**МИНОБРНАУКИ РОССИИ**

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«Волгоградский государственный социально-педагогический университет»**

**(ФГБОУ ВО «ВГСПУ»)**

**Факультет математики, информатики, физики**

**Кафедра физики, методики преподавания физики и математики, ИКТ**

**РЕФЕРАТ**

По дисциплине

«Информационные технологии в лингвистике»

**ТЕМА**

**Информационные базы данных в филологии**

**Работу выполнила**

Студентка группы ФЛ-РЛБ-11

Москаленко Екатерина

**Проверил (а)**

Сабанова Людмила Витальевна

Доцент кафедры

физики, методики преподавания физики и математики, ИКТ

Волгоград

2019

**СОДЕРЖАНИЕ**

стр.

Введение………………………………………………………………….2 - 3

1.Понятие о базах данных ………………4 - 5

2.Типы баз данных………………………5 - 6

3. Информационные базы данных в филологии ……………………...6

Заключение…………………………………………………………7

Список использованной литературы ..............................................8

Введение

Становление современного информационного общества приводит к коренным изменениям во всех сферах жизни и деятельности человека. В наши дни владение ИТ становится в один ряд с такими качествами, как умение читать и писать. Сегодня специалист с высшим образованием должен "свободно ориентироваться в мировом информационном пространстве, иметь необходимые знания и навыки поиска, обработки и хранения информации с использованием современных ИТ, компьютерных систем и сетей. ИТ – "это совокупность методов и программно-технических средств, объединенных в технологическую цепочку, обеспечивающею сбор, обработку, хранение, распределение и отображение информации с целью снижения трудоемкости процессов использования информационных ресурсов" .

Сегодня использование ИТ значительно увеличивает эффективность функционирования любой сферы человеческой деятельности. Лингвистика не является исключением. Какими бы далекими областями, на первый взгляд, ни казались ИТ и лингвистика, неоспоримая связь между ними подтверждается хотя бы фактом существования компьютерной лингвистики (КЛ ) – "направления в прикладной лингвистике, ориентированного на использование компьютерных инструментов – программ, компьютерных технологий организации и обработки данных – для моделирования функционирования языка в тех или иных условиях, ситуациях, проблемных сферах и т.д., а также всей сферы применения компьютерных моделей языка в лингвистике и смежных дисциплинах. Это и является актуальностью данной работы.

Ее целью является анализ возможностей и эффективности применения ИТ в лингвистике в целом и в исследовании автора, связанном с изучением искусствоведческих текстов на английском и русском языках, в частности. Для достижения поставленной цели необходимо решить следующие задачи :

* Сделать обзор литературы, посвященной применению ИТ в лингвистике.
* Охарактеризовать основные лингвистические программы и ресурсы, представленные в сети Интернет.
* Рассмотреть те программы и ресурсы, которые целесообразно использовать при исследовании искусствоведческих текстов.

2

В связи с тем, что в ходе исследования необходимо изучить разнообразную литературу, делая при этом определенные выводы из прочитанного, освоить специфику использования различных программ обработки текстов, показывая их эффективность на конкретных примерах и сравнивая между собой, в работе применяются методы анализа, синтеза, дедукции, индукции, сравнения.

3

**1.Понятие о базах данных**

Одной из важнейших областей применения компьютеров является переработка и хранение больших объемов информации в различных сферах деятельности человека: в экономике, банковском деле, торговле, транспорте, медицине, науке и т.д.

Существующие современные информационные системы характеризуются огромными объемами хранимых и обрабатываемых данных, сложной организацией, необходимостью удовлетворять разнообразные требования многочисленных пользователей.

Информационная система — это система, которая реализует автоматизированный сбор, обработку и манипулирование данными и включает технические средства обработки данных, программное обеспечение и обслуживающий персонал.

Цель любой информационной системы — обработка данных об объектах реального мира. Основой информационной системы является база данных. В широком смысле слова база данных — это совокупность сведений о конкретных объектах реального мира в какой-либо предметной области. Под предметной областью принято понимать часть реального мира, подлежащего изучению для организации управления его объектами и, в конечном счете, автоматизации, например, предприятие, вуз и т. д.

Создавая базу данных, пользователь стремится упорядочить информацию по различным признакам и быстро производить выборку с произвольным сочетанием признаков. При этом очень важно выбрать правильную модель данных. Модель данных — это формализованное представление основных категорий восприятия реального мира, представленных его объектами, связями, свойствами, а также их взаимодействиями.

База данных — это информационная модель, позволяющая упорядоченно хранить данные о группе объектов, обладающих одинаковым набором свойств.

Информация в базах данных хранится в упорядоченном виде. Так, в записной книжке все записи упорядочены по алфавиту, а в библиотечном каталоге либо по алфавиту (алфавитный каталог), либо в соответствии с областью знания (предметный каталог).

4

Система программ, позволяющая создавать БД, обновлять хранимую в ней информацию, обеспечивающая удобный доступ к ней с целью просмотра и поиска, называется системой управления базами данных (СУБД).

