

А. Аргюх

О СОЗДАНИИ ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ ДЛЯ ПЕРЕВОДНЫХ ЭЛЕКТРОННЫХ СЛОВАРЕЙ

Современная лингвистика характеризуется направленностью многих исследований на решение прикладных задач. Именно поэтому актуальной задачей становится создание междисциплинарных методик, позволяющих осуществлять аналитическую обработку информации и поиск способов ее структурирования для эффективного хранения и обработки. Оптимальными в этой связи представляются технологии баз данных (БД), которые нами изучены на примере переводных электронных словарей.

В настоящий период развития данной технологии в сфере лингвистики можно говорить о двух основных типах баз данных: 1) *полнотекстовые*, которые представляют собой документальные БД; 2) БД, которые являются *информационными системами фактографического типа* и содержат структурированную информацию о лингвистических единицах разного рода.

Технология БД используется в процессе создания как традиционных, так и электронных словарей, самые популярные из которых мы исследовали.

Электронный словарь «МультиЛекс»

Стратегия структурирования словаря достаточно проста. Благодаря ей создается цифровая копия известных книжных изданий. За основу словаря «МультиЛекс» взят «Новый большой англо-русский словарь» А. Д. Апресяна. При вводе слова система показывает список словарей, в которых найдены переводы этого слова, и отображает варианты перевода из первого по списку словаря. Например, для слова *surveillance* ‘слежка’ словарь Мюллера выдал два варианта перевода. В результате указанное слово найдено в 12 словарях. Для английского языка на сайте представлено 26 словарей (русско-английских и англо-русских).

Электронный словарь «Мультитран»

Словарная база создана путем сканирования, распознавания и переработки большого количества бумажных словарей и объединения полученных переводов слов в единую базу данных. Наиболее полно представлены англо-русско-английская, немецко-русско-немецкая и французско-русско-французская части словаря. Автором является Андрей Поминов, однако вклад в создание словаря может внести любой пользователь, поэтому словари могут содержать от 1 слова («Бионика») до 450 тыс. («Техника»). Для слова *surveillance* «Multitran» выдает 13 переводов и дополнительно сообщает, что это слово встречается в 982 фразах в 46 тематиках.

Таким образом, современные электронные словари не только значительно превосходят по объему книжные, но и находят искомое слово или словосочетание за несколько секунд, облегчая задачу переводчику и другим пользователям.