

ПОТЕНЦИАЛ АВТОМАТИЗИРОВАННЫХ СЛОВАРЕЙ

Абдурахманов Т, Шавкатов Ё.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Петрова Н.Е. – канд. филол. наук, доцент

Статья посвящена исследованию потенциала автоматизированных словарей. В работе называются преимущества автоматизированных словарей, их типы, обращается внимание на актуальные направления дальнейшего развития компьютерной лексикографии.

Современная лексикография невозможна без широкой компьютеризации. Сегодня традиционные методы создания словарей заменяются компьютерной обработкой лексикографических данных. Кроме очевидной экономии времени, это дает возможность поддерживать высокую точность лексикографической работы, позволяет оперативно актуализировать словари, повышает доступность лексикографических источников среди людей. Также это даёт возможность обновлять словари, быстро исправлять в них ошибки, которые ранее были не заметны авторам, и многое другое.

Потенциал автоматизированных словарей очень высок. Компьютер может сам быть создателем словаря. Компьютер быстро может найти ошибки, может обнаружить несогласованность различных отрывков текста словарной статьи (статей). Компьютер способен расположить словарный материал в любом порядке, выполнить упорядочение слов с высокой степенью точности. Учёные отмечают и следующие очевидные практические преимущества компьютера в создании словарей:

- исчерпывающая проверка различного рода ссылок и отсылочной информации;
- нормализация толкований и дефиниций;
- обеспечение последовательности среди различных кусочков информации и другое [1].

В свою очередь, компьютер – не человек, и сравниться с человеческими ассоциациями машина не может. Важно заметить, что компьютер "живёт" умом и чувствами лексикографа, который использует его для совершенствования и ускорения обработки данных. Лишь в общении, когда лексикограф и компьютер ведут диалог в режиме полного взаимопонимания, машина преобразуется, она становится незаменимой в процессе создания и использования словаря [2, с. 103].

Современная лексикография сегодня активно пользуется машинными банками данных, в частности, большими корпусами текстов (компьютерные картотеки), в которых компьютер по запросу осуществляет поиск нужных слов. Такие программы обработки текста позволяют в автоматическом режиме формировать словарные статьи, хранить словарную информацию и обрабатывать её. На этой основе формируется целое новое направление лингвистики и лексикографии – корпусная лексикография.

Создание машинных баз данных для лексикографических работ началось ещё в 60-е годы XX века и прошло два основных этапа. Корпусы первого поколения насчитывали до миллиона словоупотреблений и были репрезентативными, т.е. набирались по текстам различных жанров устной и письменной речи, каждый из которых был представлен определённой выборкой. Корпусы второго поколения создаются в результате сплошного обследования текстов и по объёму превышают корпусы первого поколения на полтора-два порядка [3].

Качество машинной базы лексикографических данных определяется не только объёмом и разнообразием источников, но и развитостью лексикографического сервиса, т.е. количеством услуг, которые она может предоставить. Уже в 1988 году было завершено новое 20-томное издание Большого оксфордского словаря. Весь корпус словаря хранится в компьютере в виде базы данных, в которой можно осуществлять поиск слов по большинству лексикографически существенных признаков и по любым комбинациям таких признаков. При этом скорость работы машины на несколько порядков превосходит скорость работы человека [4].

Исследователи компьютерной лексикографии различают онлайн-словари, программные словари, а также онлайн-программные словари. Такие словари разработаны как в России, так и в Беларуси. Онлайн-словари – вид электронных словарей, работающих на условиях интернет-доступа. К этому типу относятся Викисловарь, Lingvo, Мультигран и другие. Программные словари – лексикографические продукты, которые устанавливаются на персональный компьютер или сохраняются в виде специального электронного документа. Примером является проект «ArfaBel» – белорусский программный словарь для проверки орфографии белорусскоязычных текстов. Онлайн-программные словари – специализированные программы машинного перевода слов, словосочетаний, предложений, html-страниц, ini-файлов программ, программных интерфейсов. Примером является белорусский словарь «Белазар». К его основным возможностям относятся: перевод текста, автоматизация процесса перевода, поддержка альтернативной графики, автоматическое переключение раскладки клавиатуры и орфографический анализатор белорусского текста [5].

На современном этапе создано множество электронных лексикографических изданий на материале разных языков, различных по содержанию, объёму и особенностям описания языковых единиц. Вместе с тем многие вопросы являются актуальными и сегодня. К примеру, в связи со стремительным развитием языков словари быстро устаревают. Основная задача современной лексикографии – быстрее реагировать на новации в языках. Также не хватает электронных терминологических словарей, а существующие переводные электронные словари не всегда корректно осуществляют работу. Также не хватает словарей смешанного типа, которые соединяли в себе, к примеру, и перевод, и толкование слова, и подбор синонима. Нам кажется, что это было бы очень востребовано. Всё это как раз и позволит выполнить компьютерная лексикография.

Таким образом, в настоящее время автоматизированные словари имеют очень высокий потенциал, они бывают разных типов и популярны среди пользователей различного возраста. Однако язык и наука быстро развиваются, необходимо создание новых специальных технологий, позволяющих оптимизировать работу электронных словарей, чтобы ещё более эффективно использовать предоставляемые компьютером возможности в лексикографии. Также необходимо создавать более универсальные словари, которые совместят в себе ещё большее количество типов.

Список использованных источников:

1. Фефелова, А. Электронные словари. Их классификация [Электронный ресурс]. – Режим доступа: <https://www.stud24.ru/literature/elektronnye-slovari-ih-klassifikaciya/374779-1186647-page1.html>. – Дата доступа: 08.03.2021.
2. Дубчинский, В.В. Теоретическая и практическая лексикография / В.В. Дубчинский. – Харьков, 1998. – 160 с.
3. Роль электронных словарей в современной лексикографии [Электронный ресурс]. – Режим доступа: <https://pandia.ru/text/78/226/14233.php>. – Дата доступа: 21.03.2021.
4. Оксфордский словарь английского языка [Электронный ресурс]. – Режим доступа: https://ru.nipponkaigi.net/wiki/Oxford_English_Dictionary. – Дата доступа: 21.03.2021.
5. Дятко, Д.В. Белорусская лексикография: теория, история, современное состояние / Д.В. Дятко [Электронный ресурс]. – Режим доступа: <http://elib.bspu.by/bitstream/doc/30645/3/%D0%91%D0%B5%D0%BB%D0%BE%D1%80%D1%83%D1%81%D1%81%D0%BA%D0%B0%D1%8F.pdf>. – Дата доступа: 28.03.2021.