



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

НИУ ВШЭ - Пермь,
Научно-учебная лаборатория учебных корпусов

004.912

РАЗРАБОТКА СИСТЕМЫ ОЦЕНКИ СЛОЖНОСТИ ТЕКСТА НА АНГЛИЙСКОМ ЯЗЫКЕ

Научный семинар цикла
«Цифровые трансформации»

Работу выполнила студентка группы
ПИ-18-2 3 курса
А.А. Иксанова

Пермь, 2021



ПЛАН ДОКЛАДА

1. Постановка проблемы
2. Определение сложности текста
3. Оценка сложности текста
4. Синтаксическая сложность текста
5. Практическое применение оценки сложности текста
6. Анализ существующих решений
7. Формирование требований к приложению
8. Проектирование серверной части системы
9. Проектирование клиентской части системы
10. Проектирование шаблонов для реализации маркеров сложности текста
11. Реализация шаблонов на языке JAPE
12. Создание GATE-приложения
13. Реализация системы
14. Разработка интерфейса
15. Результат выполнения программы
16. Заключение

1. ПОСТАНОВКА ПРОБЛЕМЫ



2. ОПРЕДЕЛЕНИЕ СЛОЖНОСТИ ТЕКСТА

Сложность текста (text complexity) – это характеристика, которая зависит от внутренних параметров текста и позволяет определить, насколько он легок к восприятию.



3. ОЦЕНКА СЛОЖНОСТИ ТЕКСТА



3. ОЦЕНКА СЛОЖНОСТИ ТЕКСТА





4. СИНТАКСИЧЕСКАЯ СЛОЖНОСТЬ ТЕКСТА

Синтаксическая (грамматическая) сложность текста – раздел лингвистической сложности, направленный на изучение количества различных синтаксических структур и степени их сложности.

Фразовая

- Количество существительных и прилагательных, являющихся определениями существительного

Клаузальная

- Количество придаточных предложений в функции именной части сказуемого или обстоятельственных придаточных

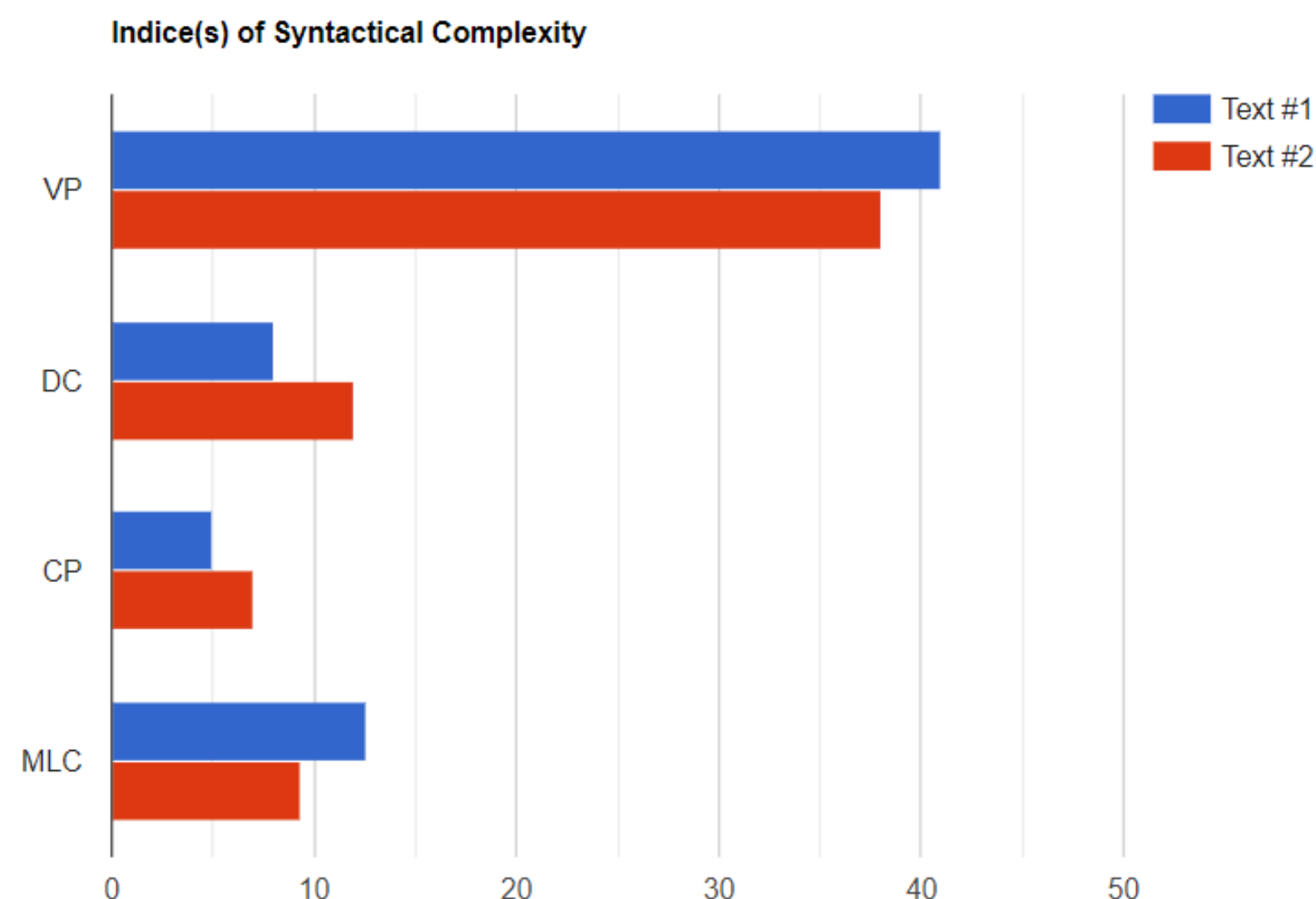
5. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ОЦЕНКИ СЛОЖНОСТИ ТЕКСТА

- Подбор учебных текстов в зависимости от возраста учащихся и их уровня владения языком;
- Проверка разрабатываемой документации на соответствие определенному уровню;
- Написание статей, нацеленных на конкретную аудиторию;
- Определение уровня владения языком.



Институт государственного и муниципального управления НИУ ВШЭ (ИГМУ) провел исследование «Оценка сложности языка законов» и констатировал, что нормативные акты становятся все более сложными. Из 592 проанализированных законодательных актов 77 оказались сложнее, чем «Критика чистого разума» Иммануила Канта, которую проходят в аспирантуре.

6. АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ



Web based L2 Syntactic Complexity Analyzer:

- + 23 метрики синтаксической сложности текста;
- + возможность сравнительного анализа двух текстов;
- + построение графика по результатам работы;
- ограничение длины текста в 1000 символов;
- отсутствие комментариев к рассчитанным числовым параметрам.

Результат сравнения двух текстов

Text #1

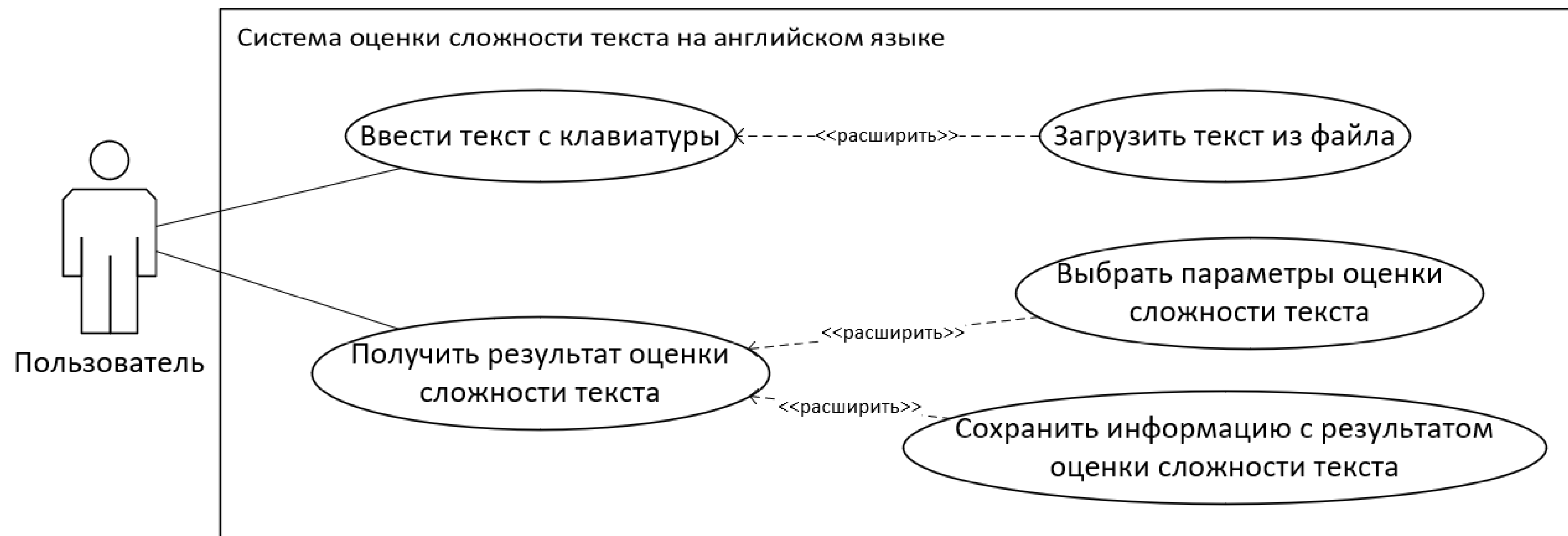
W, S, VP, C, T, DC, CT, CP, CN, MLS, MLT, MLC, C/S, VP/T, C/T, DC/C, DC/T, T/S, CT/T, CP/T, CP/C, CN/T, CN/C
562, 24, 62, 40, 27, 14, 12, 13, 86, 23.4167, 20.8148, 14.0500, 1.6667, 2.2963, 1.4815, 0.3500, 0.5185, 1.1250,
0.4444, 0.4815, 0.3250, 3.1852, 2.1500

Text #2

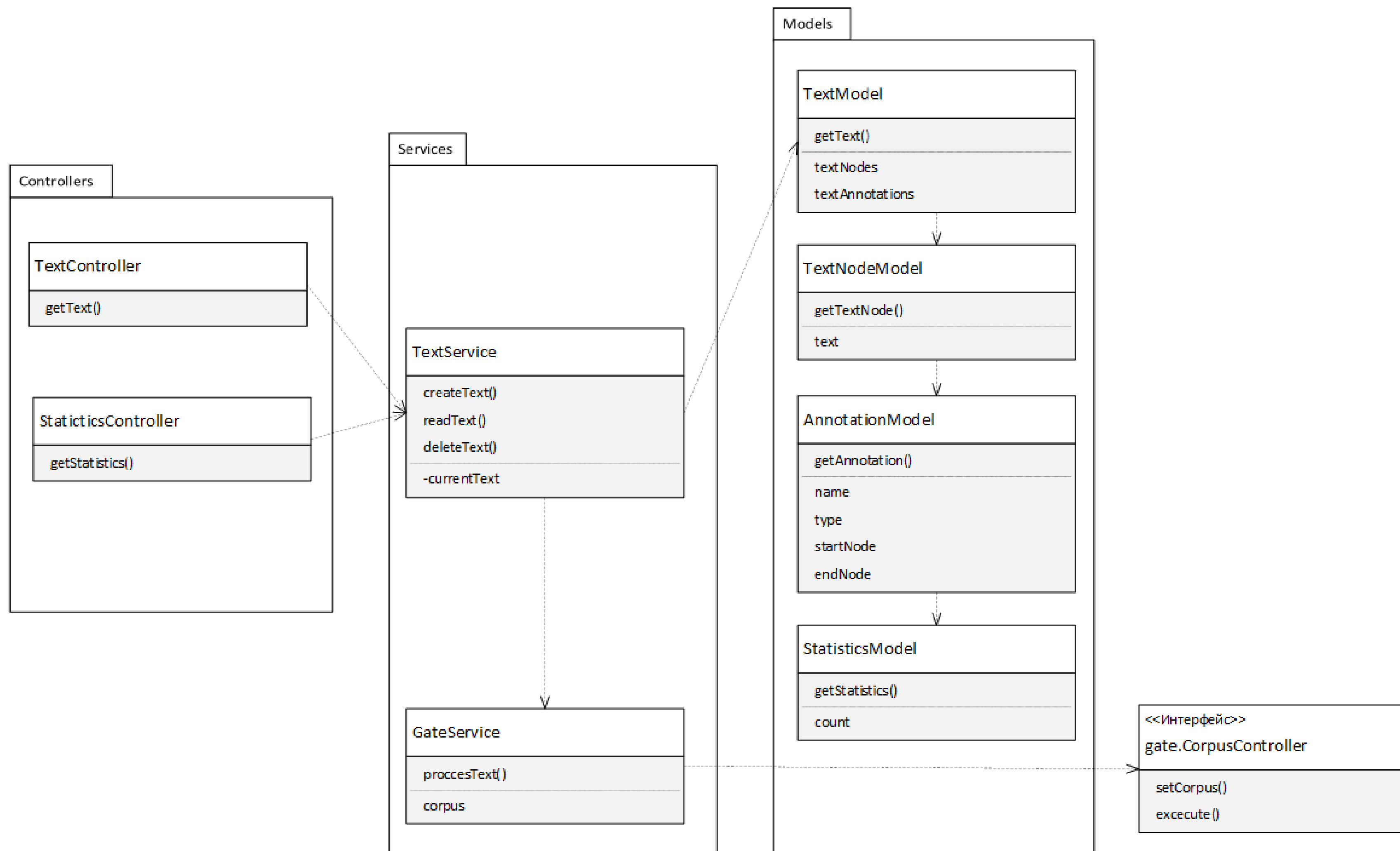
W, S, VP, C, T, DC, CT, CP, CN, MLS, MLT, MLC, C/S, VP/T, C/T, DC/C, DC/T, T/S, CT/T, CP/T, CP/C, CN/T, CN/C
301, 16, 41, 24, 16, 8, 7, 5, 48, 18.8125, 18.8125, 12.5417, 1.5000, 2.5625, 1.5000, 0.3333, 0.5000, 1.0000, 0.4375,
0.3125, 0.2083, 3.0000, 2.0000

