

В. С. Селезнева

РАЗРАБОТКА АЛГОРИТМА ДЛЯ АВТОМАТИЗИРОВАННОГО ПОИСКА КОМПОЗИТОВ

Методом сплошной выборки можно собрать искомый материал для анализа, при этом потратив значительное количество временного ресурса. Цифровые решения облегчат интеллектуальную деятельность человека, сокращая временные рамки на отведенные операции, в некоторых случаях подкрепляя достижения статистическими данными.

Основной фокус нашего внимания в данной работе смещен на рассмотрение вопросов разработки программного обеспечения для решения профессионально ориентированных задач. Исследование проходит в области пересечения лингвистики и информатики. В качестве материала для изучения выбраны композиты в текстах литературных и специализированных.

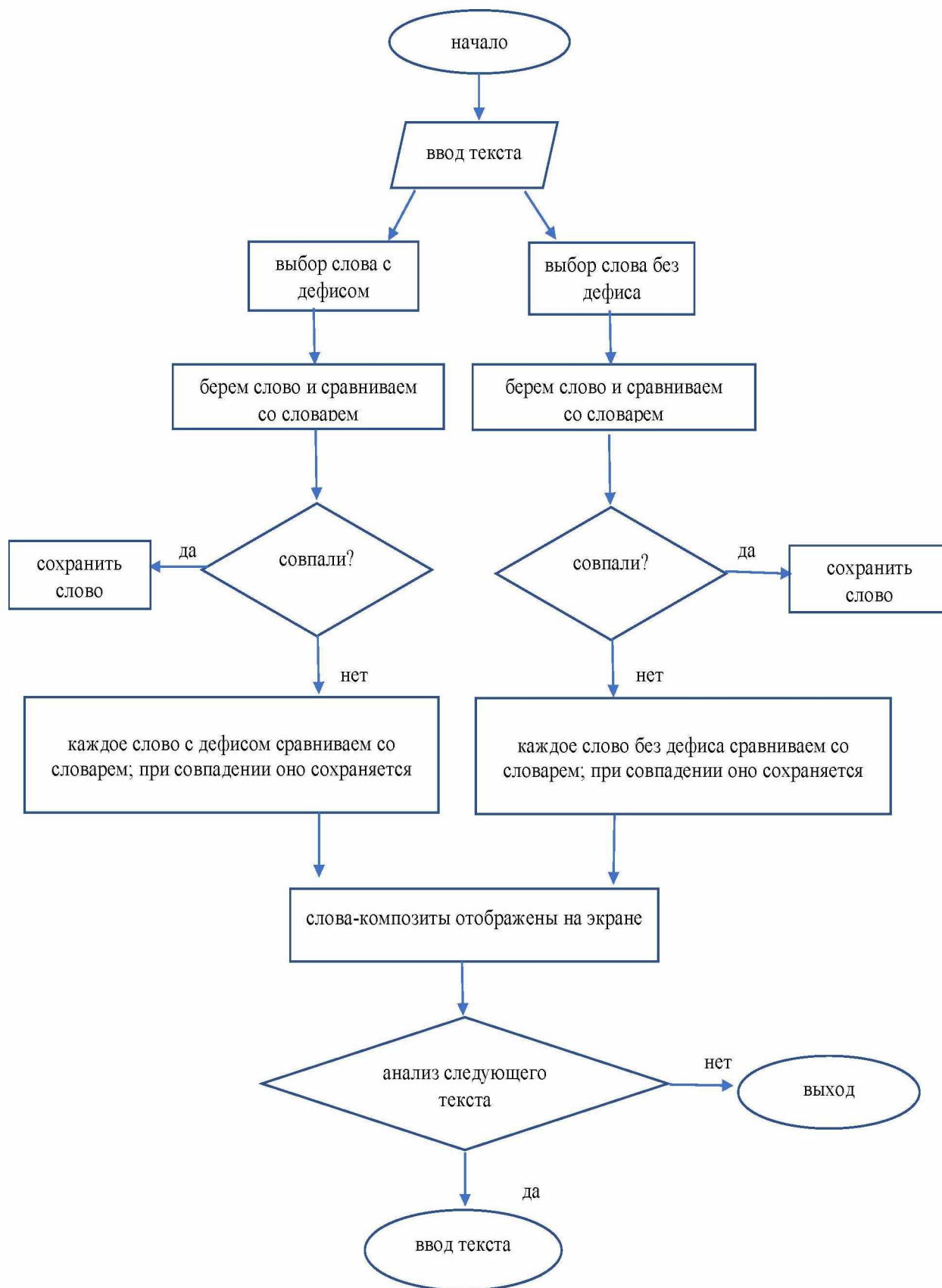
Перед началом проектирования программы необходимо провести подготовительную работу:

- выбрать «вручную» композиты для анализа и систематизировать в алфавитном порядке;
- занести в базу данных отрывки текстов, в которых были найдены композиты;
- занести в программу морфемный словарь. Для данной работы наиболее подходящим словарем является фундаментальный двухтомник А. Н. Тихонова «Словообразовательный словарь русского языка». Этот наиболее полный гнездовой словарь содержит более 145 тысяч слов. В нем описана словообразовательная структура производных слов: показано, на базе какого производящего и при помощи каких словообразовательных средств они образованы. В словаре в алфавитном порядке исходных слов представлено 12 621 гнездо.

В нашей работе предпринимается попытка разработать алгоритм автоматизированной идентификации композитов. Основная отличительная черта композитов будет делением на написание с дефисом / без дефиса. Из общего списка композитов, размещенных в алфавитном порядке, осуществляется поиск заданной словоформы композита или ее автономной корневой единицы в электронном тексте программы.

Основным примечанием поиска является распознавание композита в тексте в независимости от его полной идентичности с заданной единицей в поиске. Если поиск осуществляется по единичной корневой основе, программа находит все возможные композиты с данным корнем. Если ввести композит полностью, программа идентифицирует только его в тексте. При обнаружении искомой единицы можно осуществить переход к поиску следующих композитов.

Исходя из описанных действий можно представить следующий алгоритм к рассмотрению (рисунок).



Алгоритм для автоматизированного поиска композитов

Решая задачу автоматизированного поиска композитов, программа также может проводить анализ частотности употребления лексических единиц в составе разных композитов, а также нам предоставляется возможность для составления словообразовательных гнезд автоматизированным путем основываясь на систематизацию однокоренных слов из базы данных композитов и корпусного анализа электронных текстов.

Продолжением исследования может послужить усовершенствование разработанного алгоритма, использование его в условиях лингвистического программирования, а также поиск новых автоматизированных решений для выявления композитов в текстах. Данный алгоритм может быть универсальным, т. е. мог бы применяться на материале других текстов с дополнением или изменением некоторых параметров.