

ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА «ОРФОГРАФИЧЕСКИЙ КОНКОРДАНС»

Активное развитие новой сферы накопления, хранения, обработки и передачи информации – Интернета – требует современных подходов к созданию информационных ресурсов лингвистического характера и системе их поиска. Возможность и необходимость использования современных информационных технологий для развития русской лингвистической науки и лингводидактики трудно переоценить. Международные конгрессы и научно-практические конференции исследователей русского языка, деятельность отечественных и зарубежных ученых-лингвистов (Польша, Германия, Австрия, Югославия, Франция, США, Иран, Италия, Китай, Швеция, Испания, Болгария, Дания, Япония) свидетельствуют о востребованности лингвистической информации о русском языке в мировом научном сообществе (См.: Международный конгресс русистов-исследователей: Труды и материалы. — М., 2007). Вопросы теории и методики изучения русского языка всех уровней (от фонетики до синтаксиса), стилистической системы, языка художественных произведений, активных процессов в современном русском языке, русской языковой картины мира, сопоставительного изучения русского языка с другими языками, текстологических, социолингвистических и когнитивных аспектов исследования русского языка, его места в межнациональной коммуникации, русской компьютерной и квантитативной лингвистики, преподавание русского языка в иностранной школе – вот далеко не полный перечень вопросов и проблем, которыми занимаются русисты-исследователи в России и за рубежом. Возможность обращения ученых-лингвистов к высокотехнологичным способам хранения и обработки научной информации вносит весомый вклад в отечественную и мировую лингвистическую науку

Не менее важным является обеспечение доступа учащихся, студентов, преподавателей общеобразовательных и высших профессиональных учебных заведений к современным сетевым информационным ресурсам для развития

системы открытого образования с использованием интерактивных дистанционных технологий обучения.

Необходимость организации оперативного удобного доступа к лингвистической информации связана также с потребностями языковой практики всех членов нашего общества. Нужно прежде всего лингвистически образовывать общество, и не только в плане примитивной нормализации: говори так, а не так, хотя это тоже очень нужно, но в плане повышения общей языковой, филологической грамотности людей, привития и воспитания у них интереса к родной речи, так как язык — ворота к мышлению, развивая язык, человек развивает свою способность мыслить. Через культуру речи, культуру общения, культуру мышления лингвисты могут оказать благотворное воздействие на эволюцию общества в целом.

Информационно-поисковая система «Орфографический конкорданс» <http://lexrus.ru/default.aspx?p=3125> является новой разработкой в области компьютерной лингвистики. Её новизна обусловлена прежде всего особым подходом к решению задач языкознания и обучения русскому языку, основанным на междисциплинарном взаимодействии гуманитарных и естественнонаучных дисциплин, и жёсткой ориентацией на определённый сектор лингвистической семиосферы. Последнее (целевая направленность ресурса) позволяет с максимальной эффективностью использовать академические разработки лингвистики.

Информационно-поисковая система «Орфографический конкорданс» (ИПС «Орфографический конкорданс») включает в себя специальные программы обработки и хранения баз данных, компьютерных картотек, программ обработки текста, позволяет в режиме реального времени динамически формировать словарные статьи и текстовые материалы на основе поисковых запросов пользователей.

Система, прежде всего, ориентирована на взаимодействие с пользователями Интернета, интересующимися вопросами русского языка и

лингвистики. Модульный характер реализации системы обеспечивает, в рамках единого пользовательского интерфейса, хранение и обработку различных форматов данных, что позволяет не ограничивать возможности разработчиков по расширению и модернизации как отдельных частей ресурса, так и системы в целом. Благодаря этому, система обладает потенциалом развития.

Важно отметить факторы доступности, надёжности, удобства использования системы, ориентированной на пользователя Интернета. В то же время в ней реализованы сложные, с точки зрения пользователя, и скрытые от последнего, но необходимые алгоритмы хранения и обработки оцифрованной информации лингвистического содержания, которые позволяют использовать информационный потенциал не только в системах «человек – машина», но и «машина – машина». Последнее предоставляет возможность практически неограниченного наращивания информационной мощности системы не только с учётом существующих технологий обработки информации на ЭВМ, но и с учётом перспектив развития отраслей компьютерной лингвистики, информатики и цифровых технологий в целом.

Принципиальной позицией создателей данной разработки является реализация понятного и не отягощённого избыточной информацией интерфейса взаимодействия пользователя с системой, но в то же время позволяющего реализовывать как простые, так и очень сложные запросы к базам данных.

Совокупность предоставляемых возможностей не имеет аналогов в сфере ресурсов Интернета, как по комплексности, так и по алгоритмизации решения задач.

Краткий обзор предоставляемых возможностей системы:

✓ Создание как простых, так и сложных запросов к словарным базам данных с использованием логических операторов AND, NOT, OR (И, НЕТ, ИЛИ) в любых их комбинациях, масок замены единичных символов «?» и групп символов «*».