Рассчитанные числовые параметры

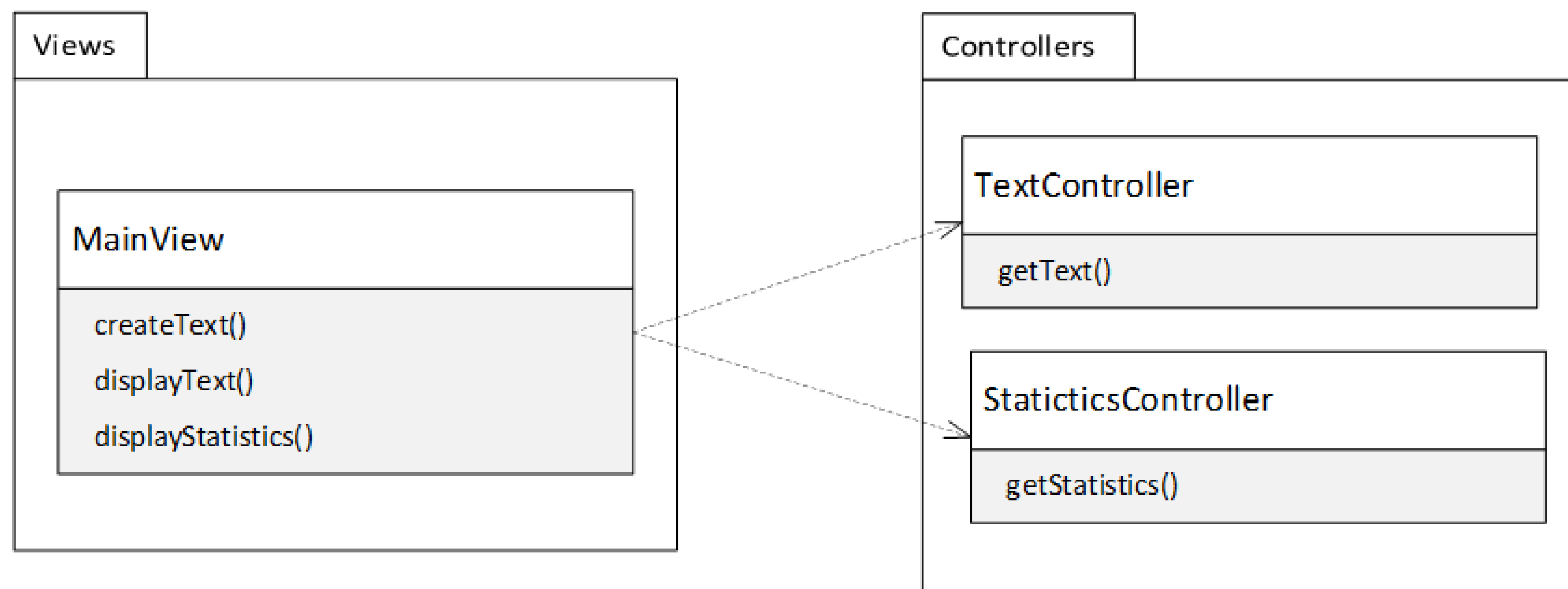
7. ФОРМИРОВАНИЕ ТРЕБОВАНИЙ К СИСТЕМЕ



8. ПРОЕКТИРОВАНИЕ СЕРВЕРНОЙ ЧАСТИ СИСТЕМЫ



9. ПРОЕКТИРОВАНИЕ КЛИЕНТСКОЙ ЧАСТИ СИСТЕМЫ





10. ПРОЕКТИРОВАНИЕ ШАБЛОНОВ ДЛЯ РЕАЛИЗАЦИИ МАРКЕРОВ СЛОЖНОСТИ ТЕКСТА

Для упрощения записи формул используются следующие сокращения:

- VV – глагол
- NN – существительное
- DT – артикль
- PP – личное местоимение
- JJ – прилагательное.

Phrasal	
Nouns	<i>NN *</i>
Attributive adjectives	<i>JJ * NN *</i>
Premodifying nouns	<i>NN * NN *</i>
Of genitives	<i>NN* of * NN*</i>

Clausal			
Finite adverbial clauses	<i>In order that</i>	<i>until</i>	
	<i>if</i>	<i>as soon as</i>	
	<i>unless</i>	<i>because</i>	
	<i>in the event that</i>	<i>since</i>	
	<i>provided that</i>	<i>where</i>	
	<i>although</i>	<i>so that</i>	
	<i>even though</i>	<i>as if</i>	
	<i>despite the fact that</i>	<i>as though</i>	
	<i>when</i>	<i>whereas</i>	
	<i>after</i>	<i>while</i>	
	<i>before</i>		
	Wh-complement clauses	<i>VV* Why * NN*</i>	<i>VV * what * NN*</i>
		<i>VV * Why * DT * NN*</i>	<i>VV * what * DT * NN*</i>
<i>VV * Why * * PP</i>		<i>VV * what * PP</i>	
<i>VV * where * NN*</i>		<i>VV * how * NN*</i>	
<i>VV * where * DT * NN*</i>		<i>VV * how * DT * NN*</i>	
<i>VV * where * PP</i>		<i>VV * how * PP</i>	
<i>VV * when * NN*</i>		<i>VV * which * NN*</i>	
<i>VV * when * DT * NN*</i>		<i>VV * which * DT * NN*</i>	
<i>VV * when * PP</i>		<i>VV * which * PP</i>	
Verb+that clauses		<i>* VV * that</i>	

11. РЕАЛИЗАЦИЯ ШАБЛОНОВ НА ЯЗЫКЕ JARE









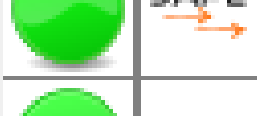


```
Phase:OfGenitives  
Input: Token  
Options: control = all
```

```
правило { Rule: OfGenitives  
(  
  ({Token.category == "NN"} | {Token.category == "NNS"} |  
    {Token.category == "NNP"} | {Token.category == "NNPS"})  
  {Token.string == "of"}  
  ({Token.category == "NN"} | {Token.category == "NNS"} |  
    {Token.category == "NNP"} | {Token.category == "NNPS"})  
): label  
-->  
действие { :label.OfGenitives = {rule= "OfGenitives" }
```

JARE-правило для поиска «of genitives»

12. СОЗДАНИЕ GATE-ПРИЛОЖЕНИЯ

Selected Processing resources

!	Name	Type
	Document Reset PR	Document Reset PR
	ANNIE English Tokeniser	ANNIE English Tokeniser
	ANNIE Sentence Splitter	ANNIE Sentence Splitter
	ANNIE POS Tagger	ANNIE POS Tagger
	AttributiveAdjectives	JAPE Transducer
	FiniteAdverbialClauses	JAPE Transducer
	OfGenitives	JAPE Transducer
	PremodifyingNouns	JAPE Transducer
	VerbThatClauses	JAPE Transducer
	Nouns	JAPE Transducer
	Wh-complementClauses	JAPE Transducer