**2.Типы баз данных**

Группу связанных между собой элементов данных называют обычно записью. Известны три основных типа организации данных и связей между ними: иерархический (в виде дерева), сетевой и реляционный.

Иерархическая БД

В иерархической БД существует упорядоченность элементов в записи, один элемент считается главным, остальные — подчиненными. Данные в записи упорядочены в определенную последовательность, как ступеньки лестницы, и поиск данных может осуществляться лишь последовательным "спуском" со ступеньки на ступеньку. Поиск какого-либо элемента данных в такой системе может оказаться довольно трудоемким из-за необходимости последовательно проходить несколько предшествующих иерархических уровней. Иерархическую БД образует каталог файлов, хранимых на диске; дерево каталогов, доступное для просмотра в Norton Соmmander, — наглядная демонстрация структуры такой БД и поиска в ней нужного элемента (при работе в операционной системе МS-DOS). Такой же базой данных является родовое генеалогическое дерево.

Сетевая БД

Эта база данных отличается большей гибкостью, так как в ней существует возможность устанавливать дополнительно к вертикальным иерархическим связям горизонтальные связи. Это облегчает процесс поиска требуемых элементов данных, так как уже не требует обязательного прохождения всех предшествующих ступеней.

Реляционная БД

Наиболее распространенным способом организации данных является третий, к которому можно свести как иерархический, так и сетевой — реляционный (англ. relation — отношение, связь). В реляционной БД под записью понимается строка прямоугольной таблицы. Элементы записи образуют столбцы этой таблицы (поля). Все элементы в столбце имеют одинаковый тип (числовой, символьный), а каждый столбец

5

— неповторяющееся имя. Одинаковые строки в таблице отсутствуют.

Преимущество таких БД—наглядность и понятность организации данных, скорость поиска нужной информации. Примером реляционной БД служит таблица на странице классного журнала, в которой записью является строка с данными о конкретном ученике, а имена полей (столбцов) указывают, какие данные о каждом ученике должны быть записаны в ячейках таблицы.

**3.Информационные базы данных в филологии**

Информационные проекты, обеспечивающие публикацию электронных версий трудов ученых, создание инновационного продукта, научных информационно - телекоммуникационных систем и сетей, унифицированной системы научных знаний и технологий, создание некоторых элементов искусственного интеллекта при построении экспертных систем и баз знаний, информационных сетей коллективного пользования с уникальными научными ресурсами. Новейшее программное обеспечение, которое позволяет облегчить процесс работы в гуманитарной области знания - филологии (Apache OpenNLP – OpenNLP - это организационный центр "open source" проектов, связанных с машинной обработкой естественного языка под эгидой Apache. OpenNLP предлагает обширный Java - инструментарий обработки текстов на основе методов машинного обучения. Он включает средства токенизации, выделения предложений, разметки частей речи, выделения имен собственных, разбора текста и разрешения перекрестных ссылок. Имеется документация на английском языке. Для скачивания доступен исходный код и бинарные компоненты (для запуска требуется установка Java VM), «Ищейка» (www.componentsoftware.com) Лингвоанализатор (http://www.rusf.ru/books/analysis/). Утилиты и программные комплексы TextAnalyst (www.analyst.ru), «Рабочее место лингвиста» (www.aot.ru) - построение семантической цепи по степени частотности словоупотребления, грамматический анализ текстов, машинный морфологический разбор слов (Morphology), синтаксический (VisualSynan) и семантический анализы (RossDev ).

**Заключение**

Использование баз данных и информационных систем – это составная часть функционирования различных преуспевающих организаций и деятельности современного человека. В связи с этим большую актуальность приобретает

6

освоение принципа построения и эффективного применения 6соответствующих технологий и программных продуктов.

В настоящее время базы данных почти во все сферы человеческой деятельности, в том числе и в филологии. Так как с помощью электронных баз данных работа персонала школы и учителей очень полезна и быстротечна. В своей работе я попробовала рассмотреть и доказать это.

7

**Список использованной литературы**

1. Асиновский А.С., Архипова Е.А., Богданова Н.В. и др. Полевая лингвистическая практика: Учеб.-метод. комплекс сложной структуры. - Ч. 1. Теоретические основы и методика сбора лингвистических данных для представления их в речевом корпусе русского языка. - С
2. Баранов А.Н. Введение в прикладную лингвистику. - М., 2001.
3. Гайдамакин Н.А. Автоматизированные информационные системы, базы и банки данных. Вводный курс: учеб. пособие. - М.: Гелиос АРВ, 2002. - 368 с.

Мишанкина Н.А., Тубалова И.В., Эмер Ю.А. Филология и информатика: специфика электронного представления региональных фольклорных текстов // Гуманитарная информатика: Сб. статей / под ред. Г.В. Можаевой. - Томск: Изд-во Том. ун-та, 2004. - Вып. I. - С. 102

8