✓ Возможность построения сложных логических запросов с использованием скобок.

✓ Автоматическая подсветка символов строгого соответствия запросов обеспечивает наглядную ориентацию в результатах работы поисковой системы.

✓ Использование собственного программного обеспечения позволило сделать систему автономной и независимой от сторонних производителей и разработчиков, что обеспечивает надёжность работы всех модулей системы как единого целого.

✓ Ориентация на российского пользователя предопределила фактор визуальной реализации интерфейса, что позволяет использовать возможности системы на любых компьютерах, подключённых к сети Интернет, не создавая дискомфорта из-за разницы времени загрузки страниц сайта или несовместимости программного обеспечения.

✓ Реализованная на надёжном хостинге 1GB.RU система авторизации пользователей позволяет гарантировать защиту содержательной информации проекта, что обеспечивает надёжную защиту авторских прав (невозможность тотального копирования данных) и защиту от злонамеренного изменения или уничтожения данных.

✓ Принцип многоуровневого администрирования системы позволяет обеспечивать работу администраторов, реализующих разные задачи в зависимости от их профессиональной принадлежности и компетенции в тех или иных вопросах, что гарантирует профессиональное администрирование на всех содержательных уровнях системы.

Существенным обстоятельством является то, что доступ ко всем перечисленным возможностям реализован через Интернет, посредством обычного Web-браузера

ИПС «Орфографический конкорданс» снабжена справочной информацией, представленной как в кратком изложении, достаточном для задания простых запросов к базам данных, так и в подробной форме,

позволяющей на наглядных примерах быстро освоить профессиональное пользование системой.

Таким образом, ИПС «Орфографический конкорданс», прежде всего ориентирована на достижение максимально эффективной и комфортной для пользователя машинной обработки данных по схеме «запрос – результат запроса» в системе «человек – машина».

Разработка программного обеспечения web-сайта базируется на технологии .NET компании Microsoft. В качестве среды выполнения используется Microsoft SQL Server 2008, MS ASP NET 4.0 и Net Framework 4.0.

Содержательная часть инженерно-технических работ ИПС «Орфографический конкорданс»:

- Создание структуры базы данных и перенос в неё имеющихся исходных текстов.
- Разработка модуля поиска, связывающего поисковую форму и встроенный поисковый механизм MS SQL Server.
- Разработка интерфейса поисковых форм, позволяющего делать разнотипные запросы к информационной базе данных.
- Разработка механизма предобработки поисковых запросов.
- Встраивание ИПС «Орфографический конкорданс» в дизайн lexrus.ru.

Информатизация современного социума, развитие соответствующих технологий, количественное увеличение объемов коммуникаций потребовали новых подходов к принципам поиска лингвистической информации в Интернете. Некоторые специалисты считают, что в современном Интернете размещено почти два миллиарда документов. Чтобы в гигантском потоке информации найти материалы лингвистического характера необходима особая система поиска, так как существующие системы Яндекс, Рамблер, Google, обеспечивающие 88% всего поиска русскоязычного Интернета, не ставят своей задачей актуализацию лингвистических аспектов найденных

слов. Недостатками существующих поисковых систем в контексте специфики лингвистической информации являются:

наличие «лишней» информации, когда на запрос пользователя, желающего уточнить значение слова или его грамматические особенности (образование падежных форм, множественного числа существительных, определенных глагольных форм и т. д.), поисковая система выдает огромное количество адресов, среди которых не всегда есть те, которые представят лингвистическую информацию.

трудности в поиске лингвистической информации на основе формальных грамматических признаков: отсутствие возможности выбирать слова по основам, корням, приставкам, суффиксам.

отсутствие системы справочной службы по вопросам грамматики, лексики, орфографии русского языка,

Преимущества ИПС «Орфографический конкорданс» состоят в том, что в ней реализована возможность предъявления пользователю лингвистического материала в системе, компоненты которой отражают три аспекта функционирования языка: язык — как совокупность текстов, язык — как структура (в конечном счете тоже сводимая к совокупности, но уже единиц и правил), и язык — как способность создавать тексты. Указанный подход к предъявлению лингвистических материалов отражает общее направление развития языкознания, в котором на рубеже 80-х годов XX века произошел «переход от структурной модели языка к коммуникативно-функциональной». Теория связывает системно-структурные и функциональные свойства языковых единиц, что находит отражение в лингвистических исследованиях.

Лингвистические материалы ИПС «Орфографический конкорданс» помогут и ученым, и рядовым носителям языка решить возникающие проблемы.

Необходимость таких материалов в мировой информационной сети продиктована потребностями развития компьютерных технологий и информатизации общества в целом.

ИПС «Орфографический конкорданс» призвана обеспечить оптимальные условия для предоставления информационных услуг отечественным и зарубежным пользователям.

Гибкая организация информации, её интегрированное представление, открытая архитектура системы являются ключевыми моментами в создании ИПС «Орфографический конкорданс».