

## ГЕНЕРАЦИЯ ТИПОВ СВЯЗЕЙ СЕМАНТИЧЕСКОЙ СЕТИ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Извлечение семантического содержания текста требует предварительного синтаксического анализа предложений. В настоящее время используется метод неполного синтаксического анализа путем сравнения простых предложений с записями в базе данных шаблонов предложений при помощи регулярных выражений. Каждый набор из утвердительного и вопросительного предложений соответствует определенному типу семантической связи. Эффективность данного метода напрямую зависит от числа шаблонов в базе данных, поэтому для анализа больших текстов требуется наполнять базу данных типов семантических связей автоматически.

В соответствии с лингвистическим методом актуального членения предложений тип смысловой связи между подлежащим и дополнительными членами предложений определяется сказуемым. Путем исследования текста электронного учебника по электрохимии было выявлено, что наиболее распространенными сказуемыми являются глаголы в форме третьего лица и краткие прилагательные. Таким образом, в основе большинства шаблонов предложений, выражающих семантические связи, стоит глагол в форме третьего лица или краткое прилагательное. Также в шаблоне при помощи специальных тегов отмечены места для подстановки подлежащего, дополнительных членов предложения, окончания и вопросительного слова.

Составление таких шаблонов становится возможным при качественно проведенном морфологическом анализе, который распознает в тексте глаголы в форме третьего лица и краткие прилагательные. Для этого во всех найденных сказуемых выполняется замена их окончаний на соответствующие теги. Затем выполняется формирование шаблонов вопросительного и утвердительного предложений. Для вопросительного предложения перед сказуемым добавляется тег вопросительного слова, а после – тег подлежащего. Шаблон утвердительного предложения дополняется тегом подлежащего перед сказуемым и тегом дополнительных членов предложения – после сказуемого. Так, для предложения «*Электродный потенциал не зависит от количества вещества, прореагировавшего на электродах*» был получен следующий SQL-запрос на вставку типа семантической связи в базу данных: *INSERT INTO шаблоны VALUES ("[B] завис[Г2] [A]", "[A] завис[Г2] [B]").*