

ЭТАПЫ ПОСТРОЕНИЯ ЛИНГВОСТАТИСТИЧЕСКОЙ МОДЕЛИ ЛИТЕРАТУРНОГО ПЕРСОНАЖА

Благодаря развитию современных технологий появляется все больше возможностей для исследований литературных текстов при помощи компьютерных инструментов. Особенно очевидными становятся изменения вектора исследований в области литературоведения. Одним из таких инструментов является тегирование.

Во время процесса тегирования исследуемый материал заключается в теги, которые представляют собой признак, релевантный для нашего исследования. Само тегирование, или же по-другому аннотирование, бывает разных видов: семантическим, синтаксическим, морфологическим и т.д. в зависимости от того, что подлежит разметке. Именно семантическое аннотирование делает возможным построение лингвостатистической модели героя литературного произведения, так как такой вид аннотирования непосредственно связан с распределением языковых единиц по определенным семантическим категориям. Под *лингвостатистической моделью* подразумевается способ описания персонажа и установления закономерностей его характеристик, при котором источником данных является частотность встречаемости отличительных черт исследуемого литературного героя.

Основными этапами построения лингвостатистической модели литературного персонажа являются: создание корпуса текстов и приведение его в машиночитаемую форму, создание системы семантических тегов, аннотирование текстового корпуса, обработка аннотированного корпуса с помощью корпусного менеджера, автоматическое построение статистик по отдельным характеристикам персонажа, сведение отдельных количественных и вербальных данных в единую модель, интерпретация построенной модели.

У такого способа исследований литературных персонажей имеются как преимущества, так и недостатки. Главным недостатком является проблема распределения отличительных черт по тегам, так как определенные характеристики литературного героя могут быть интерпретированы в зависимости от того, как воспринимает персонажа исследователь. Однако точность полученных данных в результате анализа тегированного материала, а также возможность параллельных текстов дает преимущество такому типу исследований литературных персонажей.