

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА И ПЕРСПЕКТИВЫ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Бул макалада Компьютердик лингвистика багыты боюнча кесип ээлерин даярдоонун жана лингвистикалык маселелерди чечүүдө келечектүү маалымат технологияларын түзүүнүн келечеги жана көйгөй маселелери каралат.

В данной статье рассматривается проблемы и перспективы создания перспективных информационных технологий для решения лингвистических задач и подготовка специалистов по направлению «Компьютерная лингвистика».

This article is devoted to the problems and prospects of the of creation of future IT for solving linguistic tasks and preparation of the specialists on Computer linguistics.

В жизни современного общества важную роль играют автоматизированные информационные технологии. С течением времени их значение непрерывно возрастает. Но развитие информационных технологий происходит весьма неравномерно: если современный уровень вычислительной техники и средств связи поражает воображение, то в области смысловой обработки информации успехи значительно скромнее. Эти успехи зависят, прежде всего, от достижений в изучении процессов человеческого мышления, процессов речевого общения между людьми и от умения моделировать эти процессы на ЭВМ.

Когда речь идет о создании перспективных информационных технологий, то проблемы автоматической обработки текстовой информации, представленной на естественных языках, выступают на передний план. Это определяется тем, что мышление человека тесно связано с его языком. Более того, естественный язык является инструментом мышления. Он является также универсальным средством общения между людьми – средством восприятия, накопления, хранения, обработки и передачи информации. Проблемами использования естественного языка в системах автоматической обработки информации занимается наука компьютерная лингвистика. За прошедшие полвека в области компьютерной лингвистики были получены значительные научные и практические результаты: были созданы системы машинного перевода текстов с одних естественных языков на другие, системы автоматизированного поиска информации в текстах, системы автоматического анализа и синтеза устной речи и многие другие.

Компьютерная лингвистика особенно активно развивается во всем мире в последнее десятилетие в связи с растущим влиянием Интернета.

В Кыргызстане по подготовке специалистов по направлению «Компьютерная лингвистика» впервые начали заниматься институт новых информационных технологий КГУСТА им.Н.Исанова. В 2007 году на базе кафедры Лингвистика и новые информационные технологии ИНИТ КГУСТА открылась новая кафедра «Компьютерная лингвистика», которая начала подготовку специалистов по данному направлению. В качестве головного вуза кафедрой разработано и утверждена Государственный образовательный стандарт специальности «Компьютерная лингвистика».

На основании лицензии МОиН КР кафедра ведет подготовку бакалавров и магистров по направлению «Компьютерная лингвистика» с профилями: «Компьютерная лингвистика и медиалингвистика»; «Компьютерная лингвистика и лингвистическая экспертиза»; «Компьютерная лингвистика и Межкультурная коммуникация». Учебный план направления разработан таким образом, чтобы наряду с классическими лингвистическими дисциплинами, студенты получают практические навыки работы лингвистическим материалом с применением ИТ, его анализом и приложением для

конкретных задач по автоматической обработке естественного языка. Соответственно блок специальных дисциплин включает такие курсы как "Основы искусственного интеллекта", "Автоматическая обработка естественного языка", "Информационно-поисковые машины и системы", "CASE-технологии", "Система автоматического перевода" "WEB-дизайн", "Лингвистическое конструирование", «Речевая информатика» и др.

С практическим применением компьютерной лингвистики каждый день имеют дело почти все пользователи ПК — это, например, проверка правописания, грамматики и стиля в тестовых редакторах. Давно и широко присутствуют на рынке различные средства распознавания печатных и рукописных текстов. Можно также назвать распознавание (диктовка, слитная) и синтез речи, машинный перевод текста и речи, генерацию текста, поиск нужного документа по запросу в информационных ресурсах, реферирование (смысловое сжатие), анализ содержания текстов (классификация; кластеризация; установление сходства, фильтрация нежелательных документов, анализ тональности и пр.), вопросно-ответные системы и системы логического вывода и разного рода другие системы извлечения знаний.

Центральными научными проблемами компьютерной лингвистики являются проблема моделирования процесса понимания смысла текстов (перехода от текста к формализованному представлению его смысла) и проблема синтеза речи (перехода от формализованного представления смысла к текстам на естественном языке). Эти проблемы возникают при решении ряда прикладных задач и, в частности, задач автоматического обнаружения и исправления ошибок при вводе текстов в ЭВМ, автоматического анализа и синтеза устной речи, автоматического перевода текстов с одних языков на другие, общения с ЭВМ на естественном языке, автоматической классификации и индексирования текстовых документов, их автоматического реферирования, поиска документов в полнотекстовых базах данных.

Можно отметить серьезные научные и практические достижения в области компьютерной лингвистики. Так в ряде стран (Россия, США, Япония, и др.) построены экспериментальные и промышленные системы машинного перевода текстов с одних языков на другие, построен ряд экспериментальных систем общения с ЭВМ на естественном языке, ведутся работы по созданию терминологических банков данных, тезаурусов, двуязычных и многоязычных машинных словарей (Россия, США, Германия, Франция и др.), строятся системы автоматического анализа и синтеза устной речи (Россия, США, Япония и др.), ведутся исследования в области построения моделей естественных языков.

Научное направление кафедры ЛНИТ была начата под руководством член корр. НАН КР, профессора Ж.К.Сыдыкова. Под его руководством за последние годы с применением современные информационные технологии успешно защитили диссертационные работы на соискание ученой степени кандидата филологических наук преподаватели данной кафедры: Г.Э.Жумалиева, З.П.Акжолова, С.Ж.Карабаева, Г.М.Айдыралиева и М.Ж.Тумонбаева.

Силами ППС данной кафедры с 2006 года один раз за два года проводится международная научно-практическая конференция «Исследование и преподавание государственного, официального и иностранных языков с применением информационных технологий».

Кафедра имеет научно-исследовательские связи с Институтом языкознания и литературы (ИЯиЛ) им.Ч.Айтматова НАН КР, Департаментом по государственному языку при Президенте КР и др. В 2008 году на основе договора между КГУСТА и НАН КР при ИЯиЛ НАН КР открыт филиал кафедры «Научно-исследовательский центр ЛНИТ», где ППС, соискатели, магистранты и студенты ведут научно-исследовательскую работу.

С целью развития государственного языка и его внедрение посредством информационных технологий, усовершенствования компьютерных терминологий на кыргызском языке и их внедрение в систему образования информационных терминологий с английского на кыргызский язык кафедра занимался научной темой «Исследование, разработка и внедрение в систему образования компьютерной

терминологии на государственном языке», в результате был переведен пакет языкового интерфейса MS Office 2010 на государственный язык.

Члены кафедры занимались с научными проектами «Разработка и внедрение Автоматизированной системы НАК КР», «Разработка и внедрение информационной системы МОиН КР», Разработка и внедрение информационной системы Департамента по государственному языку при Президенте КР», где созданы информационно-поисковые порталы.

Перед компьютерной лингвистикой стоят, прежде всего, задачи лингвистического обеспечения процессов сбора, накопления, обработки и поиска информации. Наиболее важными из них являются:

1. Автоматизация составления и лингвистической обработки машинных словарей;
2. Автоматизация процессов обнаружения и исправления ошибок при вводе текстов в ЭВМ;
3. Автоматическое индексирование документов и информационных запросов;
4. Автоматическая классификация и реферирование документов;
5. Лингвистическое обеспечение процессов поиска информации в одноязычных и многоязычных базах данных;
6. Машинный перевод текстов с одних естественных языков на другие;
7. Построение лингвистических процессоров, обеспечивающих общение пользователей с автоматизированными интеллектуальными информационными системами (в частности, с экспертными системами) на естественном языке, или на языке, близком к естественному;
8. Извлечение фактографической информации из неформализованных текстов.

Многие вузы Кыргызстана (КНУ, ОшТУ, КГУ) приступили к обучению студентов по направлению «Компьютерная лингвистика», где специалисты будут заниматься решением вышеуказанных проблем. По мнению экспертов, представителей академической и отраслевой среды, все необходимые условия для этого созрели: отечественная наука имеет в этой сфере серьезную основу, на рынке видно определенное повышение спроса на таких профессионалов, есть интересы к подобной деятельности со стороны молодежи, выбирающей для себя жизненный путь, имеется готовность оказания спонсорской (методической, технической, финансовой) поддержки развитию этого направления в вузах со стороны ИТ-отрасли.

Список литературы

1. Белоногов Г.Г., Зеленков Ю.Г., Новоселов А.П., Хорошилов Ал-др А., Хорошилов Ал-сей А. Метод аналогии в компьютерной лингвистике. Сб. «Научно-техническая информация» сер. 2, N 1, ВИНТИ, 2000 г.
2. Панов Д.Ю., Ляпунов А.А., Мухин И.С. Автоматизация перевода с одного языка на другой: Сессия по научным проблемам автоматизации.
3. http://www.krugosvet.ru/enc/gumanitarnye_nauki/lingvistika/KOMPYUTERNAYA_LINGVISTIKA.html - Онлайн Энциклопедия Кругосвет (описание компьютерной лингвистики, её направления)
4. http://ru.wikipedia.org/wiki/Computational_linguistics - Википедия. Свободная Энциклопедия. (использовалась при описании направлений компьютерной лингвистики (в частности машинного перевода)).