Для обращения к системе связных данных (LOD) используется язык SPARQL [1]. Но для людей не знакомых с ним это делает использование LOD абсолютно невозможным (SPARQL имеет высокий порог входа). Для решения этой проблемы исследователями, работающими над данной проблемой, разрабатываются естественно-языковые интерфейсы. Они позволяют обращаться к связным данным с помощью запросов на естественном языке. Но все работы, сделанные в данной области [2-4], в качестве естественного языка предполагают использование английского или его подъязыков. Поэтому, целью данной работы ставится разработка интеллектуального естественно-языкового интерфейса для обращения к LOD на русском языке.

В рамках данной работы будет рассмотрено, что такое система LOD, на каких принципах она основана, о том, что для представления данных в ней используется языковая система описания ресурсов RDF и приведены сведения об уже существующих проектах, в которых применяется система LOD.

Также будет необходимо изучить язык запросов к сиcтеме LOD SPARQL, его синтаксис и основные конструкций.

Для выбора представления смысла запроса на естественном языке будут рассмотрены следующие подходы к формальному представлению семантической структуры текста: абстрактному представлению смысла (АПС), грамматике Монтегю и теория К-представлений (концептуальных представлений) В.А. Фомичева [5]. АПС и грамматика Монтегю специализированы на семантическом представлении английского языка, и с их помощью возможно описать только СП простых предложений. В свою очередь, теория К-представлений дает формальный аппарат для описания семантической структуры связных текстов (дискурсов), а не только отдельных предложений, на русском языке и других языках. Учитывая все эти факторы, для формального описания СП запроса на ЕЯ выбрана теория К-представлений.

Будут рассмотрены подходы к разработке алгоритмов получения семантического представления текста на ЕЯ. Исследователи, использующие в своих работах семантическое представление, применяют последовательный лексический, синтаксический, семантический анализы с последующим построением семантического представления. В свою очередь в теории К-представлений В.А. Фомичев [5] предлагает новый метод построения семантического представления текста на ЕЯ (Фомичев 2005). В рамках нового метода синтаксический и семантический анализ проводится одновременно и не используется синтаксический уровень представления текста. Упомянутый выше метод предполагает 3 этапа: компонентно-морфологический анализ входного текста, построение матричного семантико-синтаксического представления (МССП) и непосредственно сборка семантического представления текста по его МССП.

Также потребуется рассмотрение подходов, применяемые в проектах, осуществляющих обращения к LOD на ЕЯ, для перевода запроса на ЕЯ в запрос на языке SPARQL. В частности, рассматриваются подходы, применяемые французским исследователем Себастиеном Ферре [2;3] и американскими исследователями Нели Златаревой и Девишем Амином в работе [4].