

## АЛГОРИТМ МАШИННОГО ПЕРЕВОДА

**Наргиза Бекмухамедова**

Национальный университет Узбекистана имени Мирзо Улугбека,  
Старший преподаватель

**Аннотация:** В данной статье речь идет об алгоритме машинного перевода. На примере русского языка осуществлен морфологический, синтаксический и семантический анализ текста. Алгоритм машинного перевода является по шаговым.

**Ключевые слова:** *алгоритм, компьютерные технологии, буфер, распознавание, перевод, анализ, семантика, морфология, синтактика, модуль, исходный текст*

Слово «алгоритм» сегодня является самым употребляемым. Это связано конечно с развитием компьютерных технологий. Данный термин первоначально использовался в информатике. Но, если посмотреть на это слово в широком смысле, алгоритм действий мы можем увидеть везде. В данной статье мы будем говорить об алгоритме машинного перевода, основанного на лингвистическом анализе. Но, давайте сначала вспомним определение данного термина. Алгоритм – это конечная совокупность точно заданных правил решения произвольного класса задач или набор инструкций, описывающий порядок действий исполнителя для решения некоторой задачи. Существуют следующие свойства алгоритма, такие как универсальность, т.е. (массовость), дискретность, конечность, результативность, выполнимость, т.е. (эффективность), детерминированность, т.е (определенность) и последовательность.

По шаговый алгоритм машинного перевода по В.Вороновичу:

Первый шаг-получение предложения исходного текста из файла или из буфера в памяти.

Второй шаг- задача разбиения текста на слова и предложения. Основную сложность представляют различные сокращения, инициалы, прямая речь, слова, написанные через дефис и т.д. Распознавание слов ведется с помощью специальных шаблонов. Данные шаблоны описывают различные буквенные, цифровые и буквенно-цифровые группы и символы пунктуации, которые затем будут выделяться в качестве отдельных слов. Например, как отдельные



слова будут выделены даты, записанные в их цифровом выражении, номера параграфов и подпараграфов, сокращений вместе с точками (по специальному словарю), а также слова, написанные через дефис в случае, если они распознаются специальным модулем словаря – модулем анализа сложных слов. Анализ и перевод таких слов осуществляется на основе специальных правил морфологических преобразований прилагательных. В результате анализа выделенных слов, некоторым словам (инициалам, сокращениям и т.п.) будут присвоены специальные маркеры, которые позволят разрешить многозначность при распознавании границ предложений. Также на этом этапе происходит нормализация слов с целью подготовки их для поиска по словарю.

Шаг третий - морфологический анализ. Решение данной задачи базируется на словаре исходного языка. В результате поиска по словарю каждому слову предложения приписывается множество лексико-грамматических классов: часть речи, падеж, число, род, категория и т.д., что позволяет в дальнейшем производить сравнение классов, основанное на определенных характеристиках (например, проверять согласование прилагательных и существительных). Процесс поиска слов по словарю предполагает, кроме поиска оригинального слова в случае, если оно не было найдено в словаре, поиск слов с удалением возможных префиксов. Для эффективного поиска префиксов используется древовидная структура, элементами которой являются буквы предлогов. Поиск останавливается либо когда нет дальнейшего перехода в дереве, либо когда найден предлог и слово без этого предлога существует в словаре. Кроме словаря предлогов, для каждого из языков существует таблица межязыкового соответствия, с помощью которой на этапе синтеза текста получается результирующее слово. На этапе распознавания классов производится также выделение словосочетаний, которые, согласно словарю, переводятся одним словом (словарь идиом): по барабану, зайти в тупик, kick the bucket (дословно – пнуть ведро, ‘сыграть в ящик’). Далее считается, что все такие словосочетания представляются одним словом. Это гарантирует правильность согласования и перевода словосочетания как единого целого.

Шаг четвёртый – синтаксический анализ. Сначала для каждого слова производится поиск главного слова, с которым оно должно быть согласовано в результате перевода. При этом не предполагается, что уже обязательно должна быть полностью снята многозначность. В процессе поиска главных

слов производится основное снятие многозначности. Построение синтаксического дерева производится путем последовательного распознавания заранее заданных лингвистических шаблонов и применения на их основе определенных операций. Основными операциями при распознавании шаблонов являются: а) проверка, является ли слово определенной частью речи с конкретными характеристиками (например, является ли слово существительным в родительном падеже); б) проверка, является ли некоторое слово омонимом, т.е. может ли оно принадлежать к разным частям речи (например, жаркое, love); в) проверка согласования двух слов (полное согласование – прилагательное-существительное, согласование по падежу – существительное-существительное и т.д.): красивая девушка, Татьяна Васильева; г) получение семантических характеристик управления предлогов и глаголов – каждый глагол и предлог требуют наличия соответствующих падежей у определяемых ими слов (управляют конкретным падежом). Набор этих падежей зависит от смысла этих предлогов и глаголов (например, находиться управляет только предложным падежом, а писать управляет дательным, винительным, творительным и предложным падежами).

Пятый шаг - семантический анализ исходного текста. Основная задача данного этапа – разрешение многозначности на основе полученного дерева зависимостей. Для этого первоначально производится разрешение многозначности базовых слов. Как показали исследования, целесообразным является попарное согласование рядом стоящих базовых слов в порядке, обратном положению слов в предложении. После того, как всем базовым словам поставлен в соответствие один лексико-грамматический класс, производится “досогласование” зависимых от них слов. Параметры выбора лексико-грамматических классов зависимых слов выбираются согласно типу зависимости и лексико-грамматическому классу главного слова. Пример правила для снятия многозначности: Обработка омонимии «прилагательное-причастие». Если перед омонимом прилагательное-причастие стоит запятая, то у рассматриваемого слова удаляются словоформы прилагательного. В противном случае удаляются словоформы причастия: слышался странный звук, свистящий словно ветер; с – звук свистящий.

Далее шаг шестой - осуществление перевода построенного дерева. Процесс перевода состоит из следующих шагов: а) производится пословный перевод базовых слов дерева (кроме глаголов) зависимостей с сохранением

оригинального лексико-грамматического класса (либо наиболее близкого к нему по своим характеристикам класса); б) для глаголов из списка базовых слов, имеющих в качестве исходных характеристик признак рода, перевод осуществляется в множество глаголов одной парадигмы с признаками единственного числа и рода; для остальных глагольных форм – перевод осуществляется с сохранением исходных лексико-грамматических характеристик; в) для зависимых слов результатом перевода является множество слов, которое определяется на основе типа зависимости и лексико-грамматических классов главных слов (например, результатом перевода прилагательных является вся парадигма, большинство местоимений переводятся с сохранением исходных падежей) – окончательные лексические характеристики определяются на этапе синтеза. Также на этом шаге производится анализ слов, результатом перевода которых является словосочетание. В результате перевода словосочетание должно быть согласованным.

Следующий шаг- осуществление согласования переведенного дерева. В результате перевода получается частично согласованное дерево зависимостей. Для получения полного согласования достаточно использовать процедуру, аналогичную процедуре окончательного разрешения многозначности, применяемую на этапе построения дерева. Далее на основе дерева производится построение результирующего предложения. Для этого для каждого слова в словаре результирующего языка производится поиск с целью получения конкретной словоформы, соответствующей зафиксированному лексико-грамматическому классу.

И последний шаг- запись переведённого предложения в файл или в буфер. Таким образом, осуществляется по шаговый алгоритм машинного перевода на всех языках.

### **Литература:**

1. Воронович В.В. Машинный перевод. – Минск, 2013.
2. Шаляпина З.М. Текст как объект автоматического перевода. – М., 1988.
3. Кулагина О. С. Исследования по машинному переводу. – М., 1979.
4. Семенов А.Л. Современные информационные технологии и перевод. – М., 2008.