Модули созданного приложения

I will then explore why it still persists
 We believe that coursework no longer exists.
 He also stated that globalization demanded increased cooperation among developing countries in order that the less developed among them would not be marginalized.
 Verify if your mouse works correctly and if the resolution is good.
 And you will face charges unless you cooperate.
 Those issues should also be considered in a wider context in the event that a convention was deemed unnecessary.
 Parties may use their own approaches provided that they result in equal stringency.
 All papers were distributed, although not individually presented.
 We are seeking a solution even though that may require major compromises on our side.
 This is despite the fact that delegations have repeatedly expressed their concerns and requested them to be taken into account.
 We did regular maintenance and repairs when necessary.
 The war began after increasing tension between Aro leaders and British colonialists after years of failed negotiations.
 On 5 April, three days before local elections and a day before a planned general strike, the government banned all demonstrations.
 I never understood what that meant until now.
 This information will be reported to the Assembly as soon as it becomes available.
 They didn't attack you because you created them.
 This is particularly the case since the present international setting has undergone considerable change since the third special session was held in 1988
 For that purpose it ought to consider where the claimant was domiciled, where he conducted his business and where he exercised his political rights.
 At issue is how best to tailor interventions so that they address critical areas with the broadest impact in a comprehensive manner, so that efforts are not spread too thin on the ground.
 We often talk about coordination as if it were some cure-all or abstract concept.
 The report treats those claims as though they were credible.
 It reacts quickly to needs whereas the educational system tends to change very slowly.
 These cases indicate that while accountability is possible, it occurs under exceptional circumstances.
 At this age, the meaning of life began to interest me.
 When opening a new building at the London School of Economics, the Queen of England asked why no one had predicted the crisis.

- AttributiveAdjectives
- FiniteAdverbialCl
- Nouns
- OfGenitives
- PremodifyingNouns
- Sentence
- SpaceToken
- Split
- Token
- VerbThatClauses
- WhComplementClau...
- ▶ Original markups

Разметка текста созданным приложением в среде GATE Developer



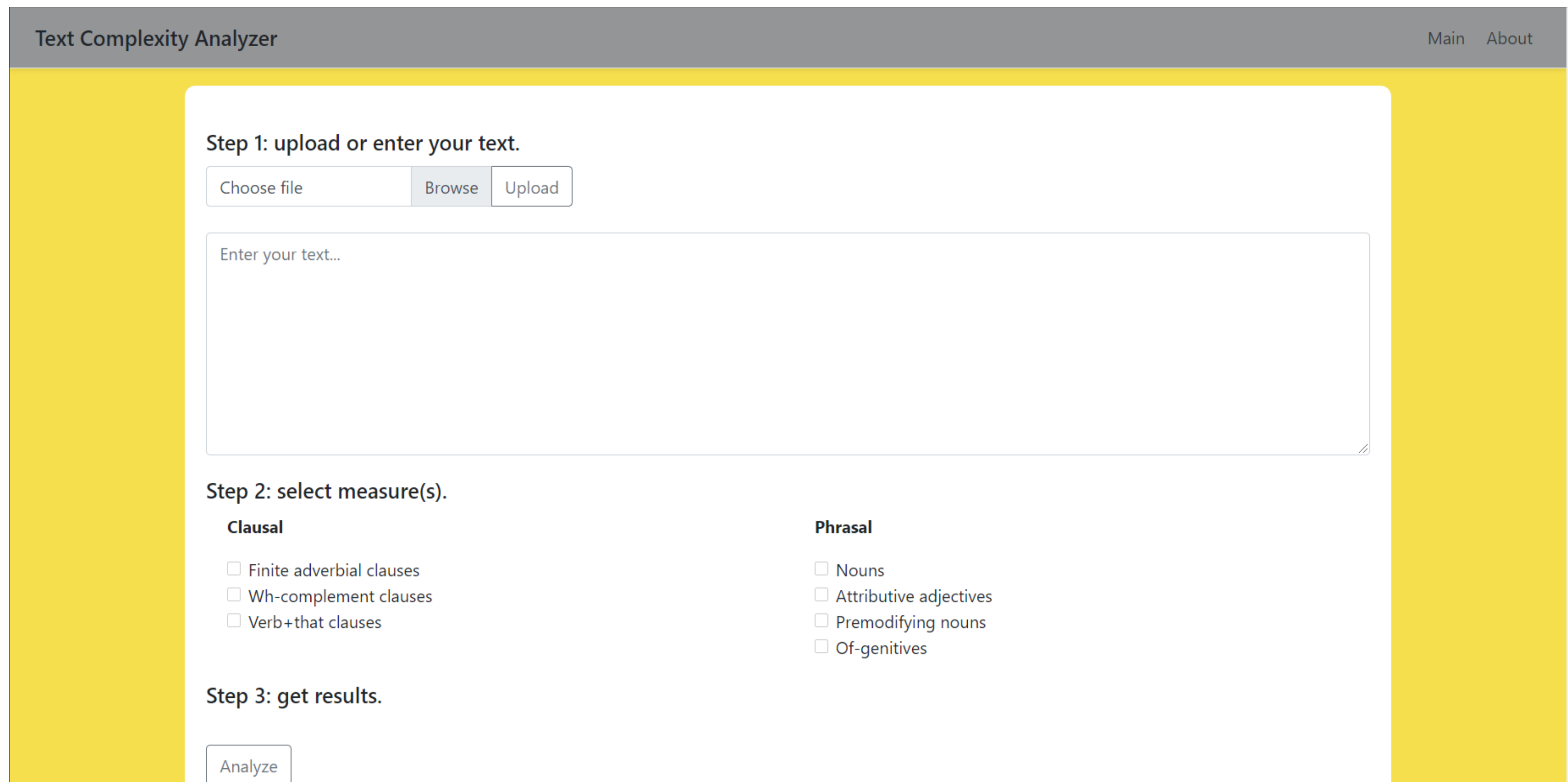
13. РЕАЛИЗАЦИЯ СИСТЕМЫ

Основные функции, реализованные в системе:

1. Загрузка текста из файла.
2. Выбор параметров оценки сложности текста.
3. Получение результатов оценки сложности текста.
4. Сохранение информации о результатах оценки в файл формата txt.



