

**Н. В. Зыгмантович**

## АВТОМАТИЧЕСКОЕ СОЗДАНИЕ ОБРАТНОГО СЛОВАРЯ ТЕКСТА

С помощью компьютера создаются разные типы словарей: терминологические, частотные, обратные, словоуказатели, конкордансы, тезаурусы и т.д. Рассмотрим подробнее автоматическое формирование обратного словаря. Отличие обратного словаря от любого другого словарного издания состоит в следующем: в обычных словарях (двуязычных, толковых и др.) слова располагаются в алфавитном порядке, в обратном словаре используется так называемый алфавитный порядок по концу слова или обратный (инверсионный) алфавитный порядок. Из любых двух слов первым в такой словарь помещается то, у которого конечная буква ближе к началу алфавита. Например, в обратном словаре русского языка на первом месте стоят слова, оканчивающиеся на *-а*, потом на *-б*, *-в*, *-г* и т.д.

Родоначальниками обратных словарей считаются средневековые арабские классические словари XIII–XIV веков. В Европе в XVIII в. обратный алфавитный порядок слов использовался еще и при составлении словарей рифм (так называемых рифмовников). В середине 50-х г. XX в. Л. Успенский высказал мысль о том, что лингвистам был бы очень полезен словарь, в котором слова располагались по алфавиту не начальных, а конечных букв. Ученый писал: «Представьте себе, что я захочу узнать что-либо, связанное не с началами, а с окончаниями слов. Ну, положим, какое значение имеет в русском языке суффикс ‘-л-’ в словах среднего рода, вроде ‘зерка-л-о’? Или каких суффиксов ‘-чик’ в нем больше: тех ли, которые образуют слова, означающие профессию, род занятий (вроде ‘лет-чик’, ‘рез-чик’, ‘пулемет-чик’), или образующих уменьшительные имена (‘маль-чик’, ‘паль-чик’ и пр.). Мне может понадобиться и сведение, какой суффикс более употребителен: ‘-чик’ или ‘-ник’ (а может быть, ‘-тель’) в тех же словах, означающих род занятий (‘гранат-о-мет-чик’ или ‘подрыв-ник’?)». Л. Успенский имел в виду обратный словарь русского языка. Первый «Обратный словарь русского языка» был создан в 1974 г. и содержал около 125 тыс. слов.

В настоящее время создано большое количество обратных словарей для разных, как правило, флективных языков. Выделить из текста все слова и расположить их в алфавитном порядке, начиная от конца слова, вручную является достаточно трудоемкой задачей. Поэтому разработаны различные компьютерные программы, позволяющие создавать обратные словари автоматически. Наиболее удобной из них является Cugwin – среда для запуска Linux-приложений из-под Windows. Определенный набор команд данной среды позволяет формировать разные типы словарей: частотные, обратные, конкордансы.

Рассмотрим алгоритм построения обратного словаря русскоязычного текста программой Cugwin. В командной строке Cugwin набирается последовательность команд, разделенных специальным знаком «*—|,*» который называется *pip*.

```
cat introduction.txt | tr -d '.,;:"- ' | tr ' ' '\n' | rev | sort | rev | uniq >
rev_dict.txt
```

Каждая из этих команд выполняет следующие действия:

1. *cat introduction.txt* – загружает в память ПК файл;
2. *tr -d '.,;:"- '* – удаляет из текста знаки препинания;
3. *tr ' ' '\n'* – заменяет каждый пробел в тексте на *Enter*, то есть преобразует текст в словарь словоформ;
4. *rev* – поменяет в каждой словоформе словаря буквы справа–налево, например, *недавно – онваден*;
5. *sort* – отсортирует преобразованный таким образом словарь по алфавиту;
6. *rev* – поменяет в каждой словоформе словаря буквы слева–направо, то есть вернет ее в исходную форму, например, *онваден – недавно*;
7. *uniq* – удалит повторяющиеся слова;
8. *> rev\_dict.txt* – сохранит полученный обратный словарь на ПК с именем.

В результате получается обратный словарь, фрагмент которого представлен ниже:

<i>касающиеся</i>	<i>отличается</i>
<i>отличающиеся</i>	<i>изучается</i>
<i>имеющиеся</i>	<i>требуется</i>
<i>являющиеся</i>	<i>формируется</i>
<i>подвергался</i>	<i>игнорируется</i>
<i>попытался</i>	<i>представляется</i>
<i>осуществлялся</i>	<i>является</i>
<i>передающемся</i>	<i>проявляется</i>
<i>обратимся</i>	<i>определяется</i>
<i>сложившимся</i>	<i>объясняется</i>
<i>изучавшегося</i>	<i>становится</i>
<i>рассматривается</i>	<i>приводится</i>
<i>сказывается</i>	<i>сводится</i>
<i>утверждается</i>	<i>относится</i>
<i>отражается</i>	<i>кажутся</i>
<i>познается</i>	<i>останутся</i>
<i>считается</i>	<i>развиваются</i>

## **А. А. Макоўскі**

### **«ЛІТАСЦЬ» У КАНЦЭПТАСФЕРЫ АНГЛІЙСКОЙ І БЕЛАРУСКАЙ КУЛЬТУР**

Пытанне ўспрыняцця рэчаіснасці народам і яе адлюстравання ў моўнай свядомасці – адна з асноўных праблем кагнітыўнай лінгвістыкі і лінгвакультуралогіі. Для вывучэння ўзаемасувязі паміж аб’ектыўнай рэальнасцю і яе адлюстраваннем у мове неабходна было перагледзець шэраг асноўных