



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**  
**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ**  
**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**  
**«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ**  
**(национальный исследовательский университет)»**

Кафедра 319 «Системы интеллектуального мониторинга»

## Принципы проектирования семантически- ориентированного естественно-языкового интерфейса для взаимодействия с системой Linked Open Data

Автор:  
студент группы МЗО-435Б-18  
Урубков В.А.  
Научный руководитель:  
Фомичев В.А.

**Москва, 2022**



# Цели и задачи

Необходимо реализовать семантически-ориентированный естественно-языкового интерфейса для взаимодействия с системой открытых данных.

Для этого необходимо:

- выбрать подход к описанию семантического представления текста на естественном языке;
- разработать алгоритм построения запроса к LOD на языке SPARQL по семантическому представлению исходного запроса на ЕЯ.

# Linked Open Data (LOD)

Система LOD – огромный размеченный ориентированный граф, состоящий из элементарных графов, являющихся тройками вида (субъект, предикат, объект).

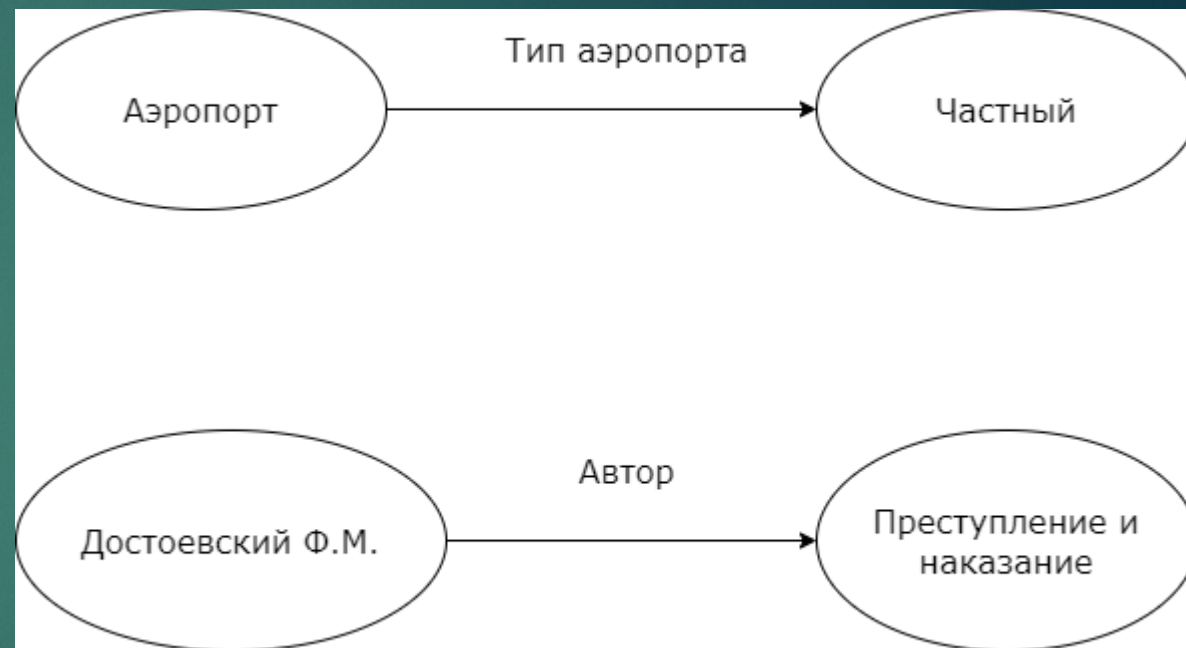




# RDF

Resource Description Framework – язык для создания распределенных баз знаний (онтологий).

Основной структурой языка RDF являются триплеты – упорядоченные тройки вида (субъект, предикат, объект)



# SPARQL

SPARQL (рекурсивный акроним SPARQL Protocol and RDF Query Language) – язык запросов к данным, представленным в формате RDF.

```
select distinct ?name
where {
    ?town rdf:type dbo:City.
    ?town dbo:country dbr:Russia.
    |
    ?town rdfs:label ?name
    filter(lang(?name) = "ru")
}
```

name
"Ульяновск"@ru
"Астрахань"@ru
"Бийск"@ru
"Челябинск"@ru



# Подходы к описанию семантического представления текстов на естественном языке

Абстрактное  
представление  
смысла

Теория К-представлений  
В.А. Фомичева

Грамматика  
Монтегю

# Сравнение подходов

Параметр	Абстрактное представление смысла	Грамматика Монтегю	Теория К-представлений В.А. Фомичева
Язык текста	Английский	Английский	Русский, Немецкий, Французский, Английский
Типы текстов	Повествовательные предложения	Повествовательные предложения и вопросы	Фразы-высказывания, повествовательные тексты, команды, вопросы
Допустимая структура текста	Отдельное предложение	Отдельное предложение	Связный текст наравне с отдельными предложениями



# Неоднозначность именования в ОНТОЛОГИЯХ

Предикаты, описывающие одно и то же отношение между объектами, могут иметь разные имена даже в рамках одной онтологии.

Например, в онтологии DBpedia города «Оттава», «Москва», «Ульяновск» и «Северодвинск» для обозначения количества населения имеют разные предикаты: «population», «populationTotal», «p» и «pop2010census», соответственно

Город	Предикат
Оттава	population
Москва	populationTotal
Ульяновск	p
Северодвинск	pop2010census



# Недостаточная связанность данных

В онтологии `уаго`, использующей систему типов `Schema`?:

- ▶ У любых объектов Автомобилей отсутствуют содержательные предикаты, хотя информация для них есть
- ▶ Объект Город связывается со страной, в которой он располагается, с помощью строки с комментарием типа «Это столица России»

`rdfs:comment`

"сталица Расеі"@be-tarask

"Столица на Русия"@bg

"capital de Rússia"@ca

"hlavní město Ruské federace"@cs

"Hovedstad i Rusland"@da

# Принципы преобразования параметров запросов к LOD

Для обеспечения перевода запроса к LOD на естественном языке в запрос на языке SPARQL необходимо заранее связывать параметры запроса (отношения и некоторые значения) с аналогичными параметрами онтологии.

Отношение	Предикаты
Колич-Жителей	population
	populationTotal
	p
	pop2010census





# Спасибо за внимание!

Принципы проектирования семантически-ориентированного естественно-языкового интерфейса для взаимодействия с системой Linked Open Data

Автор:  
студент группы МЗО-435Б-18  
Урубков В.А.  
Научный руководитель:  
Фомичев В.А.