Московский педагогический государственный университет

А. В. Гуслякова

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ЛИНГВИСТИКА XXI ВЕКА

Москва 2016

Алла Викторовна Гуслякова

Информационные технологии и лингвистика XXI века

Текст предоставлен правообладателем
http://www.litres.ru/pages/biblio_book/?art=28259959
Информационные технологии и лингвистика XXI века: Учебное пособие / А. В. Гуслякова.: МИГУ; Москва; 2016
ISBN 978-5-4263-0398-0

Аннотация

Учебное пособие «Информационные технологии и лингвистика XXI века» посвящено изучению ряда научных вопросов, связанных с проблемами компьютерного перевода; различных систем памяти переводов; правилами эффективного информационного поиска. В учебном пособие также затрагиваются вопросы изучения компьютерной лингвистики и становления новой лингвистической отрасли Интернет-лингвистики. Данное учебное пособие ориентировано на студентов языковых факультетов, институтов, вузов, изучающих дисциплины «Информационные технологии в лингвистике» и «Компьютерный перевод», а также на всех интересующихся проблемами современной лингвистики в ее взаимодействии с цифровыми технологиями и Интернетом.

Содержание

Алла Гуслякова
Информационные технологии и лингвистика XXI века: Учебное пособие

Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Московский педагогический государственный университет»

Рецензенты:

И. В. Тараканова, кандидат филологических наук, профессор Института иностранных языков МИГУ

И. В. Вашунина, доктор филологических наук, профессор Всероссийской академии внешней торговли Министерства экономического развития РФ

ВВЕДЕНИЕ

Современное общество не мыслит своего существование без интерактивной информационной среды с круглосуточным доступом в Интернет-пространство. Поразительная скорость развития и совершенствования информационных технологий сближает деятельность человека и телекоммуникационных машин, переходя за рамки рабочего пространства и превращаясь в ежедневную составляющую каждого из нас. Человек учится общаться с компьютерной средой и взаимодействовать с другими людьми посредством нее. Таким образом, возрастает роль языка, межкультурной коммуникации и лингвистической науки в целом в телекоммуникационном пространстве.

Учебное пособие «Информационные технологии и лингвистика XXI века» предлагает рассмотреть наиболее актуальные проблемы взаимодействия современного лингвистического знания и цифровых технологий. В частности, затрагиваются такие значимые направления, как роль виртуального пространства в развитии лингвистики; основы компьютерного перевода и переводческие системы; искусственный интеллект, интеллектуальные системы; Интернет-лингвистика как новое научное направление в развитии лингвистической отрасли.

Учебное пособие «Информационные технологии и линг-

вистика XXI века» включает в себя лекционный материал и задания для проведения семинарских занятий, а также вопросы и задания для самостоятельной работы студентов. Каждый раздел пособия завершается списком рекомендованной литературы. В конце учебного пособия представлены приложения, включающие информацию по ключевым особенностям профессии *лингвист*; по основным программам машинного и автоматизированного перевода текстов; краткую характеристику наиболее популярных поисковых систем и др. информацию.

Учебное пособие «Информационные технологии и лингвистика XXI века» ориентировано на студентов гуманитарных факультетов и вузов, изучающих дисциплины «Информационные технологии в лингвистике» и «Компьютерный перевод», а также может быть полезным для всех интересующихся проблемами лингвистики XXI века в ее взаимодействии с цифровыми технологиями современного тысячелетия.

Глава 1
КОМПЬЮТЕР, ИНТЕРНЕТ И ЛИНГВИСТИКА

Краткое описание: Поиск и публикация информации в Интернете. Использование компьютера в гуманитарных исследованиях. Компьютер и информационные технологии как усилители интеллектуальной деятельности. Современные парадигмы переводческого процесса.

Главной причиной, почему люди будут покупать себе домой компьютер, станет возможность быть связанными с национальной коммуникационной сетью.

Мы сейчас в самом начале этого этапа. но это будет настоящий прорыв. Примерно как телефон.
Стив Джобс, 1985 год.

Любая достаточно развитая технология неотличима от волшебства.
Артур Кларк.

Любая реальность является суммой информационных технологий.
Виктор Пелевин.

Интернет изменяет всё, чего бы ни коснулся, а касается он практически всего.
Джон Эллис.

Современный период развития цивилизации характеризуется переходом человечества от индустриального общества к информационному обществу. Основным перерабатываемым «сырьем» становится информация. Труд современников делается в меньшей степени физическим и в большей степени интеллектуальным. В наиболее развитых странах производство информации и разработка информационных технологий стало одной из самых прибыльных и стремительно растущих отраслей.

Одной из самых важных функций, реализованных в Интернете, является поиск информации. Неисчислимые объемы информации представлены в сети так, что можно потратить огромное количество времени, просто переходя из од-

ного раздела в другой и определяя, какая информация имеется в наличии. Это является первой проблемой, которая связана не столько с имеющимся оборудованием, сколько с культурой пользования и быстрым поиском нужной информации. Быстрота связи с сервером не равна быстроте получения информации по той причине, что ее может там и не быть. Поэтому начинающие исследователи первоначально пытаются связаться с широко известными серверами, хотя нужная информация лежит совсем в другом месте. Вторая проблема больше связана с исследователями старшего поколения, часть из которых не всегда может воспринимать работу с новыми информационными технологиями. В целом же подключение исследователей к данному типу источника информации ничего кроме пользы не несет.

Таким образом, появляется необходимость подготовить человека к быстрому восприятию и обработке больших объемов информации, овладению им современными средствами и технологией работы. Кроме того, новые условия работы порождают зависимость информированности одного человека от информации, приобретенной другими людьми. Поэтому недостаточно уметь самостоятельно осваивать и накапливать информацию. Необходимо научиться такой технологии работы с информацией, при которой подготавливаются и принимаются решения на основе коллективного знания. Это говорит о том, что человек должен иметь определенный уровень культуры по обращению с информацией.

Для отражения этого факта был введен термин информационная культура.

Информационная культура (education culture) – умение целенаправленно работать с информацией и использовать ее для получения, обработки и передачи компьютерную информационную технологию, современные технические средства и методы. Информационная культура проявляется в следующих аспектах.

1. В конкретных навыках по использованию технических устройств (от телефона до персонального компьютера и компьютерных сетей).

2. В способности использовать в своей деятельности компьютерную информационную технологию, базовой составляющей которой являются многочисленные программные продукты.

3. В умении извлекать информацию из различных источников: как из периодической печати, так и из электронных коммуникаций, представлять ее в понятном виде и уметь ее эффективно использовать.

4. Во владении основами аналитической переработки информации; в умении работать с различной информацией; в знании особенностей информационных потоков в своей области деятельности.

Информационная культура вбирает в себя знания из тех наук, которые способствуют ее развитию и приспособлению к конкретному виду деятельности. В первую очередь это –

информатика, кибернетика, теория информации, математика, теория проектирования баз данных, лингвистика и ряд других дисциплин.

Внедрение новых информационных технологий во все сферы современной жизни привело к тому, что умение работать на компьютере является необходимым атрибутом профессиональной деятельности любого специалиста и во многом определяет уровень его востребованности в обществе.

Компьютерные технологии не сразу нашли применение в гуманитарных науках, развиваясь, прежде всего, с учетом потребностей точных наук. Гуманитарии не рассматривали компьютер как реальный научный инструмент, способный изменить характер исследования. Применение информационных технологий долго не использовалось в гуманитарных науках. Хотя после появления микрокомпьютеров обработка текстов быстро стала наиболее распространенной сферой их применения, историки не спешили раскрыть для себя возможности нового средства. Компьютерные технологии развивались, исходя из потребностей точных наук, а связи между гуманитарными и точными науками не всегда были столь прочными, какими они становятся в настоящее время.

Ситуация начала меняться с появлением первых программ, предназначенных для контроля знаний и обучения гуманитарным дисциплинам. Другой важный аспект «компьютерной революции» был связан с возрастанием интереса к созданию баз данных.

Информационные технологии все глубже внедряются в сферу гуманитарных исследований: формируются информационные системы для различных научных направлений, компьютерная техника и медиатехнологии становятся важными средствами повышения эффективности исследований. Появляются новые направления, такие как историческая информатика, компьютерная лингвистика; компьютерные технологии используются в археологии, этнографии, графологии, истории, экономике, социологии, юриспруденции, педагогике, литературоведении, журналистике.

В последние годы все более настоятельно требуется обновление и расширение арсенала средств и методов, которые находятся в распоряжении специалиста. Многими учеными неоднократно отмечалось, что эффективность научных исследований во многом зависит от того, насколько хорошо разработана их методологическая и методическая базы. Поэтому все более актуальным становится вопрос о необходимости разработки принципиально новых подходов, инструментария гуманитарных исследований – персональных информационно-исследовательских систем, интегрированных в международные компьютерные сети.

Российские специалисты все активнее включаются в разработку перспективных проблем использования компьютерной техники в гуманитарных исследованиях, методов искусственного интеллекта, мультимедиа технологий, применения глобальной сети Интернет. Именно эти направления яв-

ляются определяющими в развитии гуманитарных наук в будущем.

На данный момент Интернет[1] является одним из наиболее престижных механизмов, используемых для общения и получения информации электронным путем. Основное его преимущество – это всеохватывающая природа информации и услуг, которые он оказывает. Исследователи могут использовать компьютерную сеть для обмена посланиями и файлами друг с другом, могут получить информацию практически из любой части мира. Однако использование Интернета в исследовательских целях становится все более и более распространенным. В глобальной сети появляется все большее количество необходимой информации, представляющей интерес для гуманитарных исследований.

Информационные технологии, основанные на Интернете, телекоммуникационных сетях и интеллектуальных компьютерных системах, открывают перед будущим поколением возможности свободного распространения знаний, различных сведений и материалов. Ему придется столкнуться с необходимостью приспосабливаться к новой социальной среде, где информация и научное знание станут основными факторами, определяющими потенциал общества и перспективы его развития. Использование единых мировых инфор-

[1] Интернет (World Wide Web (WWW) «всемирная паутина») – это *гетерогенная система*, то есть соединение разнообразных аппаратных платформ, исполняющих приложения, предназначенные для решения широкого диапазона задач.

мационных систем обеспечивает внедрение информационных технологий в образование: формируется единое образовательное пространство, возрастает потребность человека в общении, и получении доступа к общим нематериальным ресурсам, осмыслении и переработке большого объема информации.

Смысл информатизации образования заключается в создании, как для педагогов, так и для обучаемых благоприятных условий для свободного доступа к культурной, учебной и научной информации. Необходимо также понимать, что информатизация сферы образования должна опережать информатизацию других направлений общественной деятельности, поскольку именно здесь закладываются социальные, психологические, общекультурные, а также профессиональные предпосылки развития общества нового типа. Информатизация и компьютеризация становятся новыми объектами изучения, применения и использования в образовании, что дает возможность выйти на создание определенной системы образования.

Информационные технологии можно рассматривать как элемент и функцию информационного общества, направленную на регулирование, сохранение, поддержание и совершенствование системы управления нового сетевого общества. Если на протяжении веков информация и знания передавались на основе правил и предписаний, традиций и обычаев, культурных образцов и стереотипов, то сегодня главная

роль отводится технологиям. Информационные технологии упорядочивают потоки информации на глобальном, региональном и локальном уровнях. Они играют ключевую роль в формировании техноструктуры, в повышении роли образования и активно внедряются во все сферы социально-политической и культурной жизни, включая домашний быт, развлечения и досуг.

Общество с высоким уровнем развития и использования информационных технологий, развитыми инфраструктурами, обеспечивающими производство информационных ресурсов и возможность доступа к информации, называют **информационным обществом**[2] (information society). Само название «информационное общество» впервые появилось в Японии в середине 60-х годов XX века. Оно стало основным в докладе специальной группы по научным, техническим и

[2] Теория «информационного общества» была развита такими известными авторами, как М. Порат, Й. Масуда, Т. Стоуньер, Р. Карц и др. (Porat M., Rubin M. The Information Economy: Development and Measurement. Wash., 1978; Masuda Y. The information Society as Post-Industial Society. Wash., 1981; Stonier T. The Wealf of Information. L., 1983; Katz R.L. The Information Society: An International Perspective. N.Y., 1988.); в той или иной мере она получила поддержку со стороны тех исследователей, которые акцентировали внимание не столько на прогрессе собственно информационных технологий, сколько на становлении технологического или технетронного (technetronic – от греч. techne) общества (Brzezinski Zb. Between Two Ages. N.Y., 1988.), или же обозначали современный социум, отталкиваясь от возросшей или возрастающей роли знаний как «the knowledgeable society», «knowledge society» или «knowledge-value society». Сегодня существуют десятки понятий, предложенных для обозначения отдельных признаков современного общества.

экономическим исследованиям, созданной японским правительством для выработки перспектив развития экономики страны.

Специалисты, предложившие этот термин, разъяснили, что он характеризует общество, в котором в изобилии циркулирует высокая по качеству информация, а также есть все необходимые средства для ее хранения, распределения и использования. Информация легко и быстро распространяется по требованиям заинтересованных людей и организаций и выдается им в привычной для них форме. Стоимость пользования информационными услугами настолько невысока, что они доступны каждому.

Отличительными особенностями информационного общества являются: открытость, технологичность (особенность информатизации), интеллектуальность, доступ к мировым информационным ресурсам, высокая степень обеспечения безопасности, гибкость и самоорганизация выше указанных систем. В таком обществе наблюдается ускоренная автоматизация и роботизация всех отраслей производства и управления, происходят радикальные изменения социальных структур. Эти изменения приводят к расширению сферы информационной деятельности и вызывают необходимость подготовки специалистов в области разработки и сопровождения информационных технологий, требуют повышения информационной культуры граждан.

Формирование в стране информационного общества

неразрывно связано с уровнем образования в данном обществе. В истории человечества было, по крайней мере, две революции по улучшению качества и расширению доступности образования. Две предыдущие революции одновременно расширили возможности образования как системы, добавив новые средства и изменив ее структуру. Был осуществлен переход: от устного диалога времен Сократа – к образовательным формам, которые включили чтение и письмо; от ученых времен раннего Средневековья, обучающих независимых учеников тогда, когда им заблагорассудится, – к новой образовательной структуре, в которой организованные ученые и студенты работают вместе в пределах университета, колледжа, а учителя и ученики объединены в стенах школы.

Таким образом, информационные технологии вошли во все сферы нашей жизни. Компьютер является средством повышения эффективности процесса обучения, участвует во всех видах человеческой деятельности, незаменим для социальной сферы. Вот уже почти два десятка лет не утихают споры о том, какое место должен занимать компьютер в профессиональной деятельности педагога. Современные информационные технологии – это аппаратно-программные средства, базирующиеся на использовании вычислительной техники, которые обеспечивают хранение и обработку образовательной информации, доставку ее обучаемому, интерактивное взаимодействие студента с преподавателем или педа-

гогическим программным средством, а также тестирование знаний студента.

Для подавляющего большинства современных педагогов должны быть доступны компьютерные учебники, электронные обучающие программы, разнообразные системы тестирований в онлайн режиме, интерактивное взаимодействие обучающего и обучаемого.

В учебном процессе важны не информационные технологии сами по себе, а то, насколько их использование служит достижению собственно образовательных целей. При выборе технологий необходимо учитывать наибольшее их соответствие характерным чертам обучаемых, специфическим особенностям конкретных предметных областей, преобладающим типам учебных заданий и упражнений.

Образовательные технологии (educational technology) – это эффективное использование технологических инструментов в учебном процессе, еще называемым за рубежом **e-learning.**

К образовательным технологиям относятся: видео-лекции; мультимедиа-лекции и лабораторные практикумы; электронные мультимедийные учебники; компьютерные обучающие и тестирующие системы; имитационные модели и компьютерные тренажеры; консультации и тесты с использованием телекоммуникационных средств; видеоконференции.

Таким образом, главным моментом в образовательных

технологиях становится визуализация мысли, информации, знаний. Особенностью образовательных технологий является опережающий характер их развития по отношению к техническим средствам. Дело в том, что внедрение компьютера в образование приводит к пересмотру всех компонентов процесса обучения. В интерактивной среде «ученик – компьютер – преподаватель» большое внимание должно уделяться активизации образного мышления за счет использования технологий.

Таким образом, умение применять в своей деятельности современные информационные технологии становится одним из основных компонентов профессиональной подготовки любого специалиста, в том числе и специалиста переводческой сферы.

Современное переводоведение и дидактика перевода среди прочих равных по значимости вопросов уделяют пристальное внимание развитию поисковой компетенции переводчиков в эпоху инновационного технического прогресса и технологической трансформации информационного пространства переводчиков.

Изначально процесс перевода трактуется по модели: *автор – переводчик – читатель.* Однако не стоит забывать, что перевод текста требует междисциплинарного подхода. Эволюция компьютерных технологий проникает в сферу профессионального перевода и трансформирует выше представленную схему по следующему, с нашей точки зрения, прин-

ципу: *автор – переводчик (человек) – переводчик (компьютерная программа) – читатель.*

Мультимедийные программы, отражающие ролевую функцию переводчика, позволяют более эффективным образом осуществлять формирование и развитие навыков различных видов перевода с учетом деятельностных особенностей развертывания каждого из них.

Мультимедийные переводческие программы создаются на основе общей важной цели – формирование и развитие переводческих навыков, т. е. формирование общих ролевых навыков, присущих переводчику в различных стандартно-стереотипных ситуациях переводческой деятельности.

В то же самое время при работе с электронными переводчиками обнаруживается ряд погрешностей при переводе текста. Лексический анализ переведенных текстов показал, что по большей части электронные переводчики адекватно переводят простые части речи, но допускают ошибки в переводе падежей, принадлежности прилагательных, речевых оборотов, построения предложения. Недостатком некоторых переводчиков является неточность перевода слов, имеющих несколько значений. Для более адекватного перевода в перспективе можно предложить более глубокий эвристический анализ грамматического построения предложения, с улучшением качества перевода различных частей речи и их грамматических характеристик, а так же исключить конфликт словарей при переводе специализированных текстов.

Грамматический анализ текстов показывает, что электронный переводчик справляется с переводом слов во множественном и единственном числе, но имеется определенная трудность в переводе падежей и постановки глаголов в нужное число. Это объясняется различной интерпретацией падежей в русском и английском языках: в русском – через окончание, в английском – через предлоги. Итак, компьютер пока во многом не может заменить переводчика. Стоит ли тогда вообще применять системы машинного перевода? Ответ положительный. Если компьютер используется для перевода литературных текстов, то получается черновой вариант текста, так называемый подстрочник, который превращается в произведение искусства человеком, слабо владеющим языком оригинала, но являющимся хорошим литературным редактором. Если же речь идет о переводе технических текстов, то здесь при правильном выборе словаря по специальности, в рамках которой написан текст, получается вполне удовлетворительный результат, иногда не требующий последующего вмешательства. Вообще необходимость редактирования компьютерного перевода очень часто возникает в связи с проблемами, перечисленными выше. Для этого системы машинного перевода обязательно имеют средства редактирования текстов. Для некоторых заказчиков такой уровень перевода просто неприемлем, однако других машинный перевод вполне устраивает, в значительной степени потому, что часто ему просто нет реальных альтернатив.

Общеизвестно, что хороший перевод текста – это не только творческая, но и достаточно трудоемкая работа. Причем даже самый хороший перевод, как правило, нуждается в редакторской правке. Что касается творческой части, то, с нашей точки зрения, даже в будущем в соревновании компьютер-человек всегда победит «живой» переводчик. Однако для решения проблем, обусловленных трудоемкостью процесса перевода, системы машинного перевода могут оказаться хорошим помощником. Для того чтобы это лучше понять, в следующей лекции подробнее остановимся на особенностях современных систем машинного перевода и их достоинствах.

Вопросы для самоконтроля

1. Что такое информационная культура?
2. Что такое информационное общество?
3. В какой стране впервые появился термин «информационное общество» и в каком году?
4. Раскройте определение понятия «образовательные технологии».
5. Почему перевод текста требует междисциплинарного подхода?
6. Каковы базовые недостатки машинного перевода?

Задания для самостоятельного исследования

1. Подготовить доклад по перспективам развития инфор-

мационного общества XXI тысячелетия.

2. Подготовить доклад по Интернет-лингвистике. Оценить и спрогнозировать влияние Интернет-лингвистики на общее состояние современной и будущей лингвистической науки.

3. Подготовить доклад по модели представления современного переводческого процесса *«автор – переводчик (человек) – переводчик (компьютерная программа) – читатель»*.

Рекомендованная литература

1. *Баранов А.Н.* Введение в прикладную лингвистику: учеб, пособие/ А.Н. Баранов. – Московский гос. ун-т им. М. В. Ломоносова, Филологический фак. – 3-е изд. – Москва: URSS: Изд-во ЛКИ, 2007 – 358 с.

2. *Гончаров В.Н.* Информационное общество: проблемы становления и закономерности развития: монография / В.Н. Гончаров и др.; Центр развития научного сотрудничества Новосибирск: ЦРНС, 2014 – 183 с.

3. *Кирилина А.В.* Общество – язык – культура: актуальные проблемы взаимодействия в XXI: тезисы докладов Девятой Международной научно-практической конференции, Москва, 26 ноября 2014 г. /под ред. А.В. Кирилина, В.М. Хаимова, С.В. Дмитрюк // Московский ин-т лингвистики, Московский финансово-экономический ин-т. – Москва: Московский ин-т лингвистики, 2014 – 88 с.

4. *Коммисаров, В.Н.* Теория перевода (лингвистические аспекты): Учеб, для ин-тов и фак. иностр. яз. / В. Н. Комиссаров. – Репр. изд. – Москва: Альянс, 2013. – 250 с.

5. *Кутузов, А.Б.* Компьютерные технологии в формировании профессиональной компетенции переводчика / А.Б. Кутузов // Языки профессиональной коммуникации: сборник статей Третьей международной научной конференции, т.2. – Челябинск, 2007 г. [Электронный ресурс]. – URL: http://tc.utmn.ru/files/kutuzov_it.pdf

6. *Тузовский, И.Д.* Утопия-XXI: глобальный проект «Информационное общество» /И.Д. Тузовский. – Челябинск: Челяб. гос. акад. культуры и искусств, 2014. – 392 с.

