты, проводить их анализ и визуализацию. Они обеспечивают удобный доступ к корпусным данным, их классификацию и фильтрацию, что значительно упрощает процесс их изучения.

Процесс использования этих инструментов начинается с определения параметров скрейпинга в Beautiful Soup и Scrapy, что включает в себя выбор целевых веб-страниц и спецификацию данных для извлечения. После сбора данных NLTK и Stanford NLP применяются для их первичной обработки и анализа, включая разбиение текста на предложения, определение ключевых слов и фраз, а также анализ грамматической структуры.

В процессе управления данными важную роль играет организация собранных текстов, их классификация и хранение. AntConc и Sketch Engine используются для создания удобной структуры корпуса, позволяя исследователям эффективно работать с большими объемами текстов. Эти инструменты также предоставляют функциональность для поиска по корпусу, что значительно упрощает процесс извлечения необходимой информации и её анализа.

Преимущества использования этих инструментов многочисленны. Веб-скрейперы значительно ускоряют процесс сбора данных, позволяя охватить широкий спектр источников. Текстовые анализаторы автоматизируют сложные процессы лингвистического анализа, обеспечивая точность и глубину изучения языковых структур. Базы данных и менеджеры корпуса способствуют эффективному управлению и анализу больших объемов данных.

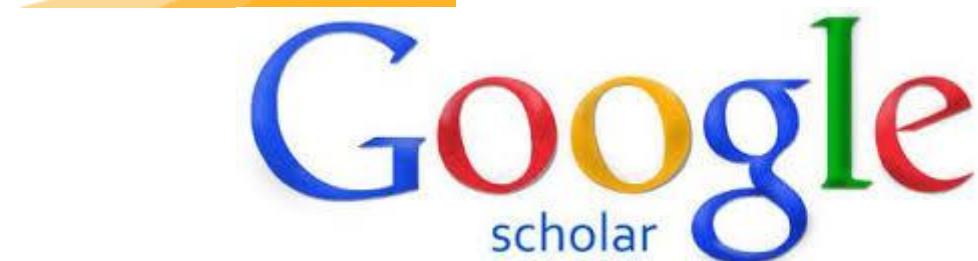
Однако существуют и недостатки. Веб-скрейперы могут столкнуться с ограничениями доступа на некоторых веб-сайтах или собирать нерелевантные данные. Текстовые анализаторы иногда не учитывают контекстуальные и культурные особенности языка, что может привести к ошибкам в интерпретации данных. Управление большими корпусами данных требует значительных вычислительных ресурсов и специализированных знаний.

Таким образом, выбор и использование программных инструментов для сбора корпусных данных является комплексным процессом, который требует баланса между автоматизацией и точностью, а также учёта специфических потребностей исследования. Эффективное применение этих инструментов обеспечивает сбор обширного и разнообразного корпуса данных, что является фундаментом для последующего анализа гипонимических отношений в русском и узбекском языках.

Заключение. В заключение данной статьи, посвящённой сбору и обработке корпусных данных на примере узбекского и русского языков, следует подчеркнуть, что процесс сбора и анализа данных представляет собой многоаспектный и многоуровневый процесс, включающий в себя как квантиitative, так и квалификативный анализ. С одной стороны, это обеспечивает глубокое понимание структурных и функциональных аспектов гипонимии в узбекском и русском языках, обогащённое детальным изучением корпусных данных, с другой стороны, разрабатывает методологические подходы к их обработке и визуализации. Это требует тщательного анализа лингвистических характеристик каждого языка и адаптации существующих алгоритмов обработки данных к особенностям обоих языков. Совместное использование этих подходов позволяет не только точно определить гипонимические отношения, но и выявить уникальные особенности каждого языка, что, в свою очередь, ведёт к созданию более эффективных инструментов для их анализа и представления в рамках интерактивной платформы. Таким образом, данная статья не просто описывает процесс, но и устанавливает основу для следующих этапов исследования, где основное внимание будет уделено разработке и тестированию самой платформы.

ЛИТЕРАТУРА:

1. Бокарева М. «Игра в текст» как способ построения постреалистического художественного дискурса //*Gospodarka i Innowacje.* – 2022. – Т. 23. – С. 407-413.
2. Ахмеджанова С. Специфика лакунарности в русской лингвистике //Евразийский журнал социальных наук, философии и культуры. – 2023. – Т. 3. – №. 3. – С. 84-88.
3. Аvezov С. О корпусной лингвистике, трудностях перевода и принципах организации параллельных корпусов текстов //«Узбекские национальные образовательные здания теоретическое и практическое создание вопросы» Международная научно-практическая конференция. – 2022. – Т. 1. – №. 1.
4. Avezov S. S. Machine translation to align parallel texts //International Scientific and Current Research Conferences. – 2022. – С. 64-66.
5. Khamidovna N. L. Expression of the Harmony of Language and Culture in World and Uzbek Lexicography //resmilitaris. – 2023. – Т. 13. – №. 1. – С. 233-244.
6. Нигматова Л. Параллельные корпусы, как эффективный метод обучения иностранного языка //«узбекские национальные образовательные здания теоретическое и практическое создание вопросы» Международная научно-практическая конференция. – 2022. – Т. 1. – №. 1.
7. Khamidovna N. G., Bakhodirovna K. S. System Relations in the Vocabulary of the Russian Language //Global Scientific Review. – 2022. – Т. 3. – С. 44-48.
8. Ахмеджанова С. Д. Русский и узбекский языки в сопоставительном аспекте //Ответственный редактор. – 2023. – С. 37.
9. Temirova D. H. et al. Genre specifics of children's game folklore //International Journal on Integrated Education. – Т. 3. – №. 3. – С. 57-62.
10. Бокарева М. Постреализм как новая парадигма художественности //Models and methods in modern science. – 2023. – Т. 2. – №. 12. – С. 160-167.



**"SCIENTIFIC REPORTS
OF BUKHARA STATE
UNIVERSITY"**

The journal was composed
in the Editorial and
Publishing Department of
Bukhara State University.

Editorial address:
Bukhara, 200117
Bukhara State University, main
building, 2nd floor, room 219.
Editorial and Publishing
Department.
<https://buxdu.uz/32-buxoro-davlat-universiteti-ilmiy-axboroti/131/131-buxoro-davlat-universiteti-ilmiy-axboroti/>
e-mail:
nashriyot_buxdu@buxdu.uz

Printing was permitted
29.04.2024 y. Paper format
60x84,1/8. Printed in express
printing method. Conditional
printing plate – 35,30.
Circulation 70. Order № 30.
Price is negotiable.
Published in the printing house
"BUKHARAHAMD PRINT" LLC
Address: Bukhara,
K.Murtazayev street, 344