

МОРФОЛОГИЧЕСКИЙ И СИНТАКСИЧЕСКИЙ АНАЛИЗАТОРЫ

В данной статье рассматриваются основные проблемы и способы реализации морфологического и синтаксического анализаторов.

I. Морфологический анализатор

Морфологический анализатор решает такие вопросы, как определение части речи, т.е. на вход падает набор слов, а на выходе получаем слово и его морфологические свойства; генерирование словоформы по заданным параметрам. Морфологический анализатор может состоять из двух модулей: первый модуль – это словарь и собственно сам анализатор, работающий со словарем по некоему алгоритму; второй модуль – это «угадывающий» анализатор, который, в случае отсутствия результатов работы первого модуля, определяет возможную часть речи. К примеру, возможная реализация первого модуля следующая: составляется словарь, в котором хранится часть речи с её морфологическими особенностями. Например, для существительного можно хранить основу слова в единственном числе, а также основу слова для множественного числа. Далее, необходимо хранить также род и склонение, чтобы была возможность сгенерировать любую форму данного существительного. Кроме этого можно хранить некоторые дополнительные признаки: одушевленность/неодушевленность и т.п. Таким же образом составляются признаки для других частей речи. А сам анализатор для определения части речи может использовать следующий алгоритм: ищется основа совпадающая с текущим словом, в случае совпадения это слово запоминается, в противном случае у слова отсекается последняя буква отсекается и предыдущие действия повторяются. Затем полученные основы пробуют привести к требуемому слову.

Сироткин Александр Владимирович, студент 4 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, salexvlad@gmail.com.

Шакур Денис Евгеньевич, студент 4 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, denisshakur@gmail.com.

Научный руководитель: Герман Олег Витольдович, доцент кафедры информационных технологий автоматизированных систем Белорусского государственного университета информатики и радиоэлектроники, кандидат технических наук, доцент, ovgerman@ut.by.

Таким образом, для требуемого слова получаем его морфологические признаки. В случае отрицательного результата работы первого модуля, запускается второй модуль, который можно реализовать на основе нейронных сетей со следующей структурой: входные нейроны будут соответствовать возможным окончаниям слов, а выходные – самостоятельным частям речи.

II. Синтаксический анализатор

Синтаксический анализатор строит структуру предложения, т.е. составляет дерево разбора предложения, которое отображает связи между словами. Существуют несколько подходов создания синтаксических анализаторов: машинное обучение с учителем и без, и метод, основанный на правилах. Но из-за того, что русский язык имеет свободный порядок слов машинное обучение с учителем использовать нецелесообразно. В русском языке выделяют три вида подчинительной связи: согласование, управление, примыкание – каждой из которых соответствуют определенный набор морфологических правил. Поэтому, зная морфологические свойства, можем выстроить возможные связи между словами.

1. Kübler. S. Dependency Parsing / S. Kübler – 2000. – Vol. 127.
2. Хабрахабр [Электронный ресурс] Электронные данные. – Режим доступа: <http://habrahabr.ru>
3. AOP [Электронный ресурс] Электронные данные. – Режим доступа: <http://aop.ru/>
4. LanguageTool [Электронный ресурс] Электронные данные. – Режим доступа: <http://languagetool.org/>