14. РАЗРАБОТКА ИНТЕРФЕЙСА



Главная страница системы



15. РЕЗУЛЬТАТ ВЫПОЛНЕНИЯ ПРОГРАММЫ

Text Complexity Analyzer Main About

Step 1: upload or enter your text.

Choose file Browse Upload

Enter your text...

Step 2: select measure(s).

Clausal

- Finite adverbial clauses
- Wh-complement clauses
- Verb+that clauses

Phrasal

- Nouns
- Attributive adjectives
- Premodifying nouns
- Of-genitives

Step 3: get results.

Analyze



16. ЗАКЛЮЧЕНИЕ

Результаты:

- Проведен анализ способов оценки сложности текста.
- Из них выбрана и реализована оценка синтаксической сложности текста.
- Сформулированы требования к системе и разработан интерфейс пользователя.
- Разработаны JARE-шаблоны для нахождения в тексте на английском языке маркеров синтаксической сложности.
- Реализована и протестирована система оценки сложности текста на английском языке.

Перспективы:

- Использование для определения сложности текстов при обучении английскому языку.
- Анализ корпусов английского языка с целью выявления особенностей синтаксической сложности текста, характерных для разного уровня владения языком.

Возможные усовершенствования:

- Выделение маркеров сложности в тексте различными цветами.
- Усовершенствование инструмента рекомендаций путем анализа экспертного корпуса текстов на английском языке и выявления закономерностей зависимости числовых показателей от уровня владения языком.

Спасибо за внимание!



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Готова ответить на ваши вопросы

«Разработка системы оценки сложности текста на
английском языке»

Иксанова Алена Анваровна

Email: iksanovaal@yandex.ru



1



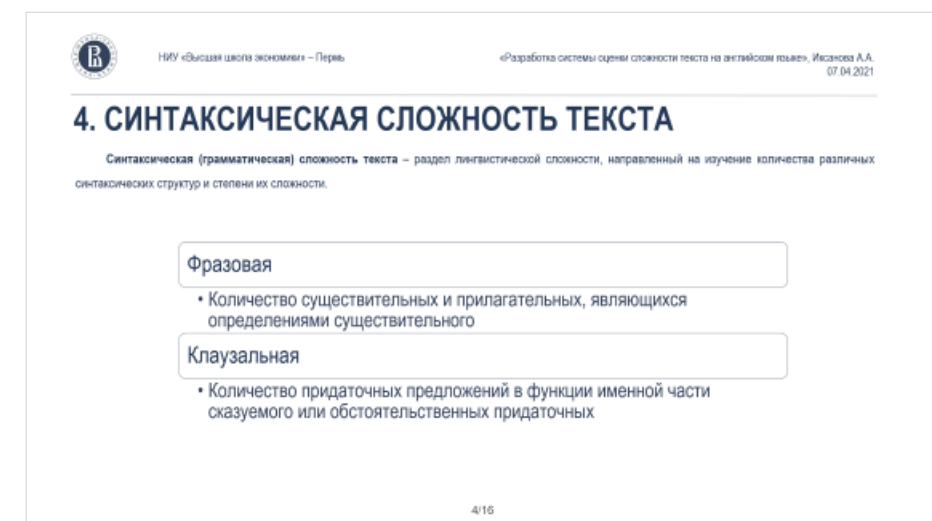
2



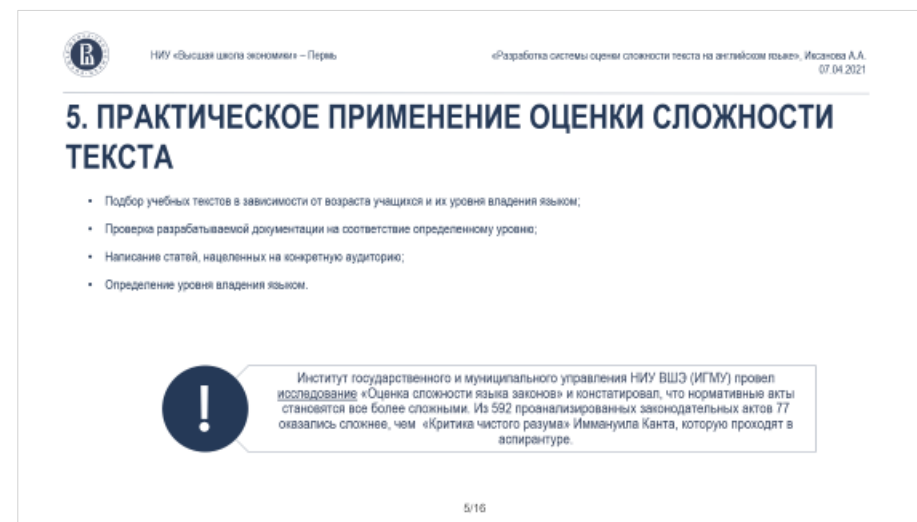
3



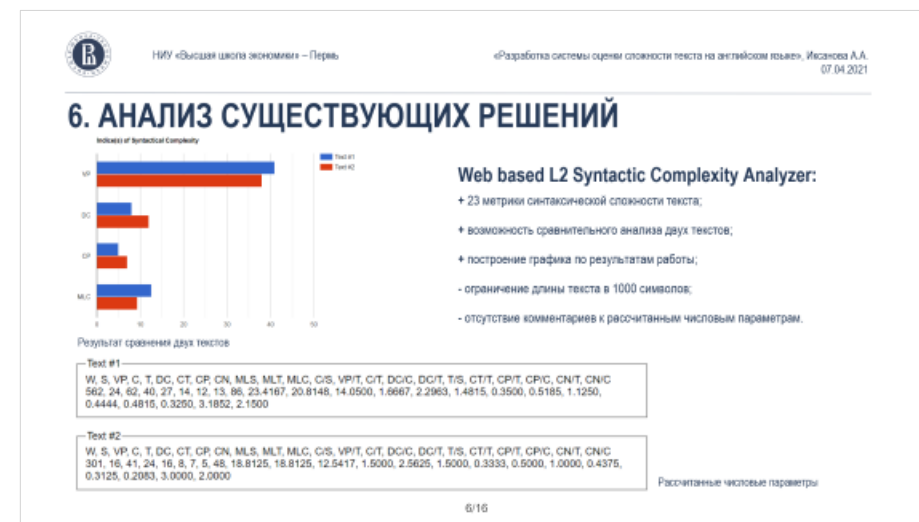
4



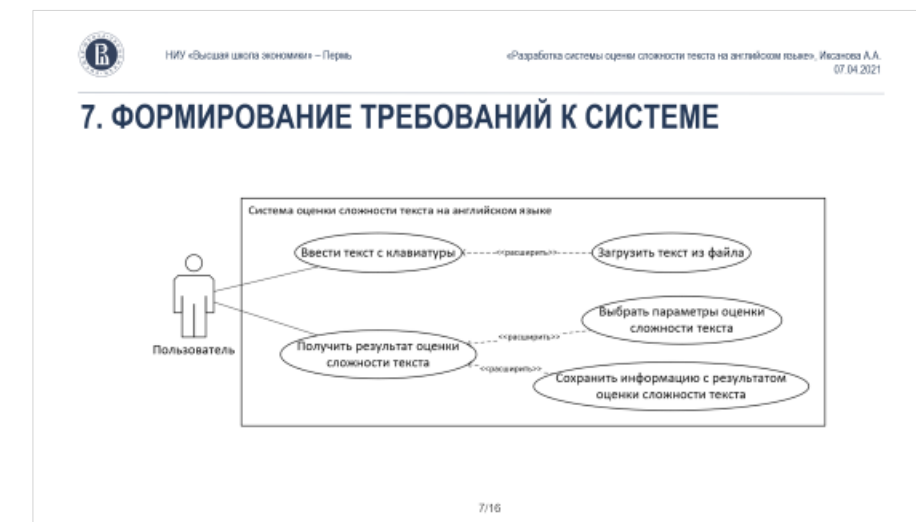
5



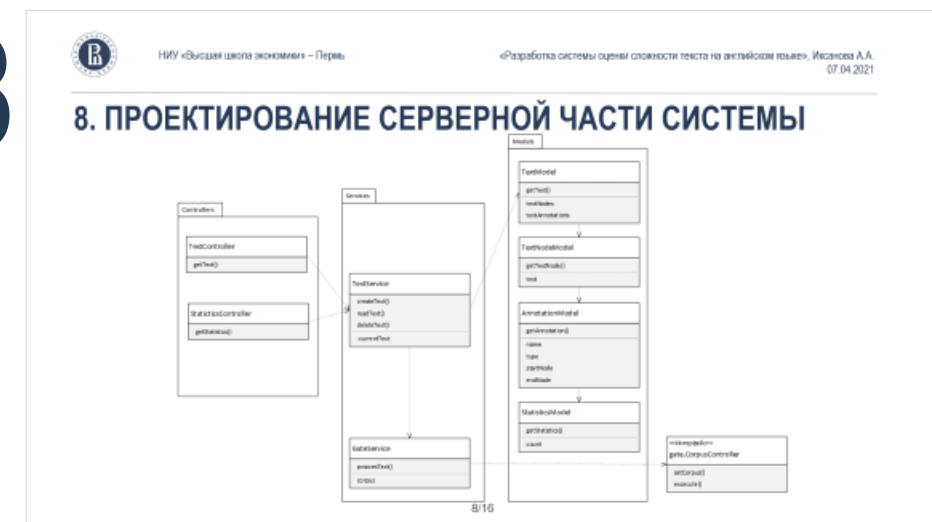
6



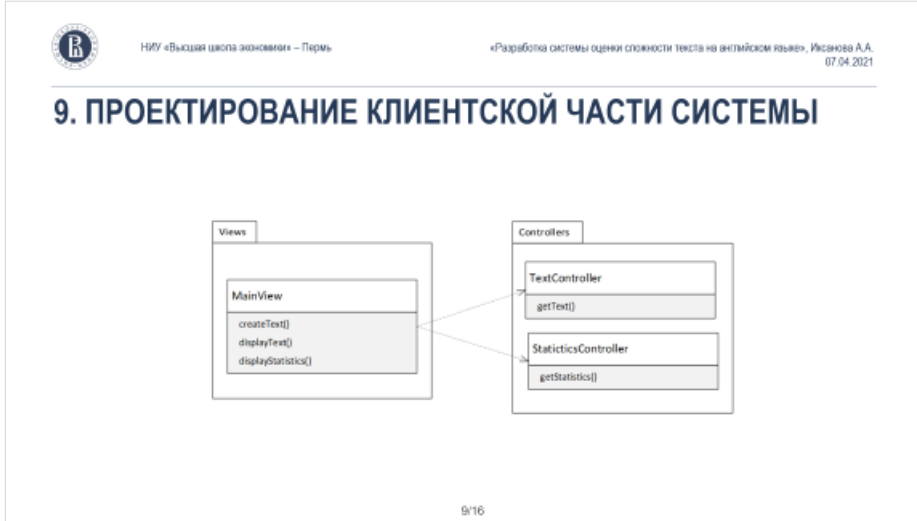
7



8



9



10

10. ПРОЕКТИРОВАНИЕ ШАБЛОНОВ ДЛЯ РЕАЛИЗАЦИИ МАРКЕРОВ СЛОЖНОСТИ ТЕКСТА

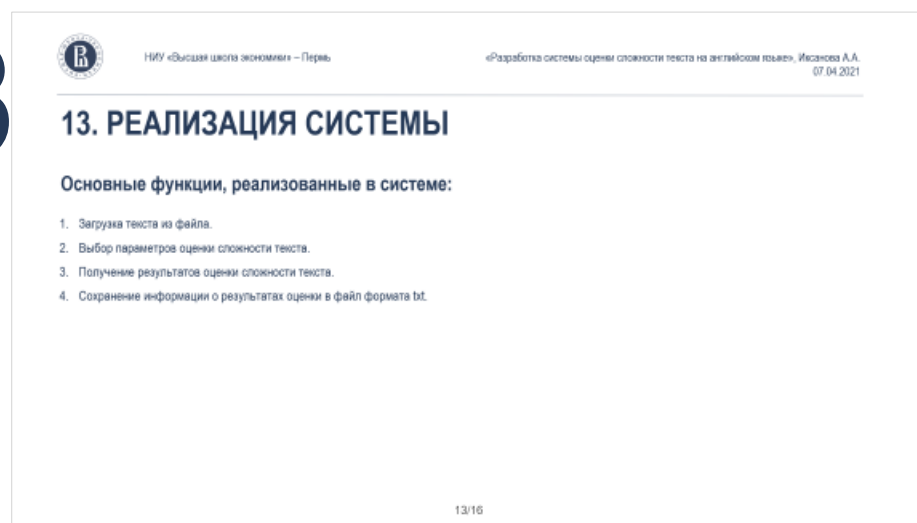
11

11. РЕАЛИЗАЦИЯ ШАБЛОНОВ НА ЯЗЫКЕ JARF

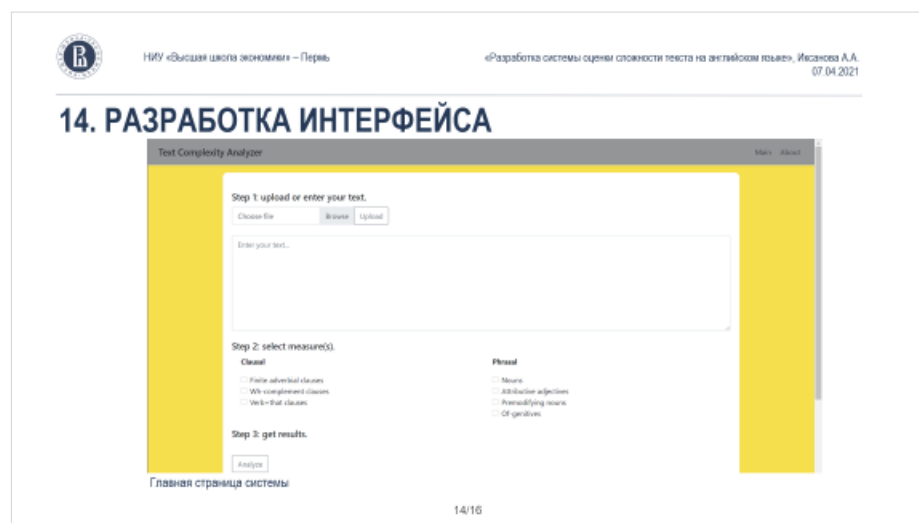
12



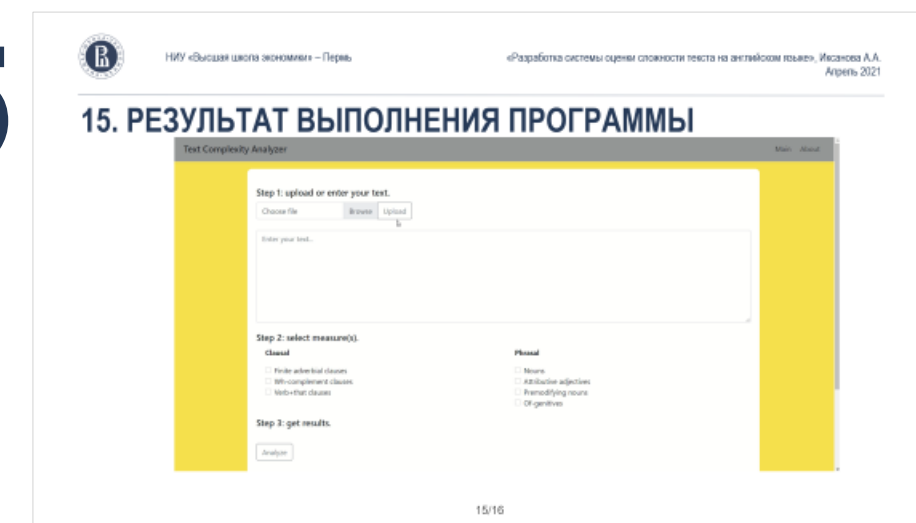
13



14



15



16

