

МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЛИНГВИСТИЧЕСКИХ ДАННЫХ

Канищева О.В.

Национальный технический университет

"Харьковский политехнический институт", г. Харьков

Разработка естественно-языковых систем, так или иначе, сталкивается с проблемой представления, хранения и интерпретации лингвистической информации. В данном случае представление данных включает в себя средства и формализмы, используемые для их представления, методов их хранения в процессе обработки, а также интерпретацию данных системой.

Существуют различные классификации подходов к представлению лингвистических данных. В этой работе будем придерживаться следующей классификации:

1) Подходы, основанные на разметке (markup-based), при которых дополнительная информация хранится непосредственно в тексте в форме дополнительной разметки (HTML, SGML или XML).

2) Подходы, основанные на аннотациях (annotation-based), при которых информация хранится отдельно и содержит ссылки на исходный текст.

3) Подходы, основанные на абстракциях (abstraction-based), при которых текст хранится только как часть некоторой структуры данных, которая в свою очередь представляет всю информацию в виде, основанном на конкретном формализме.

4) Подходы, при которых отсутствуют какие-либо ограничения на представление данных.

В данной работе рассмотрим более детально подходы, основанные на абстракциях, так как они базируются на создании структуры признаков, что является распространенным средством для представления лингвистической информации. Многие формализмы для осуществления анализа используют их.

Особый интерес представляют такие формализмы как *Minimal Recursion Semantic (MRS)* и *Robust Minimal Recursion Semantic (RMRS)*. Основная идея этих формализмов состоит в преобразовании вложенной структуры в плоскую. Таким образом, вложенная структура признаков (или предикатов) преобразуется во множество структур (которые могут быть объединены символами конъюнкции). Формализм RMRS является развитием MRS, основное отличие которого состоит в том, что структуры из нескольких признаков разбиваются на бинарные предикаты.