7. *Сокирко А.* Будущее машинного перевода / А. Сокирко // Компьютерра. – 2002. – № 21. – [Электронный ресурс]. – URL: http://old.computerra. m/offline/2002/446/18251/.

8. *Попов С.А.* Информационные технологии в лингвистике: учебное пособие / С. А. Попов, Е.Ф. Жукова. – Великий Новгород: Новгородский гос. ун-т, 2014. – 235 с.

9. *Анисимов Д.В.* Правда о машинном перевод /Д.В. Анисимов. – М.: Сам Полиграфист, 2014. – 340 с.

10. *Crystal D.* English as a Global Language. Cambridge: Cambridge University Press, 2003. – 212p.

Глава 2
КОМПЬЮТЕР И ПЕРЕВОДЧИК

Краткое описание: общие требования, автоматический перевод, автоматизированный перевод, системы памяти переводов (Translation Memory), использование систем памяти переводов Trados и OmegaT

Переводчики – почтовые лошади просвещения.
А. С. Пушкин

«Мой дед Лозинский – в моём детском и подростковом понимании – превращал мутную жесть иностранных языков в прозрачное золото русской речи. Писатели писали что-то там своё комковатое, – а он выпрямлял и разглаживал; они тыркались об стену, не находя дверей; – а он распахивал ворота; они водили

руками в тумане, – а он брал их за руку и выводил на свет. <…> Мне открылась вдруг сполна тяжесть каторжного труда переводчика, подвиг его, ничем практически не вознаграждаемый, мало кем ценимый, потаенный. Подобно Адаму, он вынужден вновь и вновь давать имена вещам, предметам, понятиям чужого, тайно и буйно цветущего за оградой сада»

Т. Толстая. Переводные картинки// Т. Толстая. Река. – М.: Эксмо, 2007.

«The word 'translation' comes, etymologically, from the Latin for 'bearing across'. Having been borne across the world, we are translated men. It is normally supposed that something always gets lost in translation; I cling, obstinately to the notion that something can also be gained.»

Salman Rushdie, Imaginary Homelands: Essays and Criticism 1981–1991

В наши дни невозможно представить себе работу переводчика без персонального компьютера. Компьютер используются как для собственно перевода, так и для решения сопутствующих задач, возникающих в деятельности переводчика, включая: изучение конъюнктуры рынка и поиск заказчиков; поиск справочной информации и изучение терминологии; подготовка коммерческих предложений; поддержание связи с заказчиками и соисполнителями; создание и обновление глоссариев и др. справочных материалов; учет заказов, планирование и учет затрат рабочего времени; выставление счетов; закупка и оплата оргтехники, справочных и расходных

материалов; бухгалтерский учет, подготовка и подача налоговых деклараций; профессиональное общение с коллегами; повышение квалификации.

Одним словом, персональный компьютер (далее ПК) является важнейшим техническим средством в инструментарии переводчика. Для того, чтобы работа переводчика с помощью ПК была эффективной с рациональным использованием времени, рабочий компьютер должен соответствовать ряду определенных требований.

I. ***Аппаратная база***

Требования к аппаратной базе (равно как и к программному обеспечению и компьютерной грамотности переводчика) определяются двумя обстоятельствами.

С одной стороны, переводчик является независимым производителем перевода. В этом качестве ему нужен текстовый процессор с проверкой орфографии на языке перевода, средства доступа к Интернету, словари и в некоторых случаях накопители переводов. Все эти программы сами по себе не отличаются особо жесткими требованиями к аппаратной базе.

С другой стороны, переводчик является подчиненным звеном в технологической цепочке заказчика, в связи с чем встает вопрос о пригодности используемых переводчиком технических средств для обработки файлов, создаваемых заказчиком, а это зачастую означает необходимость использования значительно более мощного ПК.

В целом при выборе ПК можно руководствоваться следующими соображениями:

1. расчетный срок службы – 3 года;

2. модернизация вместо замены исключается;

3. рассматриваются модели, попадающие в среднюю треть общего диапазона производительности имеющихся в продаже новых ПК.

Для письменного переводчика самое важное в компьютере – экран монитора, клавиатура и мышь, поскольку от их качества зависит не только производительность труда, но и здоровье, и сохранение работоспособности переводчика.

II. ***Операционная система***

В целом, переводчик может работать в любой операционной системе, для которой существуют офисные приложения, браузеры, словари и специализированные программы-накопители переводов. Для ОС Windows накопителей переводов существует больше, чем для других операционных систем. Кроме того, работа в приложениях для Windows связана с соображениями совместимости.

III. ***Прикладные программы, используемые в переводческой деятельности***

Программы, которыми пользуются переводчики, можно разделить на следующие основные категории.

Программы, используемые непосредственно в процессе перевода. В данной категории речь идет об офисных, верстальных и других программах общего назначения. Первым ору-

дием переводчика был и остается по сей день текстовый процессор, а также программы для обработки электронных таблиц, создания презентаций, верстки и черчения. В прошлом решения о целесообразности освоения навыков работы в тех или иных программах часто упирались в вопрос о стоимости приобретения и обновления соответствующих программ, заказы на работу в которых поступали нерегулярно. Сейчас этот вопрос в отношении многих программ отпал, поскольку их можно загрузить с сайта производителя в виде полностью функциональной пробной версии со сроком действия от 15 до 60 дней (это относится, в частности, к программам фирмы Adobe).

Далее речь идет о накопителях перевода. Начиная с середины 90-х годов, переводчики используют в работе так называемые накопители переводов (Translation Memory, TM-tools, CAT-tools).

В основе накопителя переводов лежит специализированная база, которая содержит пары предложений и (или) терминов на двух языках. Если в переводимом документе встречаются термины или предложения, совпадающие с сохраненными в базе данных предложениями или терминами полностью или в достаточно большой степени, программа предлагает переводчику использовать существующий перевод или доработать его требуемым образом.

Целесообразность и необходимость использования накопителя переводов и выбор оптимального для конкретного

переводчика накопителя в большой степени зависят от характера переводимых текстов, стиля работы переводчика, а также от особенностей взаимоотношений переводчика с базой переводов (далее БП) и особенностей технологической цепочки БП. Следует обратить внимание на то, что практически все накопители работают с обменным форматом TMX[3], благодаря чему с технической точки зрения БП может использовать для обновления своей базы данных результаты перевода, выполненного в любом накопителе. Более того, файлы формата TMX можно сформировать из текстового документа даже в отсутствие накопителя с использованием специализированного бесплатного программного обеспечения или штатных средств программы MS Word. Другое дело, что многие бюро настаивают на применении определенного накопителя (чаще всего – Trados/SDLX[4]) из соображений технологичности, удобства или в силу привычки.

---

[3] TMX (Translation Memory Exchange – обмен памятью переводов) – открытый формат файлов XML (англ, *extensible Markup Language* – расширяемый язык разметки) для обмена данными памяти переводов, которые создаются в процессе автоматизированного перевода.

[4] Trados – профессиональный продукт для компаний, использующих переводческие процессы, основанный на выявлении в переводимом документе ранее переведенных фрагментов. Продукт был разработан в 1992 году немецкой компанией Trados GmbH. В 2005 году Trados был куплен британской компанией SDL International, и в 2006 году появился совместный продукт Trados SDLX. Trados SDLX имеет разные модули, необходимые для перевода документов различных форматов (Microsoft Word, PowerPoint, HTML, FrameMaker, InterLeaf и пр.), а также для работы с терминологическими базами данных (модуль MultiTerm).

Наряду с накопителями переводов, специалисты, занимающиеся профессиональным переводом, не могут не обращаться к электронным словарям и справочникам. Популярными словарями для перевода с русского и других языков мира являются Lingvo (www.lingvo.ru), Multitran (www.multitran.ru), Контекст (www.dics.ru).

Программы для информационной поддержки и связи (браузеры, программы для локального поиска).

Одной из существенных для переводчика характеристик браузера является его быстродействие, поскольку в процессе работы приходится просматривать большие объемы текстов «по диагонали»; его удобство работы с закладками, т. е. возможность держать «под рукой» множество открытых страниц и сохранять наборы таких страниц для справки на будущее (в связи с чем, браузеры Opera и Firefox традиционно пользуются большой популярностью среди переводчиков).

Кроме того, переводчику полезно уметь работать с FTP-сайтами[5]с использованием предназначенных именно для этой цели клиентских программ (например, Smart FTP (www.smartftp.com) или Cute FTP (www.CuteFTP.com), поскольку FTP-сайты далеко не всегда реально поддерживают заявленную совместимость с наиболее популярными браузерами.

[5] FTP (*File Transfer Protocol* – протокол передачи файлов) – стандартный протокол, предназначенный для передачи файлов по TCP-сетям (например, Интернет). Использует 21й порт. FTP часто используется для загрузки сетевых страниц и дру-

Переводчикам, которые часто пользуются беспроводной сетью WiFi в отсутствии проводного соединения для доступа в Интернет, также полезно владеть навыками диагностики и решения наиболее типичных проблем, возникающих при подключении через Wi-Fi.

Помимо выбора удобного для работы браузера, переводчик может также воспользоваться программами, обеспечивающими *локальный поиск* необходимой информации (например, Copernic Desktop Search, Yahoo Desktop Search, Google desktop, XI). Программы для поиска файлов на жестком диске по заданным ключевым словам или словосочетаниям, содержащимся в тексте перевода, позволяют специалисту сэкономить рабочее время.

Программы для обработки и подготовки исходных файлов и формирования и доработки конечных файлов.

Наиболее популярной и распространенной программой является ABBYY PDF Transformer. Для обработки графической информации – Paint, CoralDraw, Microsoft Office Document Imaging. Несмотря на то, что работа с графикой не является прямой обязанностью переводчика, тем не менее, во многих случаях возможность и умение решать простейшие задачи, связанные с графическими изображениями, полезны в двух отношениях. Они экономят время переводчику и позволяют выдавать заказчику документы в готовой для распространения форме.

Программы для административного сопровождения (учет

работ, выставление счетов, налоговая отчетность).

Наиболее распространенные программы для подсчета объема текста – это PractiCount (www.practiline.com), AnyCount (www.anycount.com). С комплексом административного ПО для переводчиков можно познакомиться на сайте **www.translation3000.com**. Программа налогового учета и отчетности – 1С: **Бухгалтерия**[6]**.**

Программы вспомогательного назначения (создание резервных копий, обеспечение надежности и безопасности, преобразование файлов в требуемые форматы).

На сегодняшний день существует достаточно большое количество программ, позволяющих создавать резервные копии и обеспечивать надежность и безопасность информации на рабочем компьютере переводчика. Наиболее популярными из них являются следующие: Nero, Driver Sweeper, Drop Box, Driver Max, DVD shrink, FBackup, Яндекс. Диск; антивирусное ПО:

Kaspersky Internet Security,
Norton Security,
Avast Internet Security,
ESET NOD32 Smart Security,
Dr.Web Security Space,
AVG Internet Security,

---

[6] 1С: Бухгалтерия – собирательное название бухгалтерских продуктов фирмы «1С», относящееся к некоторым конфигурациям на платформе 1С: Предприятие версий 7.7 и 8.

Bitdefender Internet Security.

IV. ***Автоматический перевод документов***

Автоматический перевод текстов с одного языка на другой – очень сложная задача, о полном ее решении пока говорить не приходится.

Все проблемы заключаются в объеме переводимого текста. Компьютеризованный словарь вполне может справиться с переводом отдельных слов, особенно если он способен предложить несколько значений на выбор. Однако, когда речь идет о переводе целых фраз и, тем более, абзацев связного текста, все осложняется..

Для таких случаев надежного алгоритма перевода с одного языка на другой не существует. Это связано с тем, что каждая фраза языка имеет два уровня: *синтаксический* и *смысловой.* Синтаксический уровень определяет построение предложения, а смысловой – его содержание.

Для правильного смыслового перевода необходимо принимать во внимание не только конкретную фразу, но и смысл всего абзаца или даже целой главы текста. Таким образом, рассчитывать на то, что при автоматическом переводе получится полноценный документ, нельзя.

Программы автоматического перевода рассчитаны, в первую очередь, на тех, кто совсем не знает соответствующего иностранного языка, но должен ознакомиться с содержанием документа хотя бы приблизительно. Кроме того, подобные программы позволяют готовить короткие сообщения

электронной почты на иностранном языке. Такие сообщения трудно считать грамотными, но, скорее всего, корреспондент сумеет понять, что ему хотели сообщить, поэтому программу перевода текста иностранного языка на русский, можно рассматривать как средство получения простейшего черновика.

Программные средства автоматического перевода можно условно разбить на две основные категории. Первую категорию представляют *компьютерные словари.* Назначение компьютерных словарей такое же, как и у обычных словарей: предоставить значение неизвестного слова. Преимущество их состоит в быстром доступе и удобстве автоматического поиска значения выделенного слова. Автоматический словарь обычно предоставляет возможность перевода слова по нажатию выделенной комбинации клавиш.

Ко второй категории относятся *программы-переводчики*, позволяющие выполнить автоматический перевод связного текста. Они принимают текст на одном языке и выдают текст на другом языке. В ходе работы программа использует обширные словари, наборы грамматических правил и другие средства, обеспечивающие наилучшее, с точки зрения программы, качество перевода. Чем короче предложение, тем больше шансов на то, что преобразование будет правильным.

Программы-переводчики комплектуются, кроме общих словарей, специализированными словарями по разным об-

ластям человеческой деятельности и могут переводить потоком фрагмент текста или весь текст.

В России наиболее широкое распространение получили программы автоматического перевода с английского языка на русский и с русского на английский, такие как Stylus и Socrat. Stylus обеспечивает более высокое качество и более высокую гибкость при переводе. Последние версии Stylus сменили название на Promt98, Promt 2000.

Все выше указанные программы способны работать с документами в различных форматах, допускают немедленное редактирование и оригинала и перевода; могут сохранить в нужном виде как оригинал, так и перевод. Кроме того, они включают богатый набор как универсальных, так и специализированных словарей и содержат средства для управления их использованием. Перевод слов, не входящих в словари, можно определить самостоятельно и сохранить в пользовательском словаре. Программы автоматизированного перевода позволяют указать правила работы с именами собственными и другими словами, не требующими перевода. Они имеют отдельные приложения, позволяющие осуществлять пакетный перевод файлов (File Translator), быстрый перевод неформатированного текста (Qtrans), синхронный перевод Web-страниц в Интернете (WebView).

В первой четверти нового тысячелетия характер работы переводчика и требования к нему существенно изменились. В первую очередь изменения коснулись перевода на-

учно-технической, официальной и деловой документации. Сегодня уже недостаточно просто перевести текст, пользуясь компьютером как пишущей машинкой. Заказчик ожидает о переводчика, что оформление готового документа будет соответствовать внешнему виду оригинала настолько точно, насколько это возможно, при этом удовлетворять принятым в данной стране стандартам. От переводчика требуется также умение эффективно использовать ранее выполненные заказы на ту же тему, а работодатель, в свою очередь, рассчитывает на заметную экономию времени и средств при переводе повторяющихся или похожих фрагментах текста. Эти жесткие, зачастую противоречивые условия можно соблюсти лишь в том случае, если переводчик не только в совершенстве владеет родным и иностранным языком и глубоко изучил выбранную им предметную область, но и уверенно ориентируется в современных компьютерных технологиях.

Ключевой для переводчика технической документации в данных условиях является технология **Translation Memory (TM).** Системам машинного перевода уделено мало места, так как возможности их ограничены и это не позволяет рекомендовать их для применения в процессе профессиональной работы над текстом.

В последнее время устойчиво возрастает объем переводов, связанных с информационными технологиями, причем переводческим и компьютерным компаниям приходится иметь дело не только с подготовкой документации, но и с

локализацией программного обеспечения, т. е. с переводом ресурсов, содержащихся в exe- и dll- файлах. И с последующим тестированием программного обеспечения.

Что же такое машинный перевод и системы автоматизированного перевода, чем они отличаются и как могут помочь переводчику в его работе?

Машинный перевод – процесс перевода текстов (письменных, а в идеале и устных) с одного естественного языка на другой полностью специальной компьютерной программой. Таким же образом называется направление научных исследований, связанных с построением подобных систем.

Существует несколько форм организации взаимодействия ЭВМ и человека при машинном переводе:

• с постредактированием: исходный текст перерабатывается машиной, а человек-редактор исправляет результат;

• с предредактированием: человек приспосабливает текст к обработке машиной (устраняет возможные неоднозначные прочтения, упрощает и размечает текст), после чего начинается программная обработка;

• с интерредактированием: человек вмешивается в работу системы перевода, разрешая трудные случаи.

• смешанные системы (например, одновременно с пред- и постредактированием).

Кроме того, традиционно системы машинного перевода делятся на две категории: основанные на правилах (rule-based) и основанные на примерах (example-based). В пер-

вых языковая грамматика проработана глубже, языковых правил больше. Системы второго типа – самообучающиеся, они строятся на динамическом порождении языковых правил для конкретных текстовых примеров. Границы между системами example-based и rule-based не очень четкие, поскольку и те и другие используют словари (статическая информация о языке) и правила работы со словарями. Яркий представитель класса example-based – система Trados (www.trados.com), работающая фактически на одних примерах, без грамматики. Trados предназначена для больших переводческих центров, где накопилось много параллельных текстов (два текста, один из которых является переводом другого). Она позволяет не переводить дважды одно и то же предложение, а просто находит такое же или очень похожее предложение в базе параллельных текстов и выдает уже сделанный кем-то перевод. При больших массивах однотипных текстов такой подход весьма эффективен. Вообще, понятие массива документов очень важно для машинного перевода. Большинство специалистов сходится во мнении, что машинный перевод возможен только для прикладных (технических) текстов, которые могут быть заданы определенными, порой гигантскими массивами. Художественная литература, как антипод технических текстов, никогда не будет переводиться компьютером адекватно.

Вместо термина *«машинный»* в компьютерной лингвистике иногда употребляется слово «*автоматический*», что не

влияет на смысл. Однако термин **автоматизированный перевод** имеет совсем другое значение, так как при нём программа просто *помогает человеку* переводить тексты.[7]

Автоматизированный перевод предполагает такие формы взаимодействия, как частично автоматизированный перевод (например, использование переводчиком-человеком компьютерных словарей) и систему с разделением труда, то есть компьютер обучен переводить только фразы жёстко заданной структуры (но делает это так, чтобы исправлять за ним не требовалось), а всё не уложившееся в схему отдает человеку.

Принимая во внимание тот факт, что машинный перевод различной текстовой информации становится все более и более востребованным не только в среде профессионального перевода, но в других сферах общественной жизнедеятельности, осуществим небольшой экскурс в историю создания машинного перевода в России и за рубежом.

Мысль использовать ЭВМ для перевода была высказана в 1946 году в США, сразу же после появления первых электронно-вычислительных машин. Первая публичная демонстрация машинного перевода (так называемый Джорджтаунский эксперимент) состоялась в 1954 году. Несмотря на примитивность той системы (словарь в 150 слов, грамматика из

[7] В англоязычной терминологии также различаются термины англ, machine translation, MT (полностью автоматический перевод) и англ, machine-aided или англ, machine-assisted translation (MAT) (автоматизированный); если же надо обозначить и то, и другое, пишут M(A)T.

6 правил, перевод нескольких простых фраз), этот эксперимент получил широкий резонанс: начались исследования в Великобритании, Болгарии, ГДР, Италии, Китае, Франции, ФРГ, Японии и других странах; в том же 1954 году и в СССР.

К середине 1960-х в США для практического использования были предоставлены две системы русско-английского перевода:

• MARK (в Департаменте иностранной техники ВВС США);

• GAT (разработка Джорджтаунского университета, использовалась в Национальной лаборатории атомной энергии в Окридже и в центре Евратома в г. Испра, Италия).

Однако созданная для оценки подобных систем комиссия ALPAC пришла к выводу, что в силу низкого качества машинного перевода текстов эта деятельность в условиях США нерентабельна. Хотя комиссия рекомендовала продолжать и углублять теоретические разработки, в целом её выводы привели к росту пессимизма, снижению финансирования, а иногда и к полному прекращению работ по этой тематике.

Тем не менее, в ряде стран исследования продолжались, чему способствовал постоянный прогресс вычислительной техники. Особенно существенным фактором стало появление персональных компьютеров, а с ними всё более сложных словарных, поисковых систем, ориентированных на работу с данными на естественных языках. Росла и необходимость в переводе как таковом ввиду роста международных связей.

Все это привело к новому подъёму этой области, наступившему примерно с середины 70-х годов прошлого столетия. В 1980-е наступило время широкого практического использования переводческих систем, сложился рынок коммерческих разработок по этой теме.

В настоящее время также существует множество коммерческих проектов машинного перевода. Одним из пионеров в области машинного перевода была компания 8у81хап. В России большой вклад в развитие машинного перевода внесла группа под руководством профессора Р.Г. Пиотровского (Российский государственный педагогический университет им. Герцена, г. Санкт-Петербург).

Впрочем, мечты, с которыми российские и зарубежные ученые взялись полвека назад за задачу машинного перевода, в значительной мере остаются по-прежнему мечтами, поскольку высококачественный перевод текстов широкой тематики по-прежнему недостижим. Однако несомненным является ускорение работы переводчика при использовании систем машинного перевода (по оценкам конца 1980-х) приблизительно до пяти раз.

Качество же перевода зависит от тематики и стиля исходного текста. Машинный перевод художественных текстов практически всегда оказывается неудовлетворительного качества. Тем не менее для технических документов при наличии специализированных машинных словарей и некоторой настройке системы на особенности того или иного типа тек-

стов возможно получение качественного перевода, который нуждается лишь в небольшой редакторской корректировке. Чем более формализован стиль исходного документа, тем большего качества перевода можно ожидать. Самых лучших результатов при использовании машинного перевода можно достичь для текстов, написанных в техническом (различные описания и руководства) и официально-деловом стиле.

Применение машинного перевода без настройки на тематику (или с намеренно неверной настройкой) служит предметом многочисленных шуток, особенно в профессиональной среде переводчиков. Например, программа ПРОМТ переводит предложение *«My cat has given birth to four kittens, two yellow; one white and one black»* на русский язык следующим образом: *«Мой кот родил четырёх котят, два жёлтых цвета, одно белое и одного афроамериканца»*.

Профессиональная работа невозможна без надежных инструментов. Перевод и локализация[8] как область професси-

[8] Локализация часто рассматривается как «перевод высокого уровня», но это не отражает всю важность и сложность этого процесса, а также все то, что она в себя включает. Хотя, иногда сложно провести границу между переводом и локализацией, в общем, локализация проводится в значительной степени для внетекстуальных компонентов товара или услуги. В дополнение к переводу (то есть, вопросам грамматики и орфографии, которые варьируются в зависимости от страны и места, где используется один и тот же язык), процесс локализации может включать адаптацию графического компонента, символов валют, формата дат, адресов и номеров телефона, выбор цветов многих других деталей, включая пересмотр физической структуры продукта.Все эти изменения проводятся с целью, во-первых, выявить чувствительные различия и избежать возможных конфликтов с местной культурой и населением и, во-вторых, проникнуть на местный ры-

ональной активности в этом смысле не являются исключением. Любой переводчик сталкивается с проблемой согласованного применения терминологического глоссария в ходе длительного проекта или быстрого повторного использования ранее переведенного текста. По своей природе подобные рутинные задачи сравнительно легко (в отличие от машинного перевода) формализуются и программируются, поэтому оснащение рабочего места (локализатора) автоматизированными средствами является нормой в отрасли, постепенно трансформируясь в отраслевые стандарты.

Большинство таких средств построены на основе концепции *памяти перевода* (translation memory) – простой базы данных, каждая запись которой представляет собой единицу (предложение или абзац) параллельных текстов (как правило, на двух языках). Такая база данных хранит предыдущие переводы с целью их возможного повторного использования и решения задач быстрого поиска по содержимому. Несмотря на то, что программы, оснащенные памятью перевода, называются системами автоматизированного перевода (CAT, или computer-aided/ assisted translation), их не следует путать с программами машинного перевода (machine translation) – память перевода ничего не переводит сама по себе, в то время как машинный перевод основан на генерации переводов

нок, приспосабливаясь к локальным нуждам. Например, в результате локализации веб-сайт одной и той же компании может быть адаптирован к определенной стране, или издания одной и той же книги могут различаться в зависимости от места издания.

по результатам грамматического разбора исходного текста.

Как правило, запись памяти перевода состоит из двух сегментов: на исходном (source) и конечном (target) языках. Если идентичный (или похожий) сегмент на исходном языке встречается в тексте, сегмент на конечном языке будет найден в памяти перевода и предложен переводчику в качестве основы для нового перевода. Автоматически найденный текст может быть задействован как есть, отредактирован или полностью отклонен. Большинство программ используют алгоритм нечеткого соответствия (fuzzy matching), существенно улучшающий их функциональные возможности, поскольку в этом случае можно находить предложения, лишь отдаленно напоминающие искомые фразы, но тем не менее пригодные для последующего редактирования.

Преимущества от использования такого программного обеспечения поначалу могут быть неочевидны – однако по мере наполнения базы данных результаты автоматической подстановки основ для перевода будут становиться все более точными и регулярными.

Архитектура автоматизированной системы и ее функциональные возможности могут различаться. Средства поиска могут работать как с целыми сегментами, так и с отдельными словами или фразами, позволяя переводчику выполнять терминологический поиск. В систему также включают отдельную программу для работы с глоссарием, содержащим утвержденные для применения в проекте термины. Некото-

рые системы работают с программами машинного перевода. Основной рабочий интерфейс либо встраивается непосредственно в имеющийся текстовый процессор, такой как Word, либо представляет собой отдельный редактор. В состав системы обязательно включают фильтры для импорта-экспорта файлов различных форматов. Кроме того, многие системы, если не все, имеют средство для добавления в память перевода сегментов из, как правило, имеющихся у переводчика старых переведенных файлов.

Автоматизированный перевод (*Computer-Aided Translation)* – это перевод текстов на компьютере с использованием компьютерных технологий. В отличие от машинного перевода в данном случае человек осуществляет весь процесс перевода, а компьютер всего лишь помогает ему произвести готовый текст либо за меньшее время, либо с лучшим качеством.

Идея автоматизированного перевода появилась с момента появления компьютеров. Переводчики всегда выступали против стандартной в те годы концепции машинного перевода, на которую было направлено большинство исследований в области компьютерной лингвистики, но поддерживали использование компьютеров для помощи переводчикам. В 1960-е годы Европейское объединение угля и стали (предшественник современного Евросоюза) стало создавать терминологические базы данных под общим названием *Eurodicautom.* В Советском Союзе для создания баз такого

рода был создан ВИНИТИ (Всероссийский институт научной и технической информации).

В современной форме идея автоматизированного перевода была развита в 1980 году в статье Мартина Кея[9], который выдвинул следующий тезис: «by taking over what is mechanical and routine, it (computer) frees human beings for what is essentially human» (компьютер берет на себя рутинные операции и освобождает человека для операций, требующих человеческого мышления).

В настоящее время наиболее распространенными способами использования компьютеров при письменном переводе является работа со словарями и глоссариями, с системой памяти переводов (*translation memory*), содержащей примеры ранее переведенных текстов, а также использование так называемых корпусов, больших коллекций текстов на одном или нескольких языках, что дает сжатое описание того, как слова и выражения реально используются в языке в целом или в конкретной предметной области.

Для локализации программного обеспечения часто применяются специализированные средства, например, SDL Passolo 2015, которые позволяют переводить меню и сообщения в программных ресурсах и непосредственно в откомпилированных программах, а также тестировать коррект-

---

[9] Martin Kay (1980). The Proper Place of Men and Machines in Language Translation. Research report CSL-80-11, Xerox Palo Alto Research Center, Palo Alto, CA. Перепечатано в 1997 году в *Machine Translation* 12: 3-23, 1997.

ность локализации. Для перевода аудиовизуальных материалов (главным образом фильмов) также используются специализированные средства, например, Aegisub[10], которые объединяют в себе некоторые аспекты памяти переводов, но дополнительно обеспечивают возможность появления субтитров по времени, их форматирования на экране, следования видеостандартам и т. п.

При синхронном переводе использование средств автоматизированного перевода по необходимости ограничено. Одним из примеров является использование словарей, загружаемых на карманный персональный компьютер (КПК). Другим примеров может служить полуавтоматическое извлечение списков терминов при подготовке к синхронному переводу в узкой предметной области.

В узких предметных областях при большом количестве исходных текстов и устоявшейся терминологии переводчики могут использовать и машинный перевод, который может обеспечить хорошее качество перевода терминологии и устойчивых выражений в узкой области. Переводчик в этом случае осуществляет пост-редактирование полученного тек-

---

[10] Aegisub – кроссплатформенный редактор субтитров с открытым исходным кодом. Имеет расширенные возможности по созданию караоке. Включает в себя проверку орфографии и редактор переводов. Поддерживает в качестве субтитров SRT, ASS, SSA, SUB, XSS, PSB и форматированный TXT. Для тайминга в неё можно загрузить аудиофайлы в форматах WAV, MP3, OGG Vorbis, FLAC, MP4, AC3, AAC и MKA, видеофайлы – в форматах AVI, AVS, D2V, MKV, OGM, MP4, MPEG, MPG и VOB. Имеет возможность работы с анаморфорным видео.

ста. Более половины текстов внутри Еврокомиссии (главным образом юридические тесты и текущая корреспонденция) переводится с использованием машинного перевода.

Память переводов, или накопитель переводов (*translation memory)* – это база данных, содержащая набор ранее переведенных текстов. Одна запись в такой базе данных соответствует «единице перевода» (*translation unit),* за которую обычно принимается одно предложение (реже – часть сложносочинённого предложения, либо абзац). Если очередное предложение исходного текста в точности совпадает с предложением, хранящимся в базе (точное соответствие, или *exact match),* оно может быть автоматически подставлено в перевод. Новое предложение может также слегка отличаться от хранящегося в базе (неточное соответствие, *fuzzy match).* Такое предложение может быть также подставлено в перевод, но переводчик будет должен внести необходимые изменения.

Помимо ускорения процесса перевода повторяющихся фрагментов и изменений, внесенных в уже переведенные тексты (например, новых версий программных продуктов или изменений в законодательстве), системы памяти переводов также обеспечивают единообразие перевода терминологии в одинаковых фрагментах, что особенно важно при техническом переводе. С другой стороны, если переводчик регулярно подставляет в свой перевод точные соответствия, извлеченные из баз переводов, без контроля их использова-

ния в новом контексте, качество переведенного текста может ухудшиться.

В каждой конкретной системе памяти переводов данные хранятся в своем собственном формате (текстовый формат в Wordfast, база данных Access в Deja Vu), но существует международный стандарт TMX *(Translation Memory eXchange format),* который основан на XML () и который могут порождать практически все системы памяти переводов. Благодаря этому результаты работы переводчиков можно обменивать между приложениями, то есть переводчик, работающий с программой OmegaT может использовать память переводов, созданную в Irados и наоборот.

Большинство систем памяти переводов как минимум поддерживают создание и использование словарей пользователя, создание новых баз данных на основе параллельных текстов[11] (*alignment*), а также полуавтоматическое извлечение терминологии из оригинальных и параллельных текстов.

На сегодняшний день наиболее популярными программными системами автоматизированного и машинного перевода с использованием памяти переводов являются: Яндекс. Перевод, Deja Vu, OmegaT, SDLX, Trados), STAR Transit NXT, Wordfast (реализована как набор макросов для MS Word), ABBYY Lingvo, Apertium, Ectaco, Google Translate,

[11] **Параллельный текст** – методика обучения иностранному языку путем чтения текста на изучаемом языке с параллельным переводом на родной язык. Чтобы пользоваться этим методом, необходимо только заранее знать правила чтения изучаемого языка.

PROMT, Across.

Основными достоинствами выше перечисленных программ являются следующие.

1. Высокая скорость. В течение нескольких секунд получается перевод многостраничного текста. Это позволяет быстро понять смысл текста, а если система настроена на перевод текстов определенной тематики, требуется минимальная редакторская правка.

2. Низкая стоимость. При обращении к профессиональным переводчикам приходиться платить за каждую страницу переведенного текста, либо нанимать штатного переводчика, которому приходится платить зарплату. В случае с системой автоматизированного перевода платить деньги необходимо только один раз – при покупке программы, что впоследствии окупается в несколько раз.

3. Доступ к услуге. Немаловажный фактор, который многие критики систем автоматизированного перевода не принимают в расчет. Программа-переводчик всегда под рукой, а обращаться в переводческое бюро во многих случаях связано с дополнительными затратами времени и сил.

4. Конфиденциальность. Системе машинного перевода можно доверить любую (даже конфиденциальную) информацию. Программа-переводчик способна хранить в тайне любые тексты, которой ей доверяет переводчик.

5. Универсальность. Любой переводчик всегда имеет специализацию, т. е. переводит тексты по той теме, которой

он хорошо владеет. Когда переводчик художественной литературы берется за перевод, например, технических текстов, ошибок в переводе не избежать. Система автоматизированного перевода выгодно отличается тем, что она абсолютно универсальна. Нужно только грамотно подключить специализированный словарь по соответствующей тематике. Следует учесть и еще одно преимущество подобных систем: пополнение их специализированных словарей новейшими терминами значительно опережает аналогичные словари полиграфического исполнения. В ряде случаев также рекомендуется вести свой собственный словарь новых терминов или новых значений. В этом случае переводчик гарантированно получает необходимое качество перевода.

6. Перевод информации в Интернете. В виртуальном пространстве глобальной сети наиболее ярко проявляются все преимущества систем машинного перевода. Более того, в большинстве случаев переводить информацию в Интернете, если человек сам не знает нескольких языков, можно только с помощью программ-переводчиков. Именно эта потребность обусловила огромный рост интереса к системам машинного перевода сейчас в мире. Только благодаря онлайновым системам перевода появилась возможность просматривать иностранные сайты, не затрудняясь с их переводом – быстро, удобно и конфиденциально.

Коллективное использование систем машинного перевода в организациях имеет дополнительные преимущества.

1. Единообразие стиля и используемой терминологии. Как известно, затраты на постредактирование при работе коллектива переводчиков составляют около 100–140 % от стоимости перевода. Перевод, выданный системой машинного перевода, гораздо легче править, поскольку он выдержан в одном стиле. Если в тексте, какой-либо часто встречающийся термин переведен неправильно, то все эти ошибки можно исправить простой функцией *автозамены.* Когда объемный текст переводится группой переводчиков, то приходится выявлять отдельные неточности, допущенные каждым переводчиком. Редактору в этом случае требуется также «выравнивать» и стиль перевода.

2. Отсутствие затрат на форматирование. Это особенно важно при переводе электронной документации. Программа-переводчик полностью сохраняет исходное форматирование, что позволяет сэкономить время и деньги при подготовке перевода.

Вопросы для самоконтроля

1. Что такое *машинный перевод*?

2. Что такое *автоматизированный перевод*?

3. В чем отличие *автоматизированного* от *автоматического* перевода?

4. Дать определение *памяти перевода* (Translation Memory).

5. Каковы программные характеристики персонального

компьютера, необходимые для оптимального перевода?

6. Перечислить современные программы машинного и автоматизированного перевода и их базовые характеристики.

7. Каковы ключевые достоинства современных программ машинного перевода.

Задания для самостоятельного исследования

1. Подготовить доклад по перспективам развития информационного общества XXI тысячелетия.

2. Подготовить доклад по Интернет-лингвистике. Оценить и спрогнозировать влияние Интернет-лингвистики на общее состояние современной и будущей лингвистической науки.

3. Подготовить доклад по модели представления современного переводческого процесса «автор – переводчик (человек) – переводчик (компьютерная программа) – читатель».

4. Перевести на английский язык отрывок из произведения А.П. Чехова **«За двумя зайцами погонишься, ни одного не поймаешь»**[12], используя программы машинного и автоматизированного перевода текста.

5. Перевести с английского языка на русский язык отрывок из произведения **O.** Henry **«Aristocracy Versus Hash»,** [13]используя программы машинного и автоматизированного

[12] Полный текст рассказа А.П. Чехова можно прочитать в Приложении 3.

[13] Полный текст рассказа O. Henry можно прочитать в Приложении 4.

перевода текста.

Рекомендованная литература

1. *Анисимов, Д.В.* Правда о машинном переводе / Д.В. Анисимов. – Москва: Сам Полиграфист, 2014. – 340 с.

2. *Коммисаров, В.Н.* Теория перевода (лингвистические аспекты): Учеб, для ин-тов и фак. иностр. яз. / В. Н. Комиссаров. – Репр. изд. – Москва: Альянс, 2013. – 250 с.

3. *Кутузов, А.Б.* Компьютерные технологии в формировании профессиональной компетенции переводчика/А.Б. Кутузов // Языки профессиональной коммуникации: сборник статей Третьей международной научной конференции, т. 2. – Челябинск, 2007. [Электронный ресурс] – URL: http://tc.utmn.ru/files/kutuzov_it.pdf (дата обращения: 24.09.2015).

4. *Орёл М.А.* Словарь переводчику – друг, товарищ и Брут / М.А. Орел// Перевод: информационные технологии. – М.: Всероссийский центр переводов науч. – техн. лит. и документации, 2009. – С. 79–106.

5. *Попов, С.А.* Информационные технологии в лингвистике: учебное пособие / С.А. Попов, Е. Ф. Жукова; М-во образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высшего проф. образования «Новгородский гос. ун-т им. Ярослава Мудрого» Великий Новгород: Новгородский гос. ун-т,2014.-235 с.

6. *Соловьёва А.В.* Профессиональный перевод с помощью компьютера. – СПб. Литер, 2008. – 160 с.

7. *Bowker, L.* Computer-Aided Translation Technology: A Practical

Introduction. – University of Ottawa Press, 2002. – 185 p. Retrieved from http://books.google.com/books7id4y29-mc6dOOC.

8. Encyclopaedia of Translation Studies / Ed. M. Baker. – London: Routledge, 2004. – 654p.

9. *Kenny, D.* Teaching Machine Translation and Translation Technology: a Contrastive study. Retrieved from URL: http://doras.dcu.ie/15830/1/ Teaching_Machine_Translation_%26_Translation_Technology.pdf.

10. An Introduction to CAT Tools (Translation Memory). – Keypot corporation. Retrieved from URL: http://www.horsefrog.com/japanese-translator-patent/ mod/resource/view.php?id= 108.

Глава 3
ПОИСК И ПУБЛИКАЦИЯ ИНФОРМАЦИИ В ИНТЕРНЕТЕ

Краткое описание: информационный поиск, поисковая машина, поисковые системы, правила эффективного поиска информации.

В жизни, как правило, преуспевает больше других тот, кто располагает большей информацией.
Бенджамин Дизраэли

In this electronic age we see ourselves being translated more and more into the form of information, moving toward

the technological extension of consciousness.
Marshall McLuhan

В Интернете с каждым днём скапливается всё больше информации, когда-либо созданной и вновь создаваемой людьми. Равнодоступность большей части информации в Интернете уравнивает возможности доступа к этой информации как обычных пользователей Интернета и журналистов локальных СМИ, так и сотрудников мировых информационных агентств.

Благодаря Интернету перед каждым человеком ежедневно и даже ежесекундно открывается доступ к многомиллионной аудитории, которой он может передать свой информационный материал, полученный, например, с помощью обычного мобильного телефона с диктофоном и встроенной фотокамерой. Следовательно, уровень монополизации деятельности по распространению информации также снижается благодаря Интернету.

До недавнего времени ограничения в прямой коммуникации между людьми, порождаемые пространством и временем, во многом определяли потребность людей в услугах журналистов. По мере роста общего количества пользователей Интернета, а среди них – числа владеющих английским языком, эти ограничения всё в большей степени снимаются, что закономерно ведёт к уменьшению спроса на услуги журналистов. Одновременно с этим растёт объем «сырой» информации, доступной каждому отдельному пользователю

Интернета, что актуализирует проблему её отбора и редактирования. Последнее всегда входило в перечень функций журналистики, но с ростом числа пользователей Интернета эффективный информационный поиск начинает приобретать всё большую значимость не только в журналистской деятельности, но и в других разнообразных сферах общественной деятельности.

Таким образом, **информационный поиск** – это процесс поиска неструктурированной документальной информации.

Поиск информации представляет собой процесс выявления в некотором множестве документов (текстов), которые посвящены заданной теме (предмету) и удовлетворяют заранее определенному условию поиска (запросу), а также содержат необходимые (соответствующие информационной потребности) факты, сведения и данные.

Процесс поиска включает последовательность операций, направленных на сбор, обработку и предоставление необходимой информации заинтересованным лицам.

Комплекс программ, предназначенных для информационного поиска, называется **поисковой машиной.** Обычно является частью **поисковой системы** – автоматизированного программно-аппаратного комплекса с веб-интерфейсом, предоставляющего возможность поиска информации в Интернете. Самая известная поисковая система в мире – это Google, самая популярная в России – Яндекс, а одной из самых старых поисковых систем является Yahoo. Как уже бы-

ло отмечено ранее, в архитектуре поисковой системы можно выделить *поисковую машину* – ядро системы, представленное набором программных модулей; базу данных или *индекс*, хранящую информацию обо всех известных поисковой системе Интернет ресурсах; и набор сайтов, являющих собой точки входа пользователей в систему (www.google.com, www.yandex.ru, ru.yahoo.com, и т. д.). Все это соответствует классической трехуровневой архитектуре информационных систем: есть пользовательский интерфейс, бизнес логика, которая в данном случае представлена реализацией алгоритмов поиска и база данных.

Для того, чтобы найти в Интернете требуемую информацию, необходимо знать либо адрес её местоположения (например, адрес *ЬІm*\-страницы или файла), либо пользователя Интернета, который может предоставить информацию. Если мы не знаем ни адреса, ни человека, который мог бы нам помочь, то следует перейти к вопросам *«Как можно узнать адрес размещения информации?»* или *«Как найти человека, который мог бы нам помочь с поиском информации?»*. При этом не следует переоценивать возможности Интернета. Лучшие результаты может дать совмещение онлайновых и оффлайновых методов поиска информации.

Сегодня существует достаточно большое количество методов информационного поиска в Интернете и через Интернет. В каждом конкретном случае успешность поиска определяется знаниями возможных методов и навыками владе-

ния ими, знанием этнических языков, на которых эта информация может быть представлена, либо нашими социальными связями.

Выделяется 4-е этапа поиска информации.

1. определение (уточнение) информационной потребности и формулировка информационного запроса;

2. определение совокупности возможных держателей информационных массивов (источников);

3. извлечение информации из выявленных информационных массивов;

4. ознакомление с полученной информацией и оценка результатов поиска.

Поисковые запросы бывают *явные* и *неявные*. В явных вопросах конкретно указывается объект поиска. В неявных вопросах, например, *«какая сегодня погода»*, *«происходит ли сейчас что-то важное»*, *«можно ли проехать по городу»*, или, как у А. С. Пушкина в «Сказке о мертвой царевне и о семи богатырях»:

Свет мой, зеркальце! Скажи
Да всю правду доложи:
Я ль на свете всех милее,
Всех румяней и белее?

объект поиска конкретно не указывается.

Поисковые запросы также делятся в зависимости от требуемой системы поиска. *Первая группа* поисковых систем

предназначена лишь для линейного поиска информации, то есть для обнаружения в текстах фрагментов, аналогичных заданному. Следовательно, в запросе должен содержаться фрагмент текста. *Вторая группа* систем позволяет выбирать данные о связях между объектами, что требует указания в запросе на связь между теми или иными объектами.

Чтобы спланировать поиск, следует, прежде всего, определить объект поиска, сформулировать какую информацию необходимо найти. Если однозначно ответить на этот вопрос не представляется возможным, то поиск следует разделить на задачи с разными объектами. В планировании поиска также следует определить соотношение видов информации в поисковой задаче.

Например, если необходимо представить какую-то компанию, то полезными могут стать не только стандартные характеристики фирмы (данные об обороте, клиентах и пр.), но и сведения о связях ее первых лиц. И наоборот – физическое лицо можно охарактеризовать через компанию, которой оно владеет или в которой работает.

Надо определиться и с возможными форматами, файлов в которых может содержаться требуемая информация. Это может быть html-страница, текстовый документ в форматах txt, rtf, odt, doc или docx, документ pdf, презентация в форматах odp, ppt или pptx, электронная таблица в форматах ods, xls или xlsx, аудио в формате mp3, flash-ролик формата swf, видео в формате avi и т. д.

Важно отметить, что на первый взгляд поиск в интернете мало чем отличается от обычного информационного поиска, например, от обработки *SQL запроса*[14] к базе данных или от задачи поиска файла на компьютере. Так считали и разработчики первых поисковых систем в интернете, но со временем они осознали, что заблуждались.

Первое отличие поиска в Интернете от обычного состоит в том, что алгоритм поиска по той же базе данных предполагает, что ее структура заранее известна поисковой машине и автору запроса. В интернете, по понятным причинам, это не так. Интернет страницы образуют собой не структуру каталога, а сеть, что также влияет на алгоритмы поиска, а формат данных, размещаемых на интернет ресурсах, никем не контролируется.

Второе отличие, как одно из следствий первого – это то, что запрос представляется не в виде набора значений параметров (критериев поиска), а в виде текста, написанного человеком на естественном для него языке. Таким образом, перед тем, как начать поиск нужно еще понять, чего именно хочет автор запроса. Замечу, понять не другому человеку, а вычислительной машине.

Третье отличие уже менее очевидное, но не менее принципиальное: в каталоге или базе данных все элементы равноправны. В интернете имеет место конкуренция, а, следовательно, и разделение на более «благонадежных поставщиков

[14] SQL запрос – структурированные запросы к базе данных.

информации» и источников, близких по статусу к «информационному мусору». Так классифицируют ресурсы люди, и также к ним относятся поисковые машины.

И в заключении следует добавить, что область поиска – это миллиарды страниц, по несколько килобайт и более каждая. Около десятка миллионов страниц добавляется ежедневно и столько же обновляется. Все это представлено различными цифровыми форматами. К сожалению, даже современные технологии и ресурсы, имеющиеся в распоряжении лидеров рынка поисковых услуг в Интернете не позволяют им обрабатывать все это многообразие «на лету» и в полной объеме.

Принципиально важным моментом функционирования поисковой машины в Интернете является является тот факт, что поиск и отбор информации осуществляется на базе формирования запросов в собственное информационное хранилище – баз данных, называемых индексами, где хранятся досье на все известные машине сайты. Эти базы данных периодически обновляются.

Иными словами, поисковая машина работает не с оригиналом, а с проекцией области допустимых значений поиска. Поэтому последние изменения в Интернете могут отразиться в результатах поиска только после того, как соответствующие страницы будут проиндексированы – добавлены в индекс поисковой системы. Таким образом, поисковая система Интернета в первом приближении состоит из поисковой ма-

шины, базы данных или индекса (index) и точек входа в эту систему. Сама поисковая машина также является совокупностью приложений, позволяющих делать работу эффективно и быстро.

Перечислим компоненты поисковой машины.

1. *Паук или спайдер (spider).* Приложение, которое занимается скачиванием страниц Интернет-ресурсов. «Паук» запрашивает содержимое страниц точно так же, как это делает обычный интернет браузер, отправляя на сервер HTTP запрос и получая от него ответ. После того, как содержимое страницы скачано, оно отправляется индексатору и краулеру, о которых рассказывается далее.

2. *Индексатор (indexer).* Индексатор производит первоначальный анализ содержимого скачанной страницы, выделяет основные части (название страницы, описание, ссылки, заголовки и т. д.) и раскладывает все это по разделам поисковой базы данных – помещает в индекс поисковой системы. Этот процесс называют индексацией интернет ресурсов, отсюда и название самой подсистемы. На основе результатов первоначального анализа индексатор также может принять решение, что страница вообще «недостойна» находиться в индексе. Причины такого решение могут быть разными: страница не имеет названия, является точной копией другой, уже имеющейся в индексе страницы или содержит ссылки на запрещенные законодательством ресурсы.

3. Краулер (crawler). Это приложение призвано переме-

щаться по ссылкам, имеющимся на скачанной пауком странице. Краулер анализирует пути, ведущие с текущей страницы на другие разделы сайта, или на страницы внешних Интернет ресурсов и определяет дальнейший порядок обхода пауком нитей всемирной паутины. Именно краулер находит новые для поисковой машины страницы и передает их пауку. Работа краулера построена на базе алгоритмов поиска на графах в ширину и глубину.

4. Подсистема обработки и выдачи результатов (Search Engine and Results Engine). Самая важная часть любой поисковой машины. Алгоритмы работы этой подсистемы компании разработчики хранят в строгой секретности, поскольку они являют собой коммерческую тайну. Именно эта часть поисковой машины отвечает за адекватность ответа поисковой системы на запрос пользователя. Здесь можно выделить два основных компонента:

• *Подсистема ранжирования.* Ранжирование – это сортировка страниц интернет сайтов в соответствии с их релевантностью определенному запросу. Релевантность страницы – это, в свою очередь, степень соответствия содержания страницы смыслу запроса, и эту величину поисковая машина определяет самостоятельно, исходя из огромного количества параметров. Ранжирование – эта самая загадочная и спорная часть «искусственного интеллекта» поисковой машины. На ранжирование страницы, помимо ее структуры и содержимого (контента) также влияют: количество и каче-

ство ссылок, ведущих на данную страницу с других сайтов; возраст домена самого сайта; характер поведения пользователей, просматривающих страницу и многие другие факторы.

• *Подсистема выдачи результатов.* В задачи этой подсистемы входит интерпретация пользовательского запроса, его перевод на язык структурированных запросов к индексу и формирование страниц результатов поиска.

Помимо разбора самого текста запроса, поисковая машина может также учитывать *контекст запроса*, формируемый исходя из смысла ранее осуществленных пользователем запросов. Так, например, если пользователь часто посещает сайты на автомобильные темы, то на запрос со словом «Волга» или «Ока» он, вероятно, хочет получить информацию об автомобилях этих марок, а не о том, откуда начинают свое течение и куда впадают одноименные русские реки. Это называется персонализированным поиском, когда выдача на один и тот же запрос для разных пользователей существенно отличается. Таким образом, речь идет о пользовательских предпочтениях, о которых поисковая машина может «догадываться», анализируя выбираемые пользователем ссылки на страницах результатов поиска. Это еще один способ скорректировать контекст запроса: пользователь своими действиями как бы подсказывает машине, что именно он хочет найти. Как правило, поисковые машины в результаты поиска стараются добавлять страницы, релевантные запросу,

но относящиеся к довольно разным сферам жизни. Еще один важный момент, который учитывает поисковая машина, – это *регион проживания пользователя*, особенно при обработке коммерческих запросов, связанных с приобретением товаров и услуг у местных поставщиков. Если человек интересуется распродажами и скидками в торговых центрах города Москвы, в котором он проживает, то скорее всего, ему не интересно, какие акции на эту тему проводятся в Санкт-Петербурге, если только пользователь не указал этот город в тексте запроса. В первую очередь в результатах поиска должна появиться информация о распродажах в Москве. Следовательно, современные поисковые машины делят запросы на *геозависимые* и *геонезависимые.* Если поисковая система решает, что запрос Интернет пользователя геозависимый, то она автоматически добавляет к нему признак региона, который пытается определить по информации об Интернет провайдере данного пользователя.

Поисковым машинам иногда приходится анализировать, когда имели место события, описываемые на странице. Ведь информация постоянно устаревает, а пользователю нужны в первую очередь ссылки на самые последние новости, актуальные прогнозы и анонсы событий, которые еще не завершились или должны наступить в будущем. Понять, что актуальность страницы зависит от времени, и сопоставить ее с моментом выполнения запроса также требует от поисковой машины изрядной доли интеллекта.

Далее, поисковая машина ищет ближайший по смыслу ключевой запрос в индексе и формирует результаты, сортируя ссылки в порядке убывания их релевантности. Каждому ключевому запросу в индексе соответствует отдельный рейтинг страниц, релевантных ему. Не на каждое сочетание букв и цифр система заводит новый ключевой запрос, а делает это на основе анализа частоты тех или иных пользовательских запросов.

Поисковая машина может также перемешивать в результатах поиска рейтинги из разных ключевых запросов, если посчитает, что пользователю нужно именно это. Разработчики поисковых систем затрачивают большие усилия, направленные на то, чтобы «очистить» результаты своей поисковой выдачи от разного рода информационного мусора, то *есть спама* (spam).

Поисковая машина при поддержке входящих в нее приложений (пауков и краулеров) постоянно сканирует Интернет на предмет появления новых и обновления существующих страниц, поскольку неактуальная информация ценится ниже.

Поисковая машина периодически обновляет ранжирование ресурсов по их релевантности ключевым запросам, поскольку в индексе постоянно появляются новые страницы. Этот процесс называют обновлением (updating) поисковой выдачи.

В силу огромных объемов информации, размещенной во

всемирной паутине и ограниченности ресурсов самой поисковой системы, поисковая машина всегда старается загружать только самое (по ее мнению) необходимое. В ее арсенале имеются всевозможные фильтры, которые отсекают многое ненужное уже на этапе индексации или выкидывают спам из индекса по результатам обновления поисковой выдачи.

Современные поисковые системы в ходе анализа запроса стараются учитывать не только текст самого запроса, но и его окружение: контекст и предпочтения пользователя, о которых было сказано ранее, а также время запроса, регион и многое другое.

На релевантность конкретной страницы влияют не только внутренние ее параметры (структура, содержание), но и внешние параметры, такие как ссылки на страницу с других сайтов и поведение пользователя при ее просмотре.

Работа поисковых систем постоянно совершенствуется. Идеальная работа поисковой машины (для человека) возможна только в том случае, если все решения, касающиеся индексации и ранжирования будет принимать комиссия, состоящая из большого числа специалистов всех областей и направлений человеческой деятельности. Поскольку это нереально, то такую комиссию заменяют экспертные системы, эвристические алгоритмы поиска и прочие элементы искусственного интеллекта. Вероятно, работа всех этих подсистем также могла бы давать более адекватные результаты, если бы была возможность обрабатывать абсолютно все данные, име-

ющиеся в открытом доступе в интернете, но и это практически невозможно. Несовершенный искусственный интеллект и ограниченность ресурсов – две основные причины того, что результаты поисковой выдачи не всегда радуют пользователей, но все это корректируется временем. Сегодня работа наиболее известных и крупных поисковых систем таких, как **Google, Yahoo, Bing, Baidu, Яндекс, Рамблер, Nigma** вполне соответствует потребностям и ожиданиям их пользователей.

Вопросы для самоконтроля

1. Что такое *информационный поиск*? Дать определение *поисковой системе* и *поисковой машине.*

2. Перечислить основные этапы информационного поиска. Дать краткую характеристику каждому этапу.

3. В чем заключаются принципиальные отличия поиска в Интернете от обычного информационного поиска?

4. Перечислить основные приложения, входящие в поисковую машину. Охарактеризовать каждое из приложений со своими примерами.

5. Какие факторы учитывает поисковая машина при отборе запрашиваемой пользователем информации?

6. Каковы наиболее популярные и эффективные поисковые системы в современном мире?

Задания для самостоятельного исследования

1. Используя три (по выбору) поисковые системы найти список самых популярных книг первой четверти нынешнего столетия. Сравнить результаты поиска.

2. Подготовить доклад по одной из поисковых систем современности.

3. Поиск по рубрикатору поисковой системы

Поисковые каталоги представляют собой систематизированную коллекцию (подборку) ссылок на другие ресурсы Интернета. Ссылки организованы в виде тематического рубрикатора, представляющего собой иерархическую структуру, перемещаясь по которой, можно найти нужную информацию. Например:

- Бизнес и экономика;
- Общество и политика;
- Наука и образование;
- Компьютеры и связь;
- Справочники и ссылки;
- Дом и семья;
- Развлечения и отдых;
- Культура и искусство.

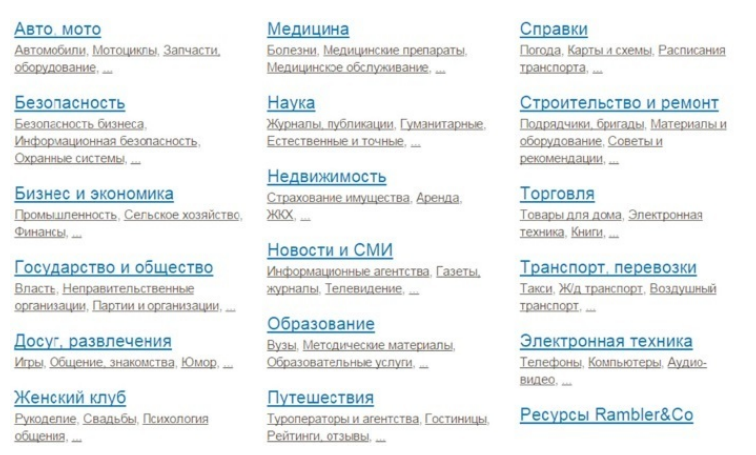

Рис. 1. Тематический рубрикат Рамблер. Топ 100 медийного Интернет-портала Рамблер

Поработайте с поисковыми каталогами российских и зарубежных медийных Интернет-порталов по интересующей тематике (используя возможности поиска по ключевым словам; расширенного поиска). Выявите сходства и различия в функционировании поисковых каталогов российских и зарубежных поисковых систем.

4. Пользуясь каталогом поисковой системы, найдите следующую информацию:

- Текст песни популярной музыкальной группы;
- Репертуар Мариинского театра на текущую неделю;
- Характеристики последней модели мобильного телефо-

на известной фирмы (по вашему выбору);

• Рецепт приготовления украинского борща с галушками;

• Долгосрочный прогноз погоды в Вашем регионе (не менее чем на 10 дней);

• Фотография любимого исполнителя современной песни;

• Примерная стоимость мультимедийного компьютера (прайс);

• Информация о вакансиях на должность преподавателя иностранных языков в Вашем регионе или городе;

• Гороскоп своего знака зодиака на текущий день;

По результатам поиска составьте письменный отчет в Word: представьте в документе найденный, скопированный и отформатированный материал.

Рекомендованная литература

1. *Ашманов, И. С.* Оптимизация и продвижение сайтов в поисковых системах / И.С. Ашманов, А.А. Иванов. – 3-е изд. – Москва: Питер, 2015 -463 с.

2. *Байков В.Д.* Интернет. Поиск информации. Продвижение сайтов /В.Д. Байков. – СПб.: БХВ-Петербург, 2000. – 288 с.

3. *Колисниченко Д.Н.* Поисковые системы и продвижение сайтов в Интернете /Д.Н. Колисниченко. – М.: Диалектика, 2007. – 272 с.

4. Основы информационной компетентности [Электронный ресурс]: учебное пособие: электронное издание/ М-во

образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высш. проф. образования Красноярский гос. пед. ун-т им. В.П. Астафьева; сост.: Н.В. Васильева Красноярск: КГПУ им. В. П. Астафьева, 2013-1 электрон, опт. диск (CD-ROM)hh.;12 см – Загл. с этикетки диска.

5. *Chu H.* Search engines for the World Wide Web: A comparative study and evaluation methodology (англ.) / H. Chu, M. Rosenthal // Proceedings of the Annual meeting – American society for information science: journal. -1996. – Vol. 33. – P. 127–135.

6. *Eric W. B.* Execution Performance Issues in Full-Text Information Retrieval. – University of Massachusetts Amherst: Computer Science Department, 1996. – 179 p. – (Technical Report 95–81).

7. *Pariser E.* The Filter Bubble: What The Internet Is Hiding From You. – NY: Penguin Group, 2011. – 257 p.

Глава 4
ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫИ ЛИНГВИСТИКА

Краткое описание: искусственный интеллект, компьютерная лингвистика, история ее развития и инструментарий.

Искусственный интеллект может оказаться

благоразумней его создателя.
Дарий (философ)

I have noticed that even people who claim everything is predetermined and that we can do nothing to change it, look before they cross the road.
Stephen Hawking

В данной главе речь пойдет о компьютерной лингвистике (computational linguistics), которая постепенно и уверенно становится основной частью технологий **искусственного интеллекта.**[15]

Компьютерная лингвистика – это научное направление в области математического и компьютерного моделирования интеллектуальных процессов у человека и животных при создании систем искусственного интеллекта, которое ставит своей целью использование математических моделей для описания естественных языков.

Компьютерная лингвистика частично пересекается с обработкой естественных языков. Однако обработка естественных языков акцентирует внимание не на абстрактные модели, а на прикладные методы описания и обработки языка для компьютерных систем.

Полем деятельности компьютерных лингвистов является

---

[15] **Искусственный интеллект (ИИ,** англ. Artificial intelligence, AI) – 1) наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ; 2) свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека

разработка алгоритмов и прикладных программ для обработки языковой информации.

Основными направлениями компьютерной лингвистики являются следующие:

1. Обработка естественного языка (англ, *natural language processing*; синтаксический, морфологический, семантический анализы текста).

2. Корпусная лингвистика, создание и использование электронных корпусов текстов

3. Создание электронных словарей, тезаурусов, онтологий (например, Lingvo). Словари используют для автоматического и автоматизированного переводов, проверки орфографии и т. д.

4. Автоматический перевод текстов посредством специализированных программ (см. приложение 2).

5. Автоматическое извлечение фактов из текста (извлечение информации; англ .*fact extraction, text mining).*

6. Автореферирование (англ, *automatic text summarization).* Эта функция включена, например, в Microsoft Word.

7. Построение систем управления знаниями (экспертные системы).

8. Создание вопросно-ответных систем (англ, *question answering systems).*

9. Оптическое распознавание символов (англ. *OCR).* Например, программа FineReader.

10. Автоматическое распознавание и синтез речи.

Кроме того, компьютерная лингвистика занимается решением следующих научных задач:

• Компьютерный анализ жанра и характеристик автора текста (более сложный, чем анализ сюжета);

• Компьютерный анализ блогосферы как источник знаний о языке (как вариант анализа корпуса текстов).

• Создание семантической паутины Интернета (формирование пространств знаний) и поиск знаний в ней.

Перечисляя все современные возможности компьютерной лингвистики, вернемся немного в историю формирования этой научной гуманитарной отрасли.

Компьютерная лингвистика родилась в январе 1954 года, когда в Джорджтаунском университете (США) был проведен первый в мире публичный эксперимент по машинному переводу. В то же самое времена под руководством крупнейшего математика и кибернетика Алексея Ляпунова начались активные работы по машинному переводу и в Москве. В созданную Ляпуновым группу вошли, в частности, тогдашние студенты и аспиранты, будущие «родители» отечественной компьютерной лингвистики Игорь Мельчук и Ольга Кулагина.

Русский термин «компьютерная лингвистика» является калькой с английского computational linguistics. Поскольку прилагательное computational по-русски может переводиться и как «вычислительный», в литературе встречается также термин «вычислительная лингвистика», однако в оте-

чественной науке он приобретает более узкое значение, приближающееся к понятию «квантитативной лингвистики».

Как особое научное направление компьютерная лингвистика начала оформляться в 1960-е годы. Хотя в России основа для успехов отечественных ученых в этой области закладывалась намного раньше. В 1920-х годах в Советской России велись интенсивные исследования по семиотике текста, в то время как широкомасштабное изучение семиотики во всем мире относится только к началу 1960-х годов (что в значительной мере связано с работами эмигрировавшего из России в 1920 году создателя структурной лингвистики Романа Якобсона). Стремительный прогресс семиотики стал основой для сближения лингвистики и математики на почве популярного в 1950-60-е годы математического структурализма. Идеи семиотики объединяли крупнейших ученых, среди которых математик Владимир Успенский и лингвист Вячеслав Иванов. Успехи формального подхода к описанию языка наглядно продемонстрировали возможность превращения чисто гуманитарной науки в логически строгую дисциплину.

Работы по кибернетике и, в частности, по структурной лингвистике, развернутые в конце 50-х годов под руководством академика Акселя Берга и члена-корреспондента АН СССР Алексея Ляпунова, вывели отечественную науку на передовые позиции. Уже в начале 1956 года в Институте прикладной математики (ИПМ) им. М. В. Келдыша зарабо-

тала первая отечественная система машинного перевода с французского на русский язык. Система ФР-I давала перевод более высокого качества, чем у американских оппонентов. Математики рассматривали алгоритмы машинного перевода как частные случаи изучаемых в кибернетике алгоритмов перекодирования.

В то же самое время в 1957 году молодой американский исследователь Ноам Хомский публикует свой научный труд «Синтаксические структуры», положения которого до сих пор в компьютерной лингвистике, в частности в автоматической обработке текста, являются доминирующими. Работы Н. Хомского послужили началом рационалистического направления в компьютерной лингвистике. Исходная точка рационализма – компьютерные модели, независимые от языка. Модели лучше всего принимаются, когда они настолько просты, насколько это возможно. Здесь можно провести параллель с идеей Ф. Соссюра отделить язык от реального мира.

Ранние исследователи машинного перевода поняли, что машина не может перевести введенный текст без дополнительной помощи. Учитывая нехватку лингвистических теорий, особенно до 1957 года, выдвигается предложение предварительно редактировать тексты таким образом, чтобы отмечать в них трудности, например, чтобы разрешить омонимию. А поскольку системы машинного перевода не могли производить правильный результат, текст на целевом языке должен был быть отредактирован, чтобы стать понятным.

Мысль о предварительном и последующем редактировании текста породила идею о том, что компьютер может быть использован для оказания помощи человеку в областях, с которыми компьютер не в состоянии пока справляться своими силами. В области машинного перевода компьютер может действовать как память-хранилище, освобождая человека от необходимости знать огромное количество слов. Израильский логик, лингвист и математик Иегошуа Бар-Хиллел (Bar Hillel) рассмотрел область исследований и пришел к выводу, что *полностью автоматический высококачественный перевод* (FAHQT – Full-Automatic High-Quality Translation) не возможен без знаний. Он также пришел к выводу, что многочисленные проекты, в которых перевод сводился главным образом к замене слов одного языка на слова другого, были изначально обречены на провал даже с учетом многочисленных заплаток и расширений. Причина проста: переводчик-человек добавляет свое понимание документа, который нужно перевести, к своим знаниям о структурах языка, с которым он работает. Там остаются некоторые конструкции, которые требуют понимания документа или пути, по которому передаются представления о мире и определенной предметной области. Во многих языках трудно понять, что имеет в виду говорящий предложения, соответствующего типу:

«Она надела красные туфли и чулки».

Сразу же возникает вопрос о цвете чулок. Были ли они тоже красными? Во многих случаях это не имеет значения, но

если система, например, анализирует свидетельские показания, значение таких деталей может существенно возрастать.

Комментарии И. Бар-Хиллеля оказали долговременное влияние на восприятие практичности систем автоматической обработки текстов и машинного перевода, в частности. Постепенно финансирование проектов в области компьютерной лингвистики в США были приостановлены.

В то же самое время в СССР компьютерная лингвистика стремительно развивалась, особенно в 1960-е годы. Однако в следующем десятилетии работы в области машинного перевода оказались под жестким государственным контролем. В отличие от атомного проекта (время уже было другое), этот контроль не сконцентрировал силы отечественных ученых, а наоборот, способствовал прекращению или замедлению многих работ. Исследования по машинному переводу в ИПМ им. М. В. Келдыша тоже практически прекратились.

В США период с 1966 по 1980 годы характеризуется разработками систем SHRDLU, LUNAR и LIFER/LADDER.

Система **SHRDLU**[16] Терри Винограда (Terry Winograd) имитировала поведение робота, манипулировавшего блоками на поверхности стола. Она могла управляться инструкциями, такими как *«Pick up the red pyramid"(«Возьми* красную пирамиду») и отвечать на вопросы типа *«What does the blue box contain*?»(«Что содержит голубой блок?»). Появление SHRDLU имело большое значение, так как оно показало, что

[16] Система SHRDLU – программа понимания естественного языка.

синтаксис, семантика и порождение выводов о мире могут быть скомбинированы так, чтобы создать систему, которая понимает естественный язык. Это была очень ограниченная система: она могла управляться только очень небольшим числом предложений. Более того, она могла понимать язык, касающийся только настоящего момента и очень небольшой части реального мира: мира блоков. Эффект, который она производила, мог быть достигнут только в весьма ограниченной области и попытка расширить систему неизбежно привела бы к снижению эффективности.

Система **LUNAR** был естественно-языковым интерфейсом к базе данных, использовавшим и расширенную сеть переходов[17] и процедурную семантику американского исследователя У.А. Вудса[18]. Система унаследовало свое имя от ба-

---

[17] Расширенная сеть переходов (ATN – Augmented Transition Network) – Расширенная Сеть Переходов представляет собой образец программного обеспечения, продемонстрировавшего возможность использования достаточно мощных грамматических средств для обработки синтаксиса. Неправильно думать о ней только как о средстве обработки синтаксиса, потому что это нечто большее, чем просто реализация поискового алгоритма. Она представила формализм для выражения знаний о предметной области (знания записывались в виде расширенной сети переходов). Был также представлен способ использования этих сетей для поиска путей решения проблем. Применительно к АОТ речь шла о знаниях синтаксиса английских предложений, а проблемой, которую система должна была решать, был синтаксический разбор этих предложений.

[18] У.А. Вудс (William Aaron Woods (born 1942), generally known as Bill Woods, is a researcher in natural language processing,continuous speech understanding, knowledge representation, and knowledge-based search technology. He is currently interested in using technology to help people organize and use information in

зы данных, содержавшей информацию (ATN – Augmented Transition Network) об образцах лунных скальных пород. Система была продемонстрирована на научной конференции по Луне в 1971 году. Ее эффективность казалась весьма впечатляющей: она сумела справиться с 78 % запросов без ошибок, причем эта цифра возросла до 90 % после исправления ошибок. Впрочем, цифры эти не должны были вводить в заблуждение, ведь не случайно система не стала предметом интенсивной эксплуатации: ученый, стремящийся использовать систему для своих повседневных рабочих нужд, быстро бы обнаружил, что ему нужны запросы, выходящие за пределы возможности системы.

Система **LIFER/LADDER** была одной из наиболее впечатляющих систем автоматизированной обработки текстов. Она была сконструирована как естественно-языковой интерфейс к базе данных кораблей ВМС США. Она использовала семантическую грамматику, в которой функционировали метки типа «КОРАБЛЬ» или «ХАРАКТЕРИСТИКА» вместо синтаксических меток по типу «существительное» или «глагол». Это означало, что система, как и SHRDLU, была тесно привязана к области, для которой была изначально сконструирована. Тем не менее, использование разработчиками семантической грамматики дало ряд преимуществ в разработке дружественного к пользователю интерфейса по сравнению с SHRDLU. Например, в систему была включена

---

organizations)

возможность определения новых словарей, определения парафразов (например, чтобы сделать возможным быстрый доступ), возможность обработки незаконченного или неполного ввода. Эти свойства сами по себе были очень впечатляющими, но исследовательская группа приступила к программе строгой оценки и опубликовала доклад, ставший настоящим сокровищем для тех, кто стремился глубоко изучать автоматическую обработку текстов. Одним из выводов доклада было то, что люди быстро подстраивались под машину и пытались использовать очень неполные предложения, заменяя нормальный естественный язык подобием неформального языка запросов.

С середины семидесятых годов во всем мире наблюдается устойчивое возрастание интереса к машинному переводу. В Москве в 1974 в институте ИНФОРМ-ЭЛЕКТРО начались работы по созданию системы франко-русского перевода (ЭТАП-1) и системы англо-русского перевода (ЭТАП-2). В том же году создается Всесоюзный центр переводов (ВЦП), в котором ряд научных коллективов работает над системами машинного перевода – АМПАР (англо-русский перевод), НЕРПА (немецко-русский перевод) и ФРАП (французско-русский перевод). С этого времени промышленные системы машинного перевода разрабатываются и широко используются в США, Европе и Японии.

Семантические системы 1970-х совершенно сознательно избегали использования синтаксической обработки, неко-

торые пытались вообще очистить от синтаксической информации свои системы. Синтаксис всегда рассматривался большинством лингвистов-теоретиков как основополагающая часть человеческого языка. Инженеры же увидели в синтаксисе полезный способ разрешения омонимии с относительно небольшим объемом необходимых знаний (по крайней мере, по сравнению с объемом знаний, требуемым для этой цели семантической обработкой). Теоретические лингвисты также выступали критически против всеобщего признания трансформационной (порождающей) грамматики Н. Хомского.

Результатом стали грамматики, оперирующие более детализированными синтаксическими характеристиками объектов информации (например, часть речи – существительное, лицо – третье, время – прошедшее, число – множественное и т. д.), а не одноатомные категории (глагол, существительное, прилагательное и т. д.). Эти грамматики могли дать гораздо более точный анализ предложения. С другой стороны, для детализированных характеристик необходимы гораздо более сложные методы сопоставления в поисковых алгоритмах. Отсюда такие грамматики получили название унификационных, так как сопоставление характеристик могло быть достигнуто посредством метода унификации.

Есть несколько грамматик, которые используют унификацию как главную операцию для комбинирования информации. Из них Грамматика Обобщенной Фразовой Структу-

ры (GPSG – Generalized Phrase Structure Grammar), была одно время очень популярной, но, видимо, самым широко используемым формализмом стала Грамматика Лексических Функционалов (EFG – Eexical Functional Grammar).

В 80-е годы в большой мере формируется фундамент современного подхода к структуре машинного перевода. Благодаря росту производительности и развитию возможностей компьютеров, разработка систем машинного перевода стала реальностью. Разработка первых систем была основана на лингвистических знаниях. Но лингвистике не удалось покрыть широкий диапазон актуальных феноменов использования языка. Анализ производился для ограниченных случаев. В 80-е годы ученые инициировали разработки в области описания грамматик. Грамматики, основанные на формальных, правильно написанных текстах, не имели полной практической ценности. Только про 60 % грамматических правил, выработанных лингвистами, можно было сказать, что они работают на практике. В реальности, различные варианты в языке были слишком многочисленны.

В начале 90-х годов IBM выпустила систему статистического машинного перевода (SMT-statistical machine translation). Данная система обнаруживала ряд английских и французских выражений, которые не могли бы быть распознаны вручную, посредством «чистой» методологии машинной трансляции. Базовыми характеристиками данной системы являлись большой объем памяти и высокая производи-

тельность компьютеров, большой объем качественных пар слов для перевода (обучающие данные).

Основным вопросом, вставшим перед компьютерной лингвистикой с 90-х годов, когда основным направлением стала статистическая обработка текстов, оказалась проблема получения достаточно большого набора реальных лингвистических данных, чтобы произвести внятный анализ или автоматически построить грамматику. Во многих случаях у исследователей недостает данных для анализа лингвистических феноменов в результате разбросанного характера данных. Например, когда они пытаются определить вероятность для каждой из пар последовательно идущих терминов, они не могут найти лингвистических данных с какими-то из анализируемых пар, хотя сочетания подобного рода являются правильными с точки зрения языка. Один из подходов состоит в улучшении точности приближения путем статистической обработки небольших рабочих образцов.

В реальности, исследователям приходится работать с низкокачественными данными в сопоставляемых для перевода парах. Когда данные не могут быть сопоставлены, используются двуязыковые корпуса в той же предметной области. Как минимум для решения проблемы требуются словари для базовых лексиконов и быстрые компьютеры с большим объемом памяти. При этом парадигмы машинного обучения выглядят на сегодняшний день мало подходящими для целей автоматической обработки текстов.

На сегодняшний день максимальную долю российского рынка систем машинного перевода составляют продукты компаний PROMT и ABBYY (Lingvo). В основу фундамента технологии перевода PROMT были заложены формализм расширенных сетей переходов (ATN – Augmented Transition Network) и использование двух «переводческих технологий» в одном продукте – технологии машинного перевода (Machine Translation) и технологии Translation Memory. Эффект от взаимного применения двух технологий позволяет обеспечить практически 100 %-ное качество перевода при работе с повторяющимися текстами.

Программы, разработанные на основе технологии машинного перевода осуществляют связный перевод текста, используя определенные лингвистические алгоритмы. Сначала система анализирует структурные элементы входного предложения, затем преобразует его в соответствии со структурой языка и синтезирует окончательный вариант. Кроме того, для повышения качества перевода программа должна уметь распознавать устойчивые выражения, а также иметь большой словарный запас. Для перевода тематических текстов обычно требуется подключать специализированные словари. С помощью систем с технологией машинного перевода можно быстро получить черновой вариант перевода, отражающий общий смысл текста.

Во второй лекции нашей работы, посвященной описанию разнообразных программ компьютерного перевода, техно-

логия **Translation Memory** была подробно описана. Повторимся, что память переводов (TM) использует базу данных, где хранятся выполненные профессионалом переводы в виде сегментов текста оригинал-перевод. Эта технология базируется на сравнении документа, который нужно перевести, с данными, хранящимися в предварительно созданной базе переводов. Память переводов работает по принципу накопления: в процессе перевода в базе сохраняется исходный сегмент (предложение) и его перевод. При обработке нового текста, поступившего на перевод, система сравнивает каждое его предложение с сохраненными в базе сегментами. Если идентичный или подобный исходному сегмент найден, то перевод этого сегмента отображается вместе с переводом и указанием совпадения в процентах. Слова и фразы, которые отличаются от сохраненного текста, выделяются подсветкой. Таким образом, переводчику остается перевести только новые сегменты и отредактировать частично совпадающие. Каждое изменение или новый перевод сохраняются в базе. В результате необходимость в повторном переводе одного и того же предложения отпадает.

Современная компьютерная лингвистика является междисциплинарной наукой. Человека окружает очень большое количество цифровой информации; существует множество бизнес-проектов, успех которых зависит от обработки информации, эти проекты могут относиться к сфере маркетинга, политики, экономики и т. д. И очень важно уметь обра-

щаться с этой информацией эффективно – главное не только быстрота обработки информации, но и легкость, с которой пользователь способен извлекать необходимые ему данные и создавать из них цельную картину.

Компьютерная лингвистика как особая прикладная дисциплина выделяется, прежде всего, по инструменту – т. е. по использованию компьютерных средств обработки языковых данных. Поскольку компьютерные программы, моделирующие те или иные аспекты функционирования языка, могут использовать самые различные средства программирования, то об общем понятийном аппарате компьютерной лингвистики говорить вроде бы не приходится. Однако это не так. Существуют общие принципы компьютерного моделирования мышления, которые так или иначе реализуются в любой компьютерной модели. В их основе лежит теория знаний, первоначально разрабатывавшаяся в области искусственного интеллекта, а в дальнейшем ставшая одним из разделов когнитивной науки. Важнейшими понятийными категориями компьютерная лингвистика являются такие структуры знаний, как «фреймы» (понятийные, или, как принято говорить, концептуальные структуры для декларативного представления знаний о типизированной тематически единой ситуации), «сценарии» (концептуальные структуры для процедурного представления знаний о стереотипной ситуации или стереотипном поведении), «планы» (структуры знаний, фиксирующие представления о возможных действиях, ведущих

к достижению определенной цели). Тесно связано с категорией фрейма понятие «сцена». Категория сцены преимущественно используется в литературе по компьютерной лингвистике как обозначение концептуальной структуры для декларативного представления актуализованных в речевом акте и выделенных языковыми средствами (лексемами, синтаксическими конструкциями, грамматическими категориями и пр.) ситуаций и их частей.

Определенным образом организованный набор структур знаний формирует «модель мира» когнитивной системы и ее компьютерной модели. В системах искусственного интеллекта модель мира образует особый блок, в который в зависимости от выбранной архитектуры могут входить общие знания о мире (в виде простых пропозиций типа «зимой холодно» или в виде правил продукций «если на улице идет дождь, то надо надеть плащ или взять зонтик»), некоторые специфические факты («Самая высокая вершина в мире – Эверест»), а также ценности и их иерархии, иногда выделяемые в особый «аксиологический блок».

Большинство элементов понятий инструментария компьютерной лингвистики омонимично: они одновременно обозначают некоторые реальные сущности когнитивной системы человека и способы представления этих сущностей, используемые при их теоретическом описании и моделировании. Иными словами, элементы понятийного аппарата компьютерной лингвистики имеют онтологический и ин-

струментальный аспекты. Например, в онтологическом аспекте разделение декларативных и процедурных знаний соответствует различным типам знаний, имеющимся у человека – так называемым знаниям ЧТО (декларативным; таково, например, знание почтового адреса какого-либо NN), с одной стороны, и знаниям КАК (процедурным; таково, например, знание, позволяющее найти квартиру этого NN, даже не зная ее формального адреса) – с другой. В инструментальном аспекте знание может быть воплощено в совокупности дескрипций (описаний), в наборе данных, с одной стороны, и в алгоритме, инструкции, которую выполняет компьютерная или какая-либо другая модель когнитивной системы, с другой.

Одна из ключевых задач современной компьютерной лингвистики – это совершенствование структуры семантических сетей, когда поиск происходит не просто по совпадению слов, а по смыслу. Ведь все сайты, так или иначе, размечены по семантике. Это может быть полезно, например, для полицейских или медицинских отчетов, которые пишутся каждый день. Анализ внутренних связей дает много нужной информации, а читать и считать это вручную невероятно долго[19].

С другой стороны, компьютерная лингвистика занимает-

[19] Речь идет о существовании тысячи текстов, которые необходимо сгруппировать, представить каждый текст в виде структуры и получить таблицу, с которой уже можно работать. Это называется обработка неструктурированной информации.

ся, например, созданием искусственных текстов. Например, существуют механизмы генерации текстов на темы, на которые человеку писать довольно-таки скучно: изменение цен на недвижимость, прогноз погоды, отчет о футбольных матчах. Заказ подобных текстов стоит немалых денежных затрат, но написаны компьютерные тексты на данные темы связным человеческим языком.

В современной России одним из наиболее успешных проектов, реализованных в области компьютерной лингвистике, является **Национальный корпус русского языка** (http://ruscorpora.ru/). Это один из лучших национальных корпусов в мире, который стремительно развивается и открывает невероятные возможности по научным и прикладным исследованиям. В современном англоязычном мире большим достижением компьютерной лингвистики является концептуальная сеть Ргатепе![20], где формально представлены все возможные связи какого-то конкретного слова с другими словами. Например, есть слово «летать» – кто может летать, куда, с каким предлогом употребляется это слово, с какими словами оно сочетается и так далее. Этот ресурс помогает связать язык с реальной жизнью, то есть проследить, как ведет себя конкретное слово на уровне морфологии и синтаксиса.

В качестве послесловия необходимо отметить, что пока не известно, какие возможности даст человечеству надвигающаяся новая компьютерная революция. Однако можно на-

[20] См. ссылку – https://framenet.icsi.berkeley.edu/fndmpal/home.

деяться, что компьютерная лингвистика перейдет на совершенно новую технологическую базу, основа которой закладывается в наше время, в эпоху научных разработок в области искусственного интеллекта.

Вопросы для самоконтроля

1. Каковы основные задачи, решаемые современной компьютерной лингвистикой?
2. Что такое искусственный интеллект?
3. Перечислить основные этапы становления и развития компьютерной лингвистики.
4. Каково приоритетное направление развития современной компьютерной лингвистике?
5. Что такое Национальный корпус русского языка и Framenet?

Задания для самостоятельного исследования

Подготовить доклад по одной из ниже перечисленных тем.

1. Компьютерная лингвистика как междисциплинарное научное направление.
2. Когнитивный инструментарий компьютерной лингвистики. «Фреймы», «сценарии» и «планы».
3. Компьютерное обеспечение представления знаний.
4. Естественные и искусственные языки. Виды искусственных языков.

5. Автоматизированный анализ: распознавание и синтез устной и письменной речи.

6. Морфологический анализ, проблемы семантического анализа, синтаксический анализ.

7. Лингвистические базы данных: модели и типы данных. Создания общих искусственных языков для представления информации.

8. Компьютерная лексикография как одно из направлений прикладной лингвистики. Словарные процессоры.

9. Основные понятия структуры словаря: словник, словарная статья, грамматические, стилистические пометы; иллюстративный материал.

10. Типология электронных словарей.

11. Тезаурусы и терминологические словари.

12. Компьютерные технологии составления и эксплуатации словарей.

13. Электронные учебники, словари, учебно-методические материалы.

14. Мультимедиа в помощь филологу.

15. Использование инновационных технологий при организации научных исследований.

Рекомендованная литература

1. *Кравченко, А.В.* От языкового мифа к биологической реальности: переосмысляя познавательные установки языкознания/ А.В. Кравченко. – Москва: Языки славян-

ских культур (ЯСК): Рукописные памятники Древней Руси, 2013. – 387 с.

2. *Болховитянов, А.В.* Алгоритмы морфологического анализа компьютерной лингвистики: учеб, пособие для студентов вузов, обучающихся по направлению 035000.62 – Издательское дело / А.В. Болховитянов, А.М. Чеповский; М-во образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высш. проф. образования Московский гос. ун-т печати им. Ивана Федорова. – Москва: МГУП, 2013.– 198 с.

3. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие для студентов высших учебных заведений, обучающихся по направлению 231300 – »Прикладная математика»/ [Большакова Е. И. и др.]; М-во образования и науки Российской Федерации, Московский гос. ин-т электроники и математики Москва: Московский гос. ин-т электроники и математики, 2011. – 272 с.

4. *Зубов А.Б.* Информационные технологии в лингвистике/ А.Б. Зубов, И.И. Зубова. М.: «Академия», – 2004. – 208 с.

5. *Кузнецов АЛ.* Образовательные электронные издания и ресурсы: методическое пособие / С.Г. Григорье, В.В. Гриншкун. – М.: Дрофа, 2009. – 156 с.

6. *Потапова Р.К.* Речь: коммуникация, информация, кибернетика: Учеб, пособие для студентов вузов, обучающихся по специальностям «Авто-матизир. системы обраб. информ.

и упр.», «Лингвистика» / Р.К. Потапова. – 3. изд., стер. – М.: УРСС, 2003. – 564 с.

7. *Соснина Е.П.* Введение в прикладную лингвистику/Е.П. Соснина. – Ульяновск, 2010. [Электронный ресурс]. – URL:http://www.twirpx.com/file/736011/ – электронный учебник.

Глава 5
ИНТЕРНЕТ-ЛИНГВИСТИКА КАК НОВОЕ НАУЧНОЕ НАПРАВЛЕНИЕ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ XXI СТОЛЕТИЯ

Краткое описание: Определение понятия «Интернет-лингвистика», становление Интернет-лингвистики как нового научного направления в лингвистическом знании, социолингвистическая, образовательная, стилистическая и практическая концепции изучения Интернет-лингвистики.

The Web is an eclectic medium, and this is seen also in its multilinguistic inclusiveness. Not only does it offer a home to all linguistic styles within a language; it offers a home to all languages – once their communities have a functioning computer technology.

David Crystal (a British linguist, academic and author)

Web screens may blossom with movies and be garnished with sound tracks but, for the moment, type is the primary vehicle for information and persuasion. Its appearance on screen is more crucial than ever. Intense competition for the user's attention means that words must attract, inform (and maybe seduce) as quickly as possible. Flawless delivery of the message to the screen is the goal. The road to success is very broad, but the surface rather uneven.

Roger Bring (author, graphic designer and educator)

Интернет-лингвистика – подраздел лингвистики, кото-

рый был сформулирован Дэвидом Кристалом. Этот подраздел науки занимается изучением новых форм употребления и использования языка, которые возникли под влиянием активного развития Интернет-пространства и иных «новых» средств передачи информации, таких как текстовые сообщения. Начиная с момента появления науки, изучающей мотивацию человеческого поведения при работе с компьютерными системами (человеко-компьютерное взаимодействие, human-computer interaction, HCI), которая, в свою очередь, привела к возникновению связи путем использования компьютера (computer-mediated communication, CMC) или Интернета (Internet-mediated communication, IMC), эксперты признали, что лингвистика играет ключевую роль в формировании этой науки, особенно в отношении восприятия web-интерфейса. Изучение развивающегося языка Интернета позволяет обеспечить дальнейшее развитие онлайн-пространства и может помочь не только лингвистам, но и самим пользователям.

На сегодняшний день существуют четыре основные концепции Интернет-лингвистики: социолингвистическая, образовательная, стилистическая и практическая. Все эти концепции взаимосвязаны и имеют влияние друг на друга.

Социолингвистическая концепция

Эта концепция связана с изучением того, как общество оценивает влияние Интернета на изменение и развитие языков. Появление Интернета колоссально изменило общение

между людьми и создало новые платформы для коммуникации (такие как текстовые сообщения, чаты, электронная почта, использование смайлов, и т. д.)

Развитие новых платформ для общения вызвало обеспокоенность в отношении использования языка. Согласно Кристалу (2005) эта озабоченность не только не беспочвенна, но и уже не раз наблюдалась в истории – она всплывает каждый раз, когда технический прорыв влияет на язык. Примером подобных открытий могут служить появление печати в 15 веке, изобретение телефона в 19 веке или распространение радиовещания и телевещания в 20 веке.

Влияние Интернета можно рассматривать на уровне частного и профессионального пользования.

На личном уровне компьютерное общение осуществляется посредством текстовых сообщений или мобильных электронных писем, что привело к значительному расширению возможностей мгновенного общения. Примеры этого могут связаны с использованием Iphone и BlackBerry.

Сейчас во многих учебных заведениях и преподавателям, и студентам предоставляются личные адреса электронной почты и аккаунты в специально созданных для учебных целей социальных сетях для ускорения обмена материалами, данными, а также для более быстрого доступа ко всей необходимой информации. Дискуссии в классе часто переходят в онлайн-пространство путем создания специализированных форумов. Например, студенты Наньянского технологическо-

го университета в образовательных целях объединяются на специализированном учебном портале *edveNTUre*, на котором они участвуют в дискуссиях, слушают и смотрят материалы, которые специально были подготовлены для них преподавателями, и выполняют онлайн-работы. Компания Apple в 2008 году запустила платформу iTunes U, которая представляет собой базу для размещения аудио- и видеокурсов от преподавателей крупнейших учебных заведений всего мира, которые пользователи этого ресурса могут смотреть и слушать абсолютно бесплатно. В числе партнеров iTunes U уже более 600 университетов из 18 стран, включая Оксфорд, Кэмбридж и Йельский Университет.

Подобная форма предоставления информации позволяет преподавателям находить новые способы общения со студенческой аудиторией, что дает возможность не только расширить аудиторию, которой адресуются материалы, но и сделать эти материалы более доступными. В Нью-Йоркском университете студенты привыкли к тому, что многие внештатные преподаватели читают им лекции по Skype, служащие библиотеки активно используют мгновенные сообщения для предоставления необходимой информации, а все услуги самой библиотеки доступны онлайн. Дальнейшее развитие подобных платформ общения с использованием компьютерных систем и их популярностью среди как преподавателей, так и студентов приведет к изменению языковых форм.

На профессиональном уровне социолингвистическая концепция проявляется в том, что практически все компании предоставляют своим сотрудникам доступ в Интернет, а также дают возможность использовать все корпоративные ресурсы и электронную почту. Подобная практика упрощает как внутреннее взаимодействие сотрудников компании, так и внешнее (с другими компаниями). На современном этапе многие создатели программного обеспечения для смартфонов стараются активно выйти на корпоративный рынок посредством создания возможностей для интеграции корпоративных ресурсов в телефон каждого сотрудника (например, компания Apple через ActiveSync позволяет сотрудникам связываться с рабочей электронной почтой, получать доступ к календарю и иным необходимым рабочим ресурсам удаленно, не находясь в офисе).

Дэвид Кристал считает, что создание новых средств связи с использованием компьютерных систем не приведет к деградации языка, а, наоборот, покажет насколько разнообразным может быть общение в Интернете.

Образовательная концепция

С образовательной точки зрения Интернет-лингвистика занимается изучением влияния Интернета на использование формального языка, в особенности на академический английский, который, в свою очередь, оказывает влияние на лингводидактику. Быстрое развитие Интернета повлекло за

собой появление новых языковых особенностей, характерных только для онлайн-пространства. Они включают в себя возрастание использования неформального письменного языка, противоречивость в стиле написания и стилистике, использование новых аббревиатур в Интернет-чатах и мгновенных текстовых сообщениях (СМС), в которых технические ограничения на количество слов привели к появлению новых аббревиатур. Подобные акронимы в своем большинстве возникают по практическим соображениям – не только из-за технических ограничений, но и сокращения времени и количества приложенных усилий на общение через эти средства коммуникации. Примеры подобных сокращений – ЛОЛ (от англ. Laughing out loud) или ОМГ (от англ. Oh my god).

Образовательная перспектива была во многом создана при исследовании влияния Интернета на обучение использованию языка. Это очень важный аспект, так как он влияет на обучение нынешних и будущих поколений студентов правильному и своевременному использованию неформального языка, который возникает в процессе использования Интернета. Эта обеспокоенность связана с использованием неформального языка в учебе или на официальных событиях. Кроме того, возникает проблема, связанная с более частым использованием студентами в институтских работах аббревиатур и сокращений, характерных для онлайн-пространства.

Лингвист и профессор Элеанор Джонсон предполагает, что широко распространенные ошибки в письме напрямую

связаны с использованием Интернета, в котором преподаватели также обнаружили грамматические и орфографические ошибки в работах студентов. Однако не существует никаких научных доказательств для подтверждения этой гипотезы. Хотя и существует обоснованное беспокойство, связанное с влиянием Интернета на академические работы студентов, оно обостряется в связи с неформальной природой новых средств связи. Лингвист и профессор Наоми С. Барон (Naomi Baron) доказывает, что Интернет (Internet-mediated communication, IMC) (интернет-чаты, сообщения и почта) имеет ничтожное влияние на правописание студентов. Недавнее исследование, опубликованное Британским психологическим обществом (British Psychological Society (BPS)) обнаружило, что студенты, которые часто отправляю CMC-сообщения, имеют более обширный словарный запас, что может привести к положительному влиянию на их развитие в области чтения.

Несмотря на то, что использование Интернета повлекло за собой использование стилистических форм, недопустимых в формальном академическом языке, было также отмечено, что использование Интернета необязательно должно навредить изучению языка, но также может и помочь. Интернет различными способами доказал, что он может помочь улучшить языковые знания, особенно при изучении иностранных языков. Интернет позволяет улучшить взаимодействие между теми, кто изучает иностранный язык и носителями языка,

предоставляя возможность исправить ошибки, а также приобрести навыки ведения переговоров и убеждения.

Стилистическая концепция

Стилистическая концепция изучает то, как Интернет и связанные с ним технологии способствовали развитию новых языковых форм, в особенности в литературе. Стилистическая концепция рассматривает Интернет как средство, через которое возникли новые языковые феномены. Новая модель языка представляет интерес для изучения, так как являет собой смесь устной и письменной речи. Например, традиционное письмо сравнивается с динамично развивающимся Интернет-языком, в котором слова появляются в разных цветах и размерах шрифтов на компьютерном экране. Кроме того, новая модель языка также содержит в себе элементы, которые нельзя встретить в обычном, естественном языке. Примером подобных проявлений может стать обрамление сообщений, которое используется в ответах на электронные письма или в обсуждениях на форумах. В ответ на письмо люди, в своем большинстве, используют сообщение отправителя в качестве рамки для написания ответа. Есть возможность выбирать ответить на письмо полностью или только на некоторые его части, оставляя фрагменты текста неиспользованными. Дискуссии на форумах также дают возможность развить новую ветвь беседы к любому из комментариев, оставленному предыдущим участником.

Предстоящие исследования также включают в себя множество новых выражений, которые появляются благодаря Интернету и другим технологиям, и их влияние не только на письменный язык, но и на устную речь. Стилистика использования Интернет-языка наиболее полно отражается в приведенных ниже средствах связи, так как при их использовании всегда существует попытка обойти технические ограничения.

Практическая концепция

Практическая концепция рассматривает Интернет с точки зрения его коммуникационных возможностей – плохого и хорошего. Интернет представляет из себя платформу, на которой пользователи могут ощутить существование «мультиязыка» (multilingualism). Хотя английский по-прежнему является доминирующим языком в Интернете, количество пользователей с другими языками постепенно возрастает. Статистика глобального использования Интернета (Global Internet usage) содержит данные о национальности, языковой принадлежности и географическом расположении пользователей Интернета. Количество используемых в Интернете языков возрастает пропорционально тому, как все больше членов языковых общностей становятся пользователями сети. Для них Интернет также является платформой, где можно оживить вымирающий язык и осведомлять о его существовании других пользователей. Интернет предоставля-

ет этим языкам возможность развиваться в двух направлениях – возрождение языков и документация языков (language documentation).

Будущее Интернет-лингвистики

С появлением более совершенных систем коммуникации посредством Интернета/использования компьютера в купе с большей готовностью людей подстраиваться под новые требования усложняющегося с технологической точки зрения мира ожидается, что большее количество пользователей продолжат изменять свой язык, чтобы иметь возможность участвовать в новых формах коммуникации.

Количество пользователей в Интернете по всему миру растет и скорость встраивания в мировую паутину различных культур, лингвистических особенностей и различий в языках отдельных групп людей. Эти лингвистические особенности каждого отдельного пользователя Интернета в будущем сыграют огромную роль в развитии Интернет-лингвистики, что станет наиболее ярко выражено в количестве языков, присутствующих в Интернете. С 2000 по 2010 год Интернет-бум настиг неанглоговорящие страны (Китай, Индию, страны Африки), что привело к проникновению других языков (не английского) в онлайн-пространство.

Особой темой для изучения может стать взаимодействие английского языка с другими языками, на основе чего рождаются новые формы общения между представителями раз-

ных культур. При смешении двух языков образуются новые стилистические формы, которые впоследствии могут перейти в другие культуры. Китайский и корейский языки уже изменились под влиянием английского языка, что привело к возникновению уникального мультиязыкового Интернет-жаргона.

В текущем своем состоянии Интернет предоставляет возможность для обучения и популяризации редких языков. Однако подобно тому, как межъязыковое взаимодействие способствовало внедрению английского языка в китайский и корейский языки и появлению в них сетевого сленга, малораспространенные языки также подвержены изменению под воздействием языков, широко используемых в интернете (таких, как, например, английский и испанский). Помимо того, что межъязыковое взаимодействие может снизить степень чистоты и подлинности редких языков, их лингвистическая близость к более распространенных языкам может также негативно повлиять на их развитие. Например, те, кто намеревается изучить язык этнического меньшинства, могут получить все необходимую информацию о нем на общеизвестном языке и остановиться на этом, тем самым отрицательно влияя на количество потенциальных носителей и так малораспространенного языка. К тому же, носители редких языков стараются приобщиться к массовой культуре и изучают более популярные языки, чтобы быть в курсе последних событий, что, в свою очередь, также приводит к посте-

пенному вымиранию их родных языков.

Вопросы для самоконтроля

1. Дайте определение понятию Интернет-лингвистика. Кто впервые ввел этот термин в научный язык?

2. Охарактеризуйте функционирование Интернет-лингвистики с точки зрения социолингвистической концепции.

3. Каковы стилистические и практические характеристики Интернет-лингвистики?

4. Каковы приоритеты развития Интернет-лингвистики в будущем?

Задания для самостоятельного исследования

1. Подготовить сообщение по современным блогам и их разновидностям (фотоблоги, видеоблоги, аудиблоги и моблоги).

2. Подготовить доклад по электронной почте – одной из самых популярных Интернет-технологий для изучения в рамках Интернет-лингвистики.

3. Подготовить доклад по лингвистическим особенностям мгновенных сообщений.

4. Подготовить доклад по документации и возрождению языков в Интернете.

Рекомендованная литература

1. *Горошко, Е.И.* Лингвистика Интернета: формирование

дисциплинарной парадигмы /Е.И. Горошко // Жанры и типы текста в научном и медийном дискурсе. – Орел: Картуш, 2007. – Вып.5. – С.223–237.

2. *Кравченко, А.В.* От языкового мифа к биологической реальности: переосмысляя познавательные установки языкознания/ А.В. Кравченко. – Москва: Языки славянских культур (ЯСК): Рукописные памятники Древней Руси, 2013. – 387 с.

3. *Болховитянов, А.В.* Алгоритмы морфологического анализа компьютерной лингвистики: учеб, пособие для студентов вузов, обучающихся по направлению 035000.62 – Издательское дело / А.В. Болховитянов, А.М. Чеповский; М-во образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высш. проф. образования Московский гос. ун-т печати им. Ивана Федорова. – Москва: МГУП, 2013.– 198 с.

ПРИЛОЖЕНИЯ

Приложение 1

Ключевые характеристики профессии *лингвист.*

Сложно переоценить значение языка в нашей жизни. Без него не было бы полноценного общения. Все мы изучали в школе родной и иностранные языки. Нас учили краткой истории языка, основам правописания, структуре. Все мы имеем базовые познания в этой области. Лингвисты – это те, кто посвятил изучению языков. Какие знания получают лингвисты в высших учебных заведениях? На самом деле они узнают много интересной информации. Они изучают эволюцию языков и диалектов, получают глубокие познания в сфере феноменов речи. На парах они узнают о родстве мировых языков. Все это очень и очень интересно.

Наука о языках имеет многовековую историю. Скажем больше, ее можно измерять тысячелетиями. Составлять алфавиты и даже классификации начали еще до нашей эры. Первооткрывателями в этой сфере были ученые Древней Греции и Китая, Арабских стран и Индии. Их труды внесли огромный вклад в развитие науки о языках. Сейчас большое количество людей считает лингвистику своим призванием. Это факт.

Описание

Профессию лингвиста относят к широко распространенным. Деятельность этого специалиста не связана с созданием конкретного продукта. При обучении в высшем учебном заведении будущие лингвисты изучают структуру и историю языка. Преподаватели их учат пониманию закономерностей в развитии народов. При глубоком понимании языков в будущем можно создать эффективные методики преподавания языка студентам-иностранцам.

Профессия лингвиста имеет свои подводные камни. Многие люди считают работу в сфере изучения языков скучной. На самом деле они неправы. Лингвистика – интересная наука. Однако в данном случае нужно иметь тягу к познаниям, обладать пытливостью и усидчивостью. Можно сказать, что лингвистика – это тяжелый труд, направленный на систематизацию полученной информации. Здесь важно умение грамотного изложения своего взгляда на те или иные языковые проблемы. Для того чтобы стать хорошим лингвистом, придется потратить немало времени на изучение мертвых языков, воссоздание текстов.

На каких специальностях учиться?

Не смотря на то, что лингвистов не так много в нашей стране, как скажем, экономистов, существует немало направлений в данной профессии. Некоторые из них мы сей-

час с вами рассмотрим.

Классифицировать лингвистов можно по нескольким направлениям.

• Тема или раздел лингвистики

В этой сфере высоко востребованы морфологи, семантисты, синтаксисты, формалисты. Если вы решите выбрать одну из этих областей знаний, то при условии успешного обучения в высшем учебном заведении, сможете достигнуть немалых высот.

• Изучаемый язык

В этом случае выделяют специалистов по конкретному языку. Это могут быть арабисты, русисты, татароведы, агнлисты. Список можно продолжать бесконечно долго. Так что каждый, кто увлекается языками, может выбрать направление по душе. Существуют специалисты по группе языков. Здесь еще более интересно, но в обучении нужно будет вложить немало сил. Итак, выделяют романистов, монголоведов, германистов, тюркологов. Нельзя не сказать об американистах и африканистах. Такие специалисты состоят в группе изучения языка по регионам.

• Теоретическое направление

К такого рода специалистам относят функционалистов, формалистов, структуралистов, когнитивистов.

Прежде чем определяться со специализацией, стоит взвесить все «за» и «против». Будущая сфера деятельности должна вам нравиться. В обратном случае вряд ли вы смо-

жете добиться успеха.

Если говорить более конкретно, то можно получать образование по следующим направлениям.

• Специальность филология. В этом случае вы можете стать бакалавром и магистром филологии.

• Специальность лингвистика (бакалавр или магистр лингвистики).

• Фундаментальная и прикладная лингвистика.

• Интеллектуальные системы в гуманитарной среде

• Перевод и переводоведение.

Чем приходится заниматься на работе

Работа лингвиста может быть самой различной. Люди, специализирующиеся на изучении языков, могут заняться преподавательской деятельностью. Лингвисты довольно высоко востребованы в сфере разработки компьютерного софта. При желании можно связать свою работу с научно-исследовательской деятельностью, заниматься подготовкой учебной литературы, писать научные статьи, принимать участие в разработке учебников. Довольно интересна научно-исследовательская деятельность.

Кому подходит данная профессия?

Если вы относите себя к числу людей, которые обладают внимательностью, хорошей памятью, терпением и усидчивостью, дедуктивным и ассоциативным мышлением, упор-

ством и усидчивостью, то эта профессия для вас. В этой профессии важна склонность к исследовательской деятельности, упорство и усидчивость, пытливый ум, стремление к совершенству. Если вы обладаете всеми этими качествами, то вам стоит выбрать лингвистику в качестве профессии.

Востребованность профессии и трудоустройство

При условии успешного получения высшего лингвистического образования специалисту гарантировано трудоустройство. Лингвисты могут работать в различных сферах: бюро переводов, учебные заведения, литературные союзы, печатные издания, музеи и музейные объединения, гостиницы и гостиничные комплексы. Также вас примут на работу в библиотеку, в издательство. Вы можете устроиться на должность секретаря-референта в международную компанию. Лингвисты сейчас нужны практически везде. Ведь в последнее время международные отношения становятся все крепче.

Карьера и перспективы

Успешная карьера зависит не только от выбранной специальности, но от личных качеств. Если вы будете проявлять упорство и настойчивость, у вас большая воля к победе, то построить успешную карьеру будет не так уж сложно. Главное, желание добиться высот во что бы то ни стало. Наибольшие шансы сделать успешную карьеру будут у тех, у кого получится устроиться в коммерческую организацию.

У лингвистов довольно много перспектив, поскольку сейчас немало сфер, в которых крайне полезны знания таких специалистов. Здесь важно правильно выбрать нишу. Тогда карьерный рост и высокая зарплата будут обеспечены.

Приложение 2

Программы машинного и автоматизированного перевода текстов

Яндекс. Переводчик – веб-сервис компании Яндекс (translate.yandex.ru), предназначенный для перевода части текста или веб-страницы на другой язык.

В сервисе используется самообучаемый алгоритм статистического машинного перевода, разработанный специалистами компании. Система строит свои словари соответствий на основе анализа миллионов переведённых текстов. Текст для перевода компьютер вначале сравнивает с базой слов, затем с базой моделей языка, стараясь определить смысл выражения в контексте.

Нужно отметить, что функция перевода полученных в результатах поиска страниц (по кнопке «Перевод») появилась ещё в 2009 году и осуществлялась на основе технологий PROMT.

Кроме того, переводчик встроен в Яндекс. Браузер и автоматически предлагает перевести текст на иностранном языке.

По состоянию на июль 2015 года доступен перевод для 63 языков.

Направление перевода определяется автоматически. Воз-

можны перевод отдельных слов, целых текстов и отдельных интернет-страниц (по ссылке). При ручном вводе текста система сама предлагает подсказки во всплывающем окне. Есть возможность двухоконного просмотра перевода и оригинала для веб-страниц. Помимо собственно машинного перевода, доступен и полный англо-русский и русско-английский словарь. Имеется приложение для устройств на базе iOS, Windows Phone и Android. Можно прослушать произношение перевода и оригинального текста (синтезированный женский голос).

Переводы предложений и слов можно добавлять в «Избранное» – соответствующий раздел располагается под полем ввода.

Переводчик от Яндекса, подобно другим инструментам автоматического перевода, имеет свои ограничения. Этот инструмент имеет целью помочь читателю понять общий смысл содержания текста на иностранном языке, он не предоставляет точных переводов.

Déjà Vu /Дежа Вю (www.atril.com) – проприетарная (частная, патентованная, в составе собственности) система автоматизированного перевода, разработанная испанской компанией Atril Language Engineering.

Déjà Vu X способен создавать проекты и работать с Microsoft Word, Rich Text Format, Windows Help, Microsoft Excel, Micosoft PowerPoint, Microsoft Access, OpenOffice/

StarOffice, Adobe FrameMaker MIF, Adobe InDesign, Adobe PageMaker tagged text, QuarkXPress XTG, Interleaf ASCII, HTML, ASP/ASP.NET, PHP, JavaScript, VBScript, HTML Help, SGML, XML, RC, C/C++/Java, Java properties, IBM TM untranslated segments, Trados WorkBench documents, TradosTag TTX, Trados TagEditor BIF, Trados TagEditor TMX, GNU PO и POT files и файлами в кодах ASCII – plain text files.

Программные продукты.

Déjà Vu X Standard

Целевая группа: переводчики-фрилансеры. Описание: данная версия предназначена для переводчиков, желающих получить простой и удобный инструмент для работы. Позволяет создавать полноценные проекты, однако в ней отсутствуют некоторые автоматизированные функции, которые есть в версии Professional.

Déjà Vu Professional

Целевая группа: переводчики-фрилансеры. Описание: классическая версия программы для переводчиков со всеми автоматизированными функциями «Интеллектуального качества»: «Предперевод», «Автопоиск» по базам данных, «Автоподстановка» перевода по всем аналогичным сегментам на проекте, «Автопополнение» баз данных, а также «Автосборка» перевода из фрагментов, хранящихся в базах данных. Использование функций Лексикона позволяет создавать глоссарии на основе переводимых Вами проектов, опре-

делить частотность употребления терминов и использовать полученные глоссарии для контроля качества.

Déjà Vu Workgroup

Целевая группа: переводческие агентства и отделы переводов. Описание: мощный инструмент, который включает все функции версии Professional, плюс широкие возможности по организации коллективной работы, управлению проектами и интеграции. Обладает всеми функциями Déjà Vu Professional, но также позволяет вести коллективную работу через создание проектов-сателлитов для переводчиков-фрилансеров или редакторов с текстами на перевод, памятью переводов и терминологией, либо через задания на перевод во внешнем формате, с которым можно работать с помощью любого текстового редактора.

Déjà Vu X TeaM Server

Гибкое решение, которое дает возможность пользователям Déjà Vu X2 Workgroup делиться базами данных в режиме реального времени по всему миру. Данное решение обеспечивает непревзойденное качество, единообразие и производительность, плюс более низкие проектные затраты. Возможные схемы лицензирования могут снизить или совсем исключить ручное администрирование проектов.

Trados (trados.com) – система автоматизированного перевода, первоначально (с 1992 года) разработанная немецкой компанией Trados GmbH. Является одним из мировых лиде-

ров в классе систем Translation Memory.

Концепция Translation Memory предполагает выявление в переводимом тексте фрагментов, переводы которых уже имеются в базе данных переводов, и за счет этого сокращение объёма работы переводчика. Фрагменты, оставшиеся непереведёнными, передаются дальше для ручной обработки переводчику или системе машинного перевода (Machine Translation, MT). Переводчик на этом этапе может выделить вновь переведённые фрагменты и занести новые пары параллельных текстов на двух языках в базу данных. Такая схема наилучшим образом работает в случае однотипных текстов, где повторяемость словосочетаний достаточно высока, то есть в случае различного рода инструкций для пользователей, технических описаний и т. и.

Основные модули

Translator's Workbench – модуль работы с базами данных памяти переводов TM (создания, обслуживания, импорта, экспорта). Создание документов перевода и их редактирование производятся в отдельных модулях.

Панель Trados в Microsoft Word – модуль для перевода документов в Microsoft Word.

TagEditor – модуль для перевода документов в формате PowerPoint, Excel, HTML, XML и т. и.

Win Align – модуль для создания памяти переводов на основе ранее переведённых двуязычных текстов.

S-Tagger – модуль для перевода документов в формате

FrameMaker и InterLeaf.

T-Window – модуль для перевода текста из буфера обмена.

MultiTerm – модуль для ведения глоссариев.

ExtraTerm – модуль для автоматического поиска терминологических кандидатов в тексте и создания глоссариев на их основе.

OmegaT (www.omegat.org) – система автоматизированного перевода, поддерживающая память переводов, написана на языке Java.

Возможности продукта включают сегментацию исходного текста на основе регулярных выражений, использование точных (англ, exact) и неточных (англ, fuzzy) соответствий с уже переведёнными фрагментами, использование словарей, поиск контекстов в базах данных переводов и работу с ключевыми словами.

Начиная с версии 2.04 OmegaT также может переводить текущий абзац текста через Google Translate.

Для работы OmegaT требуется версия Java 1.4, которая доступна для ОС Linux, Mac OS X и Microsoft Windows, Windows NT. Может работать с OpenJDK.

OmegaT поддерживает разнообразные форматы исходных документов.

Локализация OmegaT

Пользовательский интерфейс и документация OmegaT переведены на 30 языков.

Переводчики-добровольцы могут перевести пользовательский интерфейс, краткое руководство «Быстрый старт» или всё руководство пользователя (либо все три компонента).

Все языковые файлы и переводы руководства пользователя включены в стандартную поставку OmegaT.

ABBYY Lingvo (ABBYY Lingvo) – семейство электронных словарей. Создано российской компанией ABBYY. 26 августа 2014 года вышла последняя версия хб (икс шесть). 12 августа 2010 года вышла версия для Mac OS X. Объём словарных статей составляет более 8,7 млн.

Lingvo в переводе с языка эсперанто означает «язык», о чём есть статьи в словарях ABBYY Lingvo (LingvoUniversal и LingvoComputer).

Многоязычная версия охватывает 15 языков – русский, украинский, английский, немецкий, французский, испанский, итальянский, турецкий, латинский, китайский, португальский, татарский, казахский, шведский, польский и финский. Также существует Европейская версия – 130 словарей на 7 языках и англо-русско-английский электронный словарь.

В ABBYY Lingvo нет функции полнотекстового перевода, но возможен пословный перевод текстов из буфера обмена. В некоторых словарях на английском, немецком и французском большинство слов озвучены профессиональными дикторами – носителями языка.

В состав программы входит обучающий модуль Lingvo Tutor, помогающий запоминать новые слова.

Помимо существующих 150 профессиональных словарей, результата лексикографической работы сотрудников компании ABBYY и авторитетных бумажных и электронных словарей существует обширная база бесплатных пользовательских словарей для программы. Словари предварительно проверяются и находятся в общем доступе на сайте ассоциации лексикографов Lingvo.

Apertium (apertium.org) – платформа машинного перевода, которая разрабатывается при финансировании со стороны правительств Испании и Каталонии в Университете Аликанте (Universität d'Alacant). Это свободное программное обеспечение, которое бесплатно издаётся разработчиками в соответствии с условиями GNU General Public License (переводят как Универсальная общественная лицензия GNU, Универсальная общедоступная лицензия GNU или Открытое лицензионное соглашение GNU).

Apertium является системой машинного перевода, которая использует конечные преобразователи для всех своих

лексических трансформаций, а также скрытые модели Маркова для выделения частей речи или устранения противоречий в категориях слов.

Apertium в частности применяется фондом Wikimedia для разработки средств перевода статей Википедии.

Ectaco (www.ectaco.com) – американская компания, занимающаяся разработкой, созданием и производством карманных электронных переводчиков и словарей, электронных книг, лингвистического программного обеспечения.

Компания основана в 1990 году в Нью-Йорке. За время существования Ectaco было создано 7 поколений карманных словарей, разговорников и переводчиков для 45 языков. Сейчас офисы компании разбросаны по всему миру – их 16 в разных странах, в том числе в Германии, Чехии, Польше, Канаде, Австралии, России, Китае и Украине. Сейчас в компании работает более 300 человек – лингвистов, программистов и менеджеров.

Технологии

Speech Recognition (SR) – разработана собственная система распознавания речи, доступная для 22 языков, которая может быть встроена в различные электронные устройства.

Text-To-Speech (TTS) – синтез речи, дословно – «из текста в речь». Представлен почти во всех девайсах и программах производства Ectaco. Позволяет услышать произношение любых слов, фраз, цифр или буков, написанных на по-

нятном для этой системы языке – в настоящее время таких языков 17.

Machine translation (MT) – компьютерный перевод текста с одного языка на другой, или машинный перевод. Реализован двусторонний перевод между английским и 13 другими языками для нескольких платформ – Windows, Windows Mobile for Pocket PC, Microsoft Smartphone и других.

Устройства

jetBook – устройства для чтения электронных книг.

iTRAVL – электронный карманный словарь и разговорник с возможностью распознавания речи и синтеза текста, а также со встроенным аудио разговорником. Позиционируется как девайс предназначенный для путешественников.

Partner – карманные электронные словари и переводчки. Около 140 разновидностей для 30 языков, с различными кобминациями опций и с разными сочетаниями словарей и языковх пар.

SpeechGuard – линейка карманых переводчкиков, предназначенных для работы с речью. Выпущены армейский, милицейский и медицинский варианты по заказу ВВС США.

Программное обеспечение

LingvoSoft – объединяет различные лингвистические программы для перевода текстов и обучению иностранным языкам. Среди программ LingvoSoft имеются говоря-

щие и не говорящие словари, разговорники, обучающие карточки, переводчики текстов и многое другое. Приложения LingvoSoft существуют в вариантах для разных платформ.

Google Переводчик (translate.google.com) – веб-сервис компании Google, предназначенный для автоматического перевода части текста или веб-страницы на другой язык. Для некоторых языков пользователям предлагаются варианты переводов, например, для технических терминов, которые должны быть в будущем включены в обновления системы перевода.

В отличие от других сервисов перевода, таких, как Babel Fish и AOL, которые используют технологию SYSTRAN (английский), Google, как и Translate.ru, использует собственное программное обеспечение. Видимо, используется самообучаемый алгоритм машинного перевода.

Возможности

Сервис включает в себя также перевод всей веб-страницы и даже одновременный поиск информации с переводом на другой язык. Для вебдизайнеров сотрудниками компании был разработан скрипт, который позволяет организовать перевод сайта на все доступные языки.

Google Переводчик, как и другие инструменты автоматического перевода, имеет свои ограничения. Этот инструмент может помочь читателю понять общий смысл содержания

текста на иностранном языке, он не предоставляет точных переводов. Постоянно ведётся работа над качеством перевода, разрабатываются переводы на другие языки.

PROMT (www.promt.ru) – российская компания, разработчик систем машинного перевода. В настоящий момент существуют десктопные переводчики PROMT для английского, немецкого, французского, испанского, итальянского, португальского и русского языков. Причем имеются пары не только с русским языком, но и с иностранными, например англо-испанский, англо-португальский, немецко-французский, испанско-немецкий и др. Офисы компании открыты в России, США и Германии. В серверных продуктах PROMT доступны также китайский, японский, арабский и другие языки. В последние годы PROMT поставляет наряду с различными вариантами своей системы автоматического перевода и сопутствующими специализированными словарями ещё и собственные решения на основе технологии памяти переводов.

Технологии PROMT не раз удостаивались наград различных конкурсов. Так, в конкурсе, который проводится ежегодно в рамках семинара по статистическому машинному переводу под эгидой Ассоциации компьютерной лингвистики (ACL), переводчик от PROMT признан лучшим для перевода с английского языка на русский в 2013 и в 2014.

Основные возможности

Перевод слов, словосочетаний и текстов, в том числе с помощью «горячих» клавиш.

Перевод выделенной области экрана с графическим текстом.

Перевод документов разных форматов: doc(x), xls(x), ppt(x), rtf, html, xml, txt, ttx, pdf (в том числе отсканированные), jpeg, png, tiff.

Использование, редактирование и создание специализированных словарей и профилей перевода.

Подключение баз Translation Memory и глоссариев.

Интеграция в офисные приложения, веб-браузеры, корпоративные порталы и сайты.

Основные продукты

Пользовательские продукты

Сервис Translate.ru

PROMT Home

PROMT Professional

PROMT Expert

Переводчик для детей Magic Gooddy

Мобильные приложения Translate.Ru и PROMT Offline для iOS, Android, Windows Phone 8

PROMT Агент для Mac

Серверные решения для компаний на Windows и Linux

PROMT Translation Server (PTS)
PROMT Cloud
Решения для разработчиков и интеграторов (API)
PROMT Translation Server Developer Edition
Translate.ru Api.

SmartCAT – система автоматизированного перевода, включающая память переводов, машинный перевод, управление глоссариями, встроенные словари ABBYY Lingvo, функцию совместной работы переводчиков над одним документом. Программа облачная, не требует установки на компьютер, для работы нужен только браузер. Мобильное приложение ещё на стадии разработки.

Поддерживаемые форматы

SmartCAT поддерживает разнообразные форматы исходных документов.

Локализация SmartCAT

Пользовательский интерфейс и документация SmartCAT доступны на русском и английском языках. Техническая поддержка 24/7 на русском и английском языках.

STAR Transit NXT

Transit NXT – программа, предназначенная для создания, просмотра и редактирования проектов для проведения переводов и других связанных с переводом и локализаций

операций. Выпускается компанией STAR AG. На данный момент существует только версия для платформы Windows. Первая версия программы была выпущена в 2008 году.

Функции

Transit NXT позволяет выполнять большое количество самых различных операций, связанных с переводом. Например пользователь может импортировать исходный файл, при этом вся ненужная для перевода информация либо временно защищается от редактирования, либо не импортируется (но восстанавливается при экспортировании в оригинальный формат), редактировать его, используя доступные готовые переводы (при импортировании возможен частичный предварительный автоматический перевод), сохранять в виде материала для дальнейших переводов (память переводов), отправлять проект с переводом другим переводчикам, экспортировать его и т. д.

Transit NXT взаимодействует с другими приложениями от STAR, прежде всего с TermStar (система для терминологии). TermStar позволяет создавать словари, которые используются во время перевода или подготовки проектов.

Большим преимуществом программы в сравнении с неавтоматизированным переводом является использование памяти перевода, таким образом обновленные документы не нужно переводить заново – уже переведенные части будут автоматически добавлены во время предварительного пере-

вода или во время обработки файлов (программа «на лету» находит похожие или заранее переведенные сегменты и подставляет их в нужное место). В отличие от машинного перевода, Transit NXT не переводит сам. Переводы и проверки производятся пользователем программы. Программа лишь упрощает и ускоряет процесс перевода и конвертирует исходные файлы в доступный для перевода формат (и, соответственно, позволяет возвращать перевод в оригинальный формат). Готовые переводы можно использовать при последующих переводах.

В зависимости от качества предварительного перевода, различают разные степени схожести оригинала и готового переведенного материала. В терминологии NXT такие сегменты носят следующие названия:

internal repetition – данный сегмент уже был переведен, поэтому он автоматически будет переведен, когда пользователь перейдет к нему

fuzzy match – данный сегмент лишь частично переведен, например, одно из слов отличается от фразы в материалах для предварительного перевода и переводчик должен его проверить или перевести

new match – такой сегмент слишком сильно отличается от переведенных, поэтому считается «новым», то есть его необходимо перевести

Пользователь может создавать свои собственные типы сегментов (например «проверенный», или «ждет подтверждения»). Программа позволяет фильтровать сегменты по указанным критериям.

Цвета

Цвета в Transit NXT играют определенную роль, так как каждый цвет связан с определенной тематикой. Это должно помогать пользователю быстрее определять область применения той или иной функции или окна.

Следующие цвета отвечают определенным темам:

красный – всё, что связано с переводом

зеленый – всё, что связано с оригинальным текстом

синий – всё, что связано с тегами и форматированием

желтый – всё, что связано с терминологией

Файлы программы

Основные файлы программы имеют расширение PPF и TPF. PPF содержит в себе целый проект (включая исходные файлы, пары для переводов, материал с предыдущими переводами и т. д.). TPF включает только готовый перевод (языковая пара).

Приложение 3

Рассказ для машинного и автоматизированного перевода текста А.П. Чехов

«За двумя зайцами погонишься, ни одного не поймаешь»

Пробило 12 часов дня, и майор Щелколобов, обладатель тысячи десятин земли и молоденькой жены, высунул свою плешивую голову из-под ситцевого одеяла и громко выругался. Вчера, проходя мимо беседки, он слышал, как молодая жена его, майорша Каролина Карловна, более чем милостиво беседовала со своим приезжим кузеном, называла своего супруга, майора Щелколобова, бараном и с женским легкомыслием доказывала, что она своего мужа не любила, не любит и любить не будет за его, Щелколобова, тупоумие, мужицкие манеры и наклонность к умопомешательству и хроническому пьянству. Такое отношение жены поразило, возмутило и привело в сильнейшее негодование майора. Он не спал целую ночь и целое утро. В голове у него кипела непривычная работа, лицо горело и было краснее вареного рака; кулаки судорожно сжимались, а в груди происходила такая возня и стукотня, какой майор и под Карсом не видал и не слыхал.[1] Выглянув из-под одеяла на свет божий и выругавшись, он спрыгнул с кровати и, потрясая кулаками, зашагал

по комнате.

– Эй, болваны! – крикнул он.

Затрещала дверь, и пред лицо майора предстал его камердинер, куафер и поломойка Пантелей, в одежонке с барского плеча и с щенком под мышкой. Он упёрся о косяк двери и почтительно замигал глазами.

– Послушай, Пантелей, – начал майор, – я хочу с тобой поговорить по-человечески, как с человеком, откровенно. Стой ровней! Выпусти из кулака мух! Вот так! Будешь ли ты отвечать мне откровенно, от глубины души, или нет?

– Буду-с.

– Не смотри на меня с таким удивлением. На господ нельзя смотреть с удивлением. Закрой рот! Какой же ты бык, братец! Не знаешь, как нужно вести себя в моём присутствии. Отвечай мне прямо, без запинки! Колотишь ли ты свою жену или нет?

Пантелей закрыл рот рукою и преглупо ухмыльнулся.

– Кажинный вторник, ваше в<ысокоблагороди>е! – пробормотал он и захихикал.

– Очень хорошо. Чего ты смеёшься? Над этим шутить нельзя! Закрой рот! Не чешись при мне: я этого не люблю. (Майор подумал.)

Я полагаю, братец, что не одни только мужики наказывают своих жен. Как ты думаешь относительно этого?

– Не одни, ваше в-е!

– Пример!

– В городе есть судья Пётр Иваныч… Изволите знать? Я у них годов десять тому назад в дворниках состоял. Славный барин, в одно слово, то есть., а как подвыпимши, то бережись. Бывало, как придут подвыпимши, то и начнут кулачищем в бок барыню подсаживать. Штоб мне провалиться на ентом самом месте, коли не верите! Да и меня за конпанию ни с того ни с сего в бок, бывало, саданут. Бьют барыню да и говорят: «Ты, говорят, дура, меня не любишь, так я тебя, говорят, за это убить желаю и твоей жисти предел положить…»

– Ну, а она что?

– Простите, говорит.

– Ну? Ей-богу? Да это отлично!

И майор от удовольствия потёр себе руки.

– Истинная правда-с, ваше в-е! Да как и не бить, ваше в-е? Вот, например, моя… Как не побить! Гармонийку ногой раздавила да барские пирожки поела… Нешто это возможно? Гм!..

– Да ты, болван, не рассуждай! Чего рассуждаешь? Ведь умного ничего не сумеешь сказать? Не берись не за своё дело! Что барыня делает?

– Спят.

– Ну, что будет, то будет! Поди, скажи Марье, чтобы разбудила барыню и просила её ко мне… Постой!.. Как на твой взгляд? Я похож на мужика?

– Зачем вам походить, ваше в-е? Откудова это видно, штоб барин на мужика похож был? И вовсе нет!

Пантелей пожал плечами, дверь опять затрещала, и он вышел, а майор с озабоченной миной на лице начал умываться и одеваться.

– Душенька! – сказал одевшийся майор самым что ни на есть разъехидственным тоном вошедшей к нему хорошенькой двадцатилетней майорше, – не можешь ли ты уделить мне часок из твоего столь полезного для нас времени?

– С удовольствием, мой друг! – ответила майорша и подставила свой лоб к губам майора.

– Я, душенька, хочу погулять, по озеру покататься…. Не можешь ли ты из своей прелестной особы составить мне приятнейшую компанию?

– А не жарко ли будет? Впрочем, изволь, папочка, я с удовольствием. Ты будешь грести, а я рулём править. Не взять ли нам с собой закусок? Я ужасно есть хочу…

– *Я* уже взял закуску, – ответил майор и ощупал в своём кармане плётку

Через полчаса после этого разговора майор и майорша плыли на лодке к средине озера. Майор потел над вёслами, а майорша управляла рулём. «Какова? Какова? Какова?» – бормотал майор, свирепо поглядывая на замечтавшуюся жену и горя от нетерпения. «Стой!» – забасил он, когда лодка достигла середины. Лодка остановилась. У майора побагровела физиономия и затряслись поджилки.

– Что с тобой, Аполлоша? – спросила майорша, с удивлением глядя на мужа.

– Так я, – забормотал он, – баааран? Так я… я… кто я? Так я тупоумен? Так ты меня не любила и любить не будешь? Так ты… я…

Майор зарычал, простёр вверх длани, потряс в воздухе плетью и в лодке… o témpora, o mores!.. [2] поднялась страшная возня, такая возня, какую не только описать, но и вообразить едва ли возможно. Произошло то, чего не в состоянии изобразить даже художник, побывавший в Италии и обладающий самым пылким воображением… Не успел майор Щелколобов почувствовать отсутствие растительности на голове своей, не успела майорша воспользоваться вырванной из рук супруга плетью, как перевернулась лодка и…

В это время на берегу озера прогуливался бывший ключник майора, а ныне волостной писарь Иван Павлович и, в ожидании того блаженного времени, когда деревенские молодухи выйдут на озеро купаться, посвистывал, покуривал и размышлял о цели своей прогулки. Вдруг он услышал раздирающий душу крик. В этом крике он узнал голос своих бывших господ. «Помогите!» – кричали майор и майорша. Писарь, не долго думая, сбросил с себя пиджак, брюки и сапоги, перекрестился трижды и поплыл на помощь к средине озера. Плавал он лучше, чем писал и разбирал писанное, а потому через какие-нибудь три минуты был уже возле погибавших. Иван Павлович подплыл к погибавшим и стал в тупик.

«Кого спасать? – подумал он – Вот черти!» Двоих спасать ему было совсем не под силу. Для него достаточно было и

одного. Он скорчил на лице своём гримасу, выражавшую величайшее недоумение, и начал хвататься то за майора, то за майоршу.

– Кто-нибудь один! – сказал он. – Обоих вас куда мне взять? Что я, кашалот, что ли?

– Ваня, голубчик, спаси меня, – пропищала дрожащая майорша, держась за фалду майора, – меня спаси! Если меня спасёшь, то я выйду за тебя замуж! Клянусь всем для меня святым! Ай, ай, я утопаю!

– Иван! Иван Павлович! По-рыцарски!., того! – забасил, захлебываясь, майор – Спаси, братец! Рубль на водку! Будь отцом-благодетелем, не дай погибнуть во цвете лет… Озолочу с ног до головы… Да ну же, спасай! Какой же ты, право… Женюсь на твоей сестре Марье… Ей-богу, женюсь! Она у тебя красавица. Майоршу не спасай, чёрт с ней! Не спасёшь меня – убью, жить не позволю!

У Ивана Павловича закружилась голова, и он чуть-чуть не пошёл ко дну. Оба обещания казались ему одинаково выгодными – одно другого лучше. Что выбирать? А время не терпит! «Спасу-ка обоих! – порешил он – С двоих получать лучше, чем с одного. Вот это так, ей-богу. Бог не выдаст, свинья не съест. Господи благослови!» Иван Павлович перекрестился, схватил под правую руку майоршу, а указательным пальцем той же руки за галстух майора и поплыл, кряхтя, к берегу. «Ногами болтайте!» – командовал он, гребя левой рукой и мечтая о своей блестящей будущности. «Бары-

ня – жена, майор – зять… Шик! Гуляй, Ваня! Вот когда пирожных наемся да дорогие цыгары курить будем! Слава тебе, господи!» Трудно было Ивану Павловичу тянуть одной рукой двойную ношу и плыть против ветра, но мысль о блестящей будущности поддержала его. Он, улыбаясь и хихикая от счастья, доставил майора и майоршу на сушу. Велика была его радость. Но, увидев майора и майоршу, дружно вцепившихся друг в друга, он… вдруг побледнел, ударил себя кулаком по лбу, зарыдал и не обратил внимания на девок, которые, вылезши из воды, густою толпой окружали майора и майоршу и с удивлением посматривали на храброго писаря.

На другой день Иван Павлович, по проискам майора, был удалён из волостного правления, а майорша изгнала из своих апартаментов Марью с приказом отправляться ей «к своему милому барину».

– О, люди, люди! – вслух произносил Иван Павлович, гуляя по берегу рокового пруда, – что же благодарностию вы именуете?

1880

Приложение 4

Рассказ для машинного и автоматизированного перевода текста O Henry «Aristocracy Versus Hash»

The snake reporter of The Rolling Stone was wandering up the avenue last night on his way home from the Y.M.C.A. rooms when he was approached by a gaunt, hungry-looking man with wild eyes and dishevelled hair. He accosted the reporter in a hollow, weak voice.

"'Can you tell me, Sir, where **I** can find in this town a family of scrubs?'

"T don't understand exactly.'

"'Let me tell you how it is,' said the stranger, inserting his forefinger in the reporter's buttonhole and badly damaging his chrysanthemum. **'I** am a representative from Soapstone County, and **I** and my family are houseless, homeless, and shelterless. We have not tasted food for over a week. **I** brought my family with me, as **I** have indigestion and could not get around much with the boys. Some days ago **I** started out to find a boarding house, as **I** cannot afford to put up at a hotel. **I** found a nice aristocratic-looking place, that suited me, and went in and asked for the proprietress. A very stately lady with a Roman nose came in the room. She had one hand laid across her stom – across her waist, and the other held a lace handkerchief. **I** told her **I** wanted board for myself and family, and she condescended to take us. **I**

asked for her terms, and she said $300 per week.

"T had two dollars in my pocket and **I** gave her that for a fine teapot that **I** broke when **I** fell over the table when she spoke.'

"'You appear surprised,' says she. `You will please remembah that I am the widow of Governor Riddle of Georgiah; my family is very highly connected; I give you board as a favah; I nevah considah money any equivalent for the advantage of my society, I-'

"'Well, **I** got out of there, and **I** went to some other places. The next lady was a cousin of General Mahone of Virginia, and wanted four dollars an hour for a back room with a pink motto and a Burnet granite bed in it. The next one was an aunt of Davy Crockett, and asked eight dollars a day for a room furnished in imitation of the Alamo, with prunes for breakfast and one hour's conversation with her for dinner. Another one said she was a descendant of Benedict Arnold on her father's side and Captain Kidd on the other.

"'She took more after Captain Kidd.

"'She only had one meal and prayers a day, and counted her society worth $100 a week.

"'I found nine widows of Supreme Judges, twelve relicts of Governors and Generals, and twenty-two ruins left by various happy Colonels, Professors, and Majors, who valued their aristocratic worth from $90 to $900 per week, with weak-kneed hash and dried apples on the side. I admire people of fine descent, but my stomach yearns for pork and beans instead of culture.

Am I not right?'

"'Your words,' said the reporter, 'convince me that you have uttered what you have said.'

"'Thanks. You see how it is. I am not wealthy; I have only my per diem and my perquisites, and I cannot afford to pay for high lineage and moldy ancestors. A little corned beef goes further with me than a coronet, and when I am cold a coat of arms does not warm me.'

"'I greatly fear, 'said the reporter, with a playful hiccough, 'that you have run against a high-toned town. Most all the first-class boarding houses here are run by ladies of the old Southern families, the very first in the land.'

"'I am now desperate,' said the Representative, as he chewed a tack awhile, thinking it was a clove. 'I want to find a boarding house where the proprietress was an orphan found in a livery stable, whose father was a dago from East Austin, and whose grandfather was never placed on the map. I want a scrubby, ornery, low-down, snuff-dipping, back-woodsy, piebald gang, who never heard of finger bowls or Ward McAllister, but who can get up a mess of hot cornbread and Irish stew at regular market quotations.'

"'Is there such a place in Austin?'

"The snake reporter sadly shook his head. 'I do not know,' he said, 'but I will shake you for the beer. '

"Ten minutes later the slate in the Blue Ruin saloon bore two additional characters: 10."

Приложение 5

Краткая характеристика наиболее популярных поисковых систем *Google, Яндекс, Рамблер, Yahoo, Bing, Baidu, Nigma*

Google (www.google.com) – крупнейшая поисковая система интернета, принадлежащая корпорации Google Inc. Первая по популярности система (77,05 %), обрабатывает 41 млрд 345 млн запросов в месяц (доля рынка 62,4 %), индексирует более 25 миллиардов веб-страниц (на закрытой конференции в начале мая 2014 представитель Google упомянул, что на данный момент проиндексировано 60 триллионов документов, и как можно заметить, в результате тестов, счетчик в поиске Google ограничен числом 25 270 000 000, также на это число при выдаче влияют фильтры, встроенные в алгоритм ранжирования выдачи). Поддерживает поиск в документах форматов PDF, RTF, PostScript, Microsoft Word, Microsoft Excel, Microsoft PowerPoint и других.

Сленг «Гуглить»

Из-за популярности поисковой системы в английском языке появился неологизм to google или to Google (аналог в русском компьютерном сленге – гуглить), использующийся для обозначения поиска информации в Интернете с помощью Google. Именно с таким определением глагол занесён в наиболее авторитетные словари английского языка – Оксфордский словарь английского языка и Merriam-Webster, хо-

тя в других источниках приводятся примеры его использования для обозначения поиска вообще чего-либо в Интернете.

Первым, кто использовал слово как глагол, был сам Лэрри Пэйдж, 8 июля 1998 года подписавший одно из своих сообщений для списка рассылки: «Have fun and keep googling!». Американское диалектическое сообщество назвало глагол «to google» словом десятилетия.

Опасаясь возможной утраты товарного знака, Google не одобряет использование глагола google, особенно когда подразумевается поиск в Интернете вообще. Например, 23 февраля 2003 года компания направила письмо «прекратить и воздерживаться» (англ, cease and desist) Полу МакФедрису, основателю Word Spy – сайта, отслеживающего неологизмы. Также, в своей статье в «Вашингтон пост», Фрэнк Арэнс обсуждал письмо, полученное от юристов Google, иллюстрирующее «правильное» и «неправильное» употребление глагола google. В ответе на эту статью лексикографы словаря Merriam-Webster заметили, что записали глагол to google со строчной буквы, но для обозначения поисковой системы Google употребили заглавную букву (англ, to use the Google search engine to seek online information – пользоваться Google для поиска информации в Интернете), впрочем, редакторы оксфордского словаря не стали сохранять обе «версии» для истории. В 2006 году Google выпустил публичное заявление с требованием «использовать слова, образованные от Google, только когда речь идет о Google Inc. или его серви-

сах».

Яндекс (**www.yandex.ru**) – российская ИТ-компания, владеющая одноимённой системой поиска в Сети и интернет-порталом. Поисковая система «Яндекс» является четвёртой среди поисковых систем мира по количеству обработанных поисковых запросов (свыше 6,3 млрд в месяц на начало 2014 года). По состоянию на 5 июля 2015 года, согласно рейтингу Alexa.com, сайт yandex.ru по популярности занимает 19-е место в мире и первое место в России.

Поисковая система Yandex.ru была официально анонсирована 23 сентября 1997 года, и первое время развивалась в рамках компании CompTek International. Как отдельная компания «Яндекс» образовалась в 2000 году. В мае 2011 года Яндекс провёл первичное размещение акций, заработав на этом больше, чем какая-либо из Интернет-компаний со времён IPO поисковика Google в 2004 году.

Приоритетным направлением компании является разработка поискового механизма, но за годы работы «Яндекс» стал мультипорталом. В 2013 году «Яндекс» предоставляет более 50 сервисов. Некоторые из них – Яндекс. Поиск, Яндекс. Карты, Яндекс. Маркет, Поиск по блогам, Яндекс. Пробки – доминируют на рынке.

Бренд

Поисковый продукт «Яндекс» появился в 1993 году. Название системы – Яндекс, Япбех, – придумали вместе Арка-

дий Волож и Илья Сегалович.

Есть несколько вариантов происхождения названия:

Слово «Яндекс», или латиницей «Yandex», расшифровывается как Yet another indexer (англ, ещё один индексатор; очередной индексатор). Затем Волож заменил «Ya» на «Я» (сделав слово «Япбех») с целью подчеркнуть российское происхождение бренда.

Слово «Япбех» получилось в результате замены первой буквы в слове «Index».

Слово «Яндекс» расшифровывается как «Языковой индекс».

По трактовке Артемия Лебедева, название поисковика созвучно «Яньдекс», где янь – мужское начало.

Яндекс как наименование поисковых и иных продуктов на письме не должно выделяться кавычками. Подобное написание идёт от изначально гибридного названия Япбех, на которое не могли распространяться правила русского языка. Кроме того, за годы существования поисковика его имя стало нарицательным и позволяет употреблять его с маленькой буквы – яндекс как синоним поиска (ср. ксерокс, мерседес, браунинг и т. д.). Напротив, в значении юридического лица – ООО «Яндекс» – слово должно заключаться в кавычки как название организации.

Рамблер (www.rambler.ru) – популярный премиальный медийносервисный интернет-портал. Такое же название но-

сила поисковая система «Рамблер-Поиск», существовавшая в 1996–2011 годах.

«Рамблер» стоял у истоков российского интернета. Появившись в 1996 году, он быстро завоевал огромную популярность и оставался ведущим игроком на рынке поиска России вплоть до 2001 года. «Рамблер» запустил первый в рунете рейтинг-классификатор (Rambler Top 100), первый интернет-портал, первым среди отечественных интернет-компаний вышел на биржу.

В 2012 году философия портала была полностью переосмыслена – «Рамблер» стал медиапорталом персонализированных новостей.

Состоит в группе компаний Rambler&Co, образованной в мае 2013 года в результате объединения активов «Афиши-Рамблер» (ранее входила в холдинг «ПрофМедиа» Владимира Потанина) и SUP Media Александра Мамута.

«Рамблер» четырежды завоёвывал «Премию Рунета».

По данным на июль 2013 года, «Рамблер» занимал 11-е место по популярности среди сайтов России (по другим данным – 9-е).

Сегодняшний «Рамблер» вовсе не «странник» (от англ. Rambler – «странник», «бродяга»), блуждающий по сети в поисках ответов. Это индивидуальная картина дня и помощь в главных аспектах жизни. На «Рамблере» можно прочитать важные новости, разобраться в сфере финансов, недвижимости и авто, отправиться в путешествие, посмотреть попу-

лярные видео, купить билеты в кино или театр, собрать ребенка в садик и школу, познакомиться, узнать точный прогноз погоды и весело провести выходные. «Рамблер» – портал, которому доверяют.

Месячная аудитория «Рамблера» в 2015 году составляет 26 миллионов человек.

Yahoo (www.yahoo.com) – американская компания, владеющая второй по популярности (7,57 %) в мире поисковой системой (при этом в США и Канаде в соответствии с соглашением с Майкрософт от 2009 года и по состоянию на 2012 год поиск на сайте Yahoo! осуществляется поисковой машиной Bing) и предоставляющая ряд сервисов, объединённых интернет-порталом Yahoo! Directory; портал включает в себя популярный сервис электронной почты Yahoo! Mail, один из старейших и наиболее популярных в Интернете.

Согласно статистике Alexa Internet, в феврале-апреле 2012 г. Yahoo! – четвёртый по посещаемости веб-сайт в сети Интернет, и примерно 28 % посещений состоят из просмотра только одной страницы.

Bing (www.bing.com) (рус. Бинг) – поисковая система, разработанная международной корпорацией Microsoft. Bing был представлен генеральным директором Microsoft Стивом Балмером. Ранее имела следующие наименования и адреса:

MSN Search (http://search.msn.com/) – с момента появле-

ния в 1998 году и до 11 сентября 2006 года;

Windows Live Search (http://search.live.com/) – до 21 марта 2007 года;

Live Search (http://www.live.com/) – до 1 июня 2009 года.

Кроме того, с октября 2006 до января 2009 года действовал сайт Ms. Dewey (www.msdewey.com), а с августа 2007 до 30 июня 2009 года – Tafiti (tafiti.com), основанные на тех же технологиях Live Search, но имевшие иной, экспериментальный интерфейс.

В настоящее время сайт Bing занимает 2-е место в списке самых популярных поисковых сайтов по объёму трафика, в отличие от которых обладает рядом эксклюзивных возможностей, таких как просмотр результатов поиска на одной странице (вместо пролистывания многочисленных страниц результатов поиска), а также динамическое корректирование объёма информации, отображаемой для каждого результата поиска (например, только название, краткая или большая сводка).

Логотип

Сменил 4 логотипа. Нынешний – 5-й по счёту.

В 1994–1995 годах логотипом было слово «Yahoo» чёрного цвета и написано шрифтом Times New Roman.

В 1995–1997 годах логотипом было слово «Yahoo!» коричневого цвета и прыгающими буквами, шрифт поменялся на жирный.

В 1997–2009 годах логотипом было слово «Yahoo!» крас-

ного цвета и поменялся шрифт на обычный.

В 2009–2013 годах в логотипе слово «Yahoo!» стало фиолетового цвета.

С 2013 по настоящее время логотипом является слово «Yahoo!» фиолетового цвета и шрифт расширили.

С осени 2013 «Yahoo» начала проводить политику препятствования пользователям входа в свои почтовые ящики без предоставления им дополнительных персональных данных, что нарушает политику конфиденциальности.

С 7 августа 2013 года на протяжении 30 дней логотип менялся каждый день в рамках 30 days of change. Окончательный вариант логотипа был представлен 5 сентября

Baidu (www.baidu.com) (**кит**. **упр**. ##, **пиньинь**: Bàidù, Байду) – китайская компания, предоставляющая веб-сервисы, основным из которых является поисковая система с таким же названием – лидер среди китайских поисковых систем. По количеству обрабатываемых запросов поисковый сайт «Байду» стоит на 2 месте в мире (с долей в глобальном поиске 18.03 %). С запуском японской версии уверенно обогнал Bing.

В индексе Байду содержится свыше 740 млн. веб-страниц, 80 млн. изображений и 10 млн. медиафайлов.

Baidu также имеет онлайн-энциклопедию – Энциклопедию Байду, которая обогнала Китайскую Википедию.

В настоящее время выпускает (совместно с немецкими

производителями) Baidu Antivirus 2013 Beta. Антивирусная программа сочетает в себе движок Baidu Antivirus и облачный движок Baidu Cloud Security вместе с антивирусным движком Avira Antivirus для предоставления комплексной защиты от всех видов онлайн-угроз. Baidu Antivirus 2013 имеет статус экспериментальной (Beta) программы.

Энциклопедия Байду, или Байдупедия

20 апреля 2006 года ведущий китайский поисковик baidu.com заявил о запуске альтернативного проекта – «Байдупедии» (####). Уже через три недели она обогнала китайскую Википедию по числу статей. В настоящее время Байдупедия содержит более 3 500 000 статей. Правки, вносимые в Байдупедию, становятся видны не сразу, а проходят через модераторов и, предположительно, цензоров. В ней нет статей о Фалуньгун или независимости Тайваня. В то же время, в Байдупедии есть статья о Википедии, в которой Википедия описывается в нейтрально-положительном ключе. Длительное время в ней содержалась информация о её блокировке в КНР, и даже давались ссылки на зеркала Википедии, по которым можно было зайти на её главную страницу. Через некоторое время ссылки на зеркала были убраны, была оставлена лишь ссылка на официальную главную страницу http://zh.wikipedia.org/ (кит.), по которой, однако, из Китая зайти в Википедию до снятия блокировки было нельзя. Интерфейс сделан максимально удобным для пользователей из Китая.

Nigma (Нигма. РФ) (www.nigma.ru) – российская интеллектуальная метапоисковая система, первая кластеризующая поисковая система в Рунете. Проект создан при поддержке факультетов ВМК и психологии МГУ, а также Стэнфордского университета. Название «Nigma» (один из родов пауков семейства Dictynidae, en: Nigma) было выбрано по ассоциации со Всемирной паутиной.

На момент появления Нигма. РФ в проекте участвовало 2 человека, а именно Виктор Лавренко и Владимир Чернышов, которые познакомились на кафедре АСВК факультета вычислительной математики и кибернетики МГУ им М.В. Ломоносова в 2004-м году. В 2005-м году на сервисе появилась функция кластеризации. В 2007-м году Владимир Чернышов отправился в Стенфордский университет, где под руководством научного руководителя основателей Google разрабатывал алгоритмы для Нигмы.

Nigma осуществляет поиск как по своему индексу, так и по индексам Google, Yahoo, Bing, Яндекс, Rambler, AltaVista, Aport. По состоянию на 28 февраля 2009 года в суммарном индексе всех этих поисковых систем находилось более 7,16 млрд русскоязычных документов.

На основе введённого пользовательского запроса Нигма формирует список документов, разделённых на несколько множеств (кластеров). Пользователь может уточнить, в каком множестве продолжить поиск, тем самым улучшив ре-

левантность результатов поиска. Пользователь также может исключить ненужные ему множества сайтов, например, документы, пришедшие с интернет-магазинов (для них формируется специальный кластер).

Список кластеров выводится слева от списка результатов поиска. Для каждого кластера указывается образующая его фраза и количество документов в кластере. Пользователь может управлять кластерами при помощи специальных ссылок под списком кластеров.

Nigma поддерживает русскую морфологию. Используется морфологический модуль для русского языка собственной разработки.

Nigma позволяет производить простейшие арифметические преобразования и решать математические задачи, с учётом различных единиц измерения и распознаванием математических и физических констант. Также поддерживаются запросы на конвертацию валют, решение систем уравнений и построение графиков функций. Математическая система была запущена в октябре 2008 года.

В декабре 2008 года появилась поддержка запросов по неорганическим химическим реакциям, как по исходным, так и по конечным веществам реакции. Впоследствии были добавлены поиск химических реакций и поддержка органической химии.

На 2011 год система позволяет производить поиск по более чем 12 000 неорганических реакций. Вещества можно

задавать как в виде названий («хлорид натрия», «каменная соль»), так и в виде формул **(«NaCl»).**

Кщта эффективна для обзорного поиска при сборе информации. Она позволяет быстро найти открытые сведения из различных областей и проверить, не пропустили ли вы что-либо важное, используя другие поисковики.

Приложение 6

Правила формирования запросов в поисковых системах (на примере поисковой системы Яндекс)

Правила формирования запроса в поисковой системе Яндекс

1. Ключевые слова в запросе следует писать строчными (маленькими) буквами. Это обеспечит поиск всех ключевых слов, а не только тех, которые начинаются с прописной буквы.

2. При поиске учитываются все формы слова по правилам русского языка, независимо от формы слова в запросе. Например, если в запросе было указано слово «знаю», то условию поиска будут удовлетворять и слова «знаем», «знаете» и т. и.

3. Для поиска устойчивого словосочетания следует заключить слова в кавычки. Например, «фонема».

4. Для поиска по точной словоформе перед словом надо поставить восклицательный знак. Например, для поиска слова «сентябрь» в родительном падеже следует написать «¡сентября».

5. Для поиска внутри одного предложения слова в запросе разделяют пробелом или знаком &. Например, «приключенческий роман» или «приключенческий&роман». Несколько набранных в запросе слов, разделенных пробелами, означа-

ют, что все они должны входить в одно предложение искомого документа.

6. Для того, чтобы были отобраны только те документы, в которых встретилось каждое слово, указанное в запросе, необходимо поставить перед каждым из них знак плюс «+». Если вы, наоборот, хотите исключить какие-либо слова из результата поиска, поставьте перед этим словом минус «-». Знаки «+» и «-» надо писать через пробел от предыдущего и слитно со следующим словом. Например, по запросу «Волга – автомобиль» будут найдены документы, в которых есть слово «Волга» и нет слова «автомобиль».

7. При поиске синонимов или близких по значению слов между словами можно поставить вертикальную черту «|». Например, по запросу «ребенок | малыш | младенец» будут найдены документы с любым из этих слов.

8. Вместо одного слова в запросе можно подставить целое выражение. Для этого его надо взять в скобки. Например, «(ребенок | малыш | дети | младенец) +(уход | воспитание)».

9. Знак «~» (тильда) позволяет найти документы с предложением, содержащим первое слово, но не содержащим второе. Например, по запросу «книги – магазин» будут найдены все документы, содержащие слово «книги», рядом с которым (в пределах предложения) нет слова «магазин».

10. Если оператор повторяется один раз (например, & или ~), поиск производится в пределах предложения. Двойной оператор (&&,—) задает поиск в пределах документа. На-

пример, по запросу «дева— астрология» будут найдены документы со словом «дева», не относящиеся к астрологии.

11. Вернемся к примеру с аквариумными рыбками. После прочтения нескольких предлагаемых поисковой системой документов становится понятно, что поиск информации в Интернете следует начинать не с выбора аквариумных рыбок. Аквариум – сложная биологическая система, создание и поддержание которой требует специальных знаний, времени и серьезных капиталовложений.

На основании полученной информации человек, производящий поиск в Интернете, может кардинально изменить стратегию дальнейшего поиска, приняв решение изучить специальную литературу, относящуюся к исследуемому вопросу.

Для поиска литературы или полнотекстовых документов возможен следующий запрос: **«+(аквариум | аквариумист | аквариумистика) +начинающим +(советы | литература) +(статья | тезис | полнотекстовый) – (цена | магазин | доставка | каталог)».**

После обработки запроса поисковой машиной результат оказался весьма успешным. Уже первые ссылки приводят к искомым документам.

Теперь можно подытожить результаты поиска, сделать определенные выводы и принять решение о возможных действиях:

Прекратить дальнейший поиск, так как в силу различных причин содержание аквариума вам не под силу.

Прочитать предлагаемые статьи и приступить к устройству аквариума.