Грамматики и синтаксический анализ

Евгений Борисов

Уровни сложности при автоматической обработке текстов

Прагматика - смысловые контексты

Семантика - смыслы последовательностей слов

Синтаксис - правила формирования последовательностей слов

Лексика - отдельные слова и устойчивые словосочетания

Задача автоматического синтаксического разбора

Применение:

- машинный перевод
- извлечение информации
- диалоговые системы

Грамматика — способ описания языка

- грамматика зависимостей
- грамматика составляющих

Грамматика зависимостей (Dependency grammar)

Главные члены предложения

Подлежащее — предмет. кто? что?

Сказуемое — что делать? что сделать? каков?

inventory DET P-ATTR Of COMP-P functions ATTR syntactic An inventory of syntactic functions is taken to be primitive.

Образец разбора предложения

В саду расцвели красные и белые розы.



Это предложение – повествовательное, невосклицательное. Основа предложения – розы (подлежащее) расцвели (сказуемое). В предложении есть второстепенные члены, поэтому оно распространённое. Розы (какие?) красные и белые – однородные определения, произносятся с интонацией перечисления. Расцвели (где?) в саду – обстоятельство.

Второстепенные члены предложения

Определение — признак предмета. какой? чей? который?

Обстоятельство — время, место, способ действия. где? когда? куда? откуда? почему? зачем? как?

Дополнение — предмет. кого? чего? кому? чему? кого? что? кем? чем? о ком? о чём?

Грамматика составляющих (Constituency grammar)

предложение (П; sentence, S).

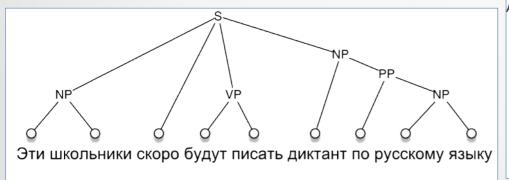
<u>именная группа</u> (группа существительного, ИГ; noun phrase, NP) — возглавляется существительным;

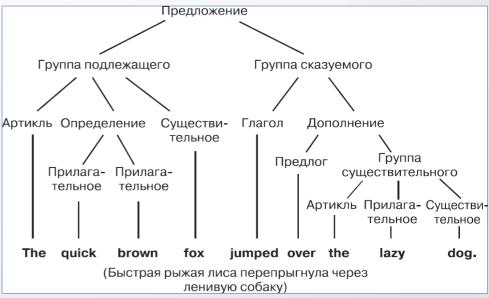
группа прилагательного (ГПрил; adjectival phrase, AP) — возглавляется прилагательным;

наречная группа (НарГ; adverbial phrase, AdvP) — возглавляется наречием;

предложная группа (ПрГ; prepositional phrase, PP) — возглавляется предлогом;

глагольная группа (ГГ; verb phrase, VP) — возглавляется глаголом;





[s[NP Эти школьники] скоро[vP будут писать][NP диктант[PP по [NP русскому языку]]]]

Литература

```
git clone https://github.com/mechanoid5/ml_nlp.git
```

Турдаков Д.Ю. Основы обработки текстов. лекция 7. Формальные грамматики и синтаксический анализ. ИСП РАН, 2017 https://www.youtube.com/watch?v=TkMtUm-D6aE

Steven Bird, Ewan Klein, and Edward Loper Analyzing Text with the Natural Language Toolkit https://www.nltk.org/book/

D.Jurafsky, J.H. Martin Speech and Language Processing. third edition, 2020