Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Волгоградский государственный технический университет»

Утверждаю

Зав. кафедрой ПОАС

Орлова Ю.А.

« » 2020 г.

Качество и надёжность программного обеспечения

"Расшифровать сообщение на азбуке Морзе. "

Техническое задание

|  |  |
| --- | --- |
| СОГЛАСОВАНО:  Руководитель работы  доцент кафедры ПОАС  Сычёв О.А.  « » 2020 г. | РАЗРАБОТЧИК:  Студент группы ПрИн-266  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Тарпатина Е.С.  « » 2020 г. |
|  | НОРМОКОНТРОЛЁР:  доцент кафедры ПОАС  Сычёв О.А.  « » 2020 г. |

2020 г.

Содержание

[1 Введение 3](#_Toc509404267)

[2 Основания для разработки 3](#_Toc509404268)

[3 Назначение разработки 3](#_Toc509404269)

[4 Требования к программе 4](#_Toc509404270)

[4.1 Требования к функциональным характеристикам 4](#_Toc509404271)

[4.2 Требования к надёжности 6](#_Toc509404272)

[4.3 Требования к составу и параметрам технических средств 8](#_Toc509404273)

[4.4 Требования к информационной и программной совместимости 8](#_Toc509404274)

[5 Требования к программной документации 9](#_Toc509404275)

[6 Стадии и этапы разработки 10](#_Toc509404276)

[7 Порядок контроля и приемки 10](#_Toc509404277)

[Приложение 1. Список ключевых глаголов 11](#_Toc509404278)

[Приложение 2. Описание формата ввода графа на языке dot, совместимого с программой 12](#_Toc509404279)

[Приложение 3. Минимальные данные файла xml, необходимые для работы программы 14](#_Toc509404280)

# 1 Введение

Наименование программы – «Программа для расшифровки сообщения на азбуке Морзе» (далее «Программа»).

Программа предназначена к применению в исследованиях по моделированию и анализу текстов.

# 2 Основания для разработки

2.1 Основания для проведения разработки

Разработка проводится на основании задания на курсовой проект по дисциплине «Качество и надёжность программного обеспечения», выданного доцентом кафедры ПОАС Сычевым О.А., именуемым в дальнейшем Заказчиком, 13 февраля 2018 г.

2.2 Наименование и условное обозначение темы разработки

Наименование темы разработки – "Определение говорящих персонажей для участков прямой речи в отрывке художественной прозы на английском языке".

# 3 Назначение разработки

Разработка может быть использована в ходе исследования литературных прозаических текстов на английском языке для определения лексики, используемой персонажем.

# 4 Требования к программе

## 4.1 Требования к функциональным характеристикам

Программа должна выполнять следующие функции:

1) выделять имя говорящего, если оно следует в тексте за прямой речью как пояснение к ней с одним из ключевых глаголов (см. Приложение 1);

2) выделять имя говорящего, если оно находится непосредственно перед самой репликой с одним из ключевых глаголов (см. Приложение 1) и предваряет её;

3) выделять имя говорящего, если местоимение, которое обозначает говорящего и идет в тексте с одним из ключевых глаголов (см. Приложение 1), верно относится программой «Stanford CoreNLP» к самому имени;

4) если в списке персонажей только два персонажа, то программа должна определять говорящего, если в своей речи он обратился по имени к собеседнику или использовал уникальный для собеседника спецификатор в качестве обращения;

5) определять имя говорящего или группу возможных персонажей по объективному спецификатору, если объективный спецификатор идёт в тексте с одним из ключевых глаголов (см. Приложение 1) и предваряет или следует за репликой как пояснение к ней;

6) выделять по субъективному спецификатору в обращении персонажа или группу персонажей, которые могли произнести реплику;

7) определять субъект речи в случае вложенной прямой речи, если это возможно сделать при помощи предыдущих способов, используя текст только прямой речи, включающей в себя данную прямую речь;

8) если в списке персонажа только два персонажа и предыдущие правила не определили говорящего, то считается, что субъекты речи чередуются (моментом смены персонажа считается переход на новый абзац вне кавычек);

9) на вход подаётся разбор данного текста, списком персонажей и текст с выделенной прямой речью;

10) на выходе имеем текст с выделенными субъектами прямой речи и ключ.

Программа не должна проверять связи, установленные «Stanford CoreNLP», вывод о соответствии местоимения имени собственному делается на основе только связей, выявленных парсером. Определение с помощью данных парсера считается менее надёжным, чем методы под номерами 1,2,4,5,6,7. Если ответ получен любым из более надёжных методов, то данные парсера игнорируются.

Методы 1 и 2 считаются наиболее достоверными. По причине наибольшей достоверности данные методы применяются первыми. Если метод 1 дал ответ, то далее проверка не проводится. Если данные получены одним из этих методов, то данные, полученные от других методов, не учитываются.

Если методы 1 и 2 не дали результата, то используется метод 4. Если метод дал результат, то дальнейшие проверки не проводятся.

Если результат не получен методом 4, то проводится анализ методами 5 и 6. Пересечением полученных списков определяется список возможных субъектов речи.

Если методы 5 и 6 не дали однозначный результат, то используется метод 3. Если субъект, полученный в результате анализа методом 3 есть в пересечении списков, полученных методами 5 и 6, то этот персонаж и считается субъектом речи. Иначе данные метода 3 игнорируются и в ответ попадает только группа предполагаемых субъектов прямой речи.

Если методами 5 и 6 не удалось выделить группу возможных персонажей, то применяется метод 8.

Субъективным спецификатором является словосочетание, которым персонажа называет некоторый другой персонаж или группа персонажей и которое не может быть применено к нему другими персонажами.

Объективным спецификатором является словосочетание, которое по той или иной причине используется в произведении всеми персонажами по отношению к одному персонажу.

Обращением является имя собственное или субъективный спецификатор, заключённый между двумя запятыми, если обращение находится в середине предложения; ограниченный запятой после него, если обращение – первое слово в предложении; ограничено запятой слева, если обращение – последнее слово в предложении. Так же к обращению могут быть приписаны прилагательные, обращения «Mr.», «Sir», «Esq.», «Mrs.», «Ms.», «Miss», «Madam» и местоимение «you».

## 4.2 Требования к надёжности

В случае возникновения ситуаций ввода некорректных данных программа должна выдать соответствующее ошибке сообщение и корректно завершиться.

Ситуации некорректных входных данных и выдаваемые сообщения об ошибках указаны в таблице 1.

Таблица 1. Ситуации некорректных входных данных и выдаваемые сообщения об ошибках.

|  |  |
| --- | --- |
| Ошибка | Выдаваемое сообщение |
| Отсутствует файл, заданный пользователем как файл с результатами парсера | Ввод некорректен: не найден разбор текста |
| Отсутствует файл списка имён персонажей | Ввод некорректен: не найден список имён персонажей |
| Отсутствует файл с текстом с пометками о прямой речи | Ввод некорректен: не найден текст |
| Файл менее чем на 50% состоит из букв | Ввод некорректен: текст не является художественным |

Продолжение таблицы 1

|  |  |
| --- | --- |
| Ошибка | