III МЕЖДУНАРОДНАЯ НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ

# НАУКА И ОБРАЗОВАНИЕ В СОВРЕМЕННОМ МИРЕ: ВЫЗОВЫ XXI ВЕКА

**Нур-Султан (Астана), 10-12 июля 2019**

# ПРИНЦИПЫ СОЗДАНИЯ МОРФОЛОГИЧЕСКОГО МОДУЛЯ ДЛЯ ПРОГРАММЫ АВТОМАТИЧЕСКОГО АНАЛИЗА УЗБЕКСКОГО ТЕКСТА

**Абжалова Манзура Абдурашетовна**
Препадователь Навоийского государственного
горного института
Навои, Узбекистан

**Аннотация.** Автоматический морфологический анализ — процедура, позволяющая из формы слова извлечь информацию о его грамматических признаках. В данном статье приводится пути для разработки модулей автоматического морфологического анализа.

**Ключивые слова:** морфологический анализ,модуль, морфологический параметр,автоматический анализ, словоформа.

**Abstract:**The automatic morphological analysis — the procedure allowing from a form of a word to take information on its grammatical signs. In given to article it is brought ways for development of modules of the automatic morphological analysis.

**Key words:** morphological analysis, module, morphological parameter, automatic analysis, word form.

В русском энциклопедическом словаре дается полное описание автоматического анализа текста: Автоматический анализ текста (АА), операция, которая заключается в том, что из данного текста на естественном языке извлекается содержащаяся в этом тексте грамматическая и семантическая информация, выполняемая по некоторому алгоритму в соответствии с заранее разработанным описанием данного языка. В автоматическом анализе этап морфологического анализа имеет важное значение. *Автоматический морфологический анализ* – обеспечивает определение нормальной формы, от которой была образована данная словоформаи набора параметров, приписанных данной словоформе. Это делается для того, чтобы ориентироваться в дальнейшем только на нормальную форму, а не на все словоформы, использовать параметры, например, для проверки согласования слов[62].

Цель морфологического анализа, это определить:

1) основную форму слова;

2) полную морфологическую характеристику слова (морфологические параметры слова).

Автоматический морфологический анализ помогает избежать лингвисту, обрабатывать все словоформу одного слова.

Узбекский язык является агглютинативным. В нем часть информации об употреблении слова добавляется в конец слова в виде аффиксов, то есть словоформа образуется от основы, которому аффиксы связываются по порядку: основа+словообразующийаффикс+формаобразующийаффикс+словоизменяющий аффикс. В основном, основа остается так она есть. Например, ***иш****+чи+лар+имиз* «наше работники».

Морфологические параметры – это пара: имя параметра, значения параметра[63]. Именем параметра может служить число, падеж, время, склонение, наклонения и другие

[62]Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютернаялингвистика (учеб. пособие) — М., 2011. — стр.106

[63]Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютернаялингвистика (учеб. пособие) — М., 2011. — стр.110.

признаки слов, принятые данным языке. Значение параметра – это конкретное значение, которое может принимать данный элемент языка. Например, время может быть прошедшим, настоящим и будущим; степени сравнения прилагательных может быть сравнительная степень, превосходная степень; число – единственным и множественным и т.д.

Есть такие случае что значение параметра определить невозможно или в этом нет необходимости. Например, в узбекском языке существуют слова, которые имеют только форму множественного числа: *армия, халқ* (народ), *қўшин* (войско), *тўда* (шайка, стадо) и т.д. И ещё, если некоторым словам прибавляется аффикс множественного числа ***-лар***, то они обозначают разновидность сорта, уважение, ирония.Например: *унлар (муки), севинчлар (радости), шакарлар (сахары).*

Самые большие возможности и высокое качество анализа текстов можно получить, проведя его полный анализ. Однако сложности, возникающие при создании подобного анализа таковы, что на практике до сих пор не реализованы все теоретические положения, разработанные на данный момент. Основными проблемами здесь являются сложность работать над огромным количеством комбинацией аффиксов. Так, например, в одно слово можно сказать *китоб+лар+им+да+ги+лар* «то (вин. падеж), что лежит на моих книгах». За счет добавления аффиксов у одного существительного может появиться несколько тысяч словоформ, хранить которые в морфологическом словаре также будет просто невозможно, так как аффиксы будут добавляться после любой имеющейся словоформы, не содержащей аффиксов, хотя и в строго определенном порядке. В связи с этим было бы проще ввести ряд дополнительных параметров и «откусив» соответствующий аффикс, добавить оставшейся словоформе параметр с заданным для данного аффикса значением. Для того чтобы решить проблему, вначале намечается анализировать текстов на узбекском языке в стиле деловой, публицистичной и научной. В этих стилях узбекского языка формообразование происходит прибавлением специальных аффиксов и в основном один аффикс имеет лишь одно значение: ***kitoblarim***-*lar* – аффикс множественного числа, -*im* – аффикс принадлежности.

Разрабатывая лингвистические модули морфологического анализа создаётся группы, которым внесены параметры от которых характеризуется словоформа: основа слова (от которого была образована), часть речи основы, морфологические параметры.

Есть и другой подход. Тут нам понадобится толковый словарь узбекского языка. Из него берём все литературные слова. Для того чтобы в процессе лингвистической обработки не уделять время диалектичным и старинным словам (они будут внесены по отдельным группам, чтобы в дальнейшим работать над этими словами). После этого определяем их части речи. В ряде этого смотря значению слов мы должны распределить их по группам. Например: слова *китоб (книга), олма (яблоко), чинор (платан), торт (торт)* существительные; *китоб – учебное, олма – овощное, чинор – дерево, торт – сладость* и т.д.

Эта нужна для того, чтобы:

1) определив группы слов, мы просматриваем комбинаторы аффиксов, которые они могут связать в себя. Число комбинаторы аффиксов одного слово в узбекском языке может быть и больше три тысяч и менее ста. В таком случае, невозможна физически разработать словоформы каждому слову. Такой подход требует много времени и внимательность.

2) Создавая точные именные группы внутри части речи мы помогаем решать проблемы неологизмов[64]. После того как ЛП передастся своим пользователям, они сами тоже могут обновлять свои программу анализатора. То есть, если в языке появится новое

[64] Неологизм (от греч. neos – новый и logos - слово) – слово, значение слова или словосочетание, недавно появившиеся в языке. Из этого определения ясно, что понятие неологизма изменчиво во времени и относительно: неологизмом слово остается до тех пор, пока говорящие ощущают в нем новизну.

слово, пользователь через запросы эту слову вместит нужную группу по его значению. После этого программа анализатор будет считать правильным словоформ неологизма.

Рассмотрев то, над чем работает морфология, перейдем к тем методам, с помощью которых она реализуется. Различают два вида морфологических словарей: словарные и бессловарные. Словарная морфология предполагает наличие словаря, в котором каждой словоформе сопоставлены нормальная форма и набор параметров. Т.е. у нас хранится полный словарь слов, и мы не можем проанализировать слова, отсутствующие в словаре. Самым простым решением проблемы создания морфологического словаря является таблица, в которой в первой колонке будет записана словоформа, во второй – нормальная форма, а в третьей – набор параметров. Для морфологического анализа в таком словаре необходимо просто найти все соответствующие словоформы и выдать найденные результаты.

**Список литературы:**

1. Mark Hawthorne. The Computer in Literary Analysis: Using TACT with Students // Computers and the Humanities. Vol.28. N1. 1994. P.19-27

2. Большакова Е.И. и Носков А.А.. Программные средства анализа текста на основе лексико-синтаксических шаблонов языка LSPL. В Программные системы и инструменты: Тематический сборник, № 11 / Под ред. Королева Л.Н., страницы 71–73. М.: Изд. отдел факультета ВМиК МГУ; МАКС Пресс, 2010.

3. Леонтьева Н. Н. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие – М.: Академия, 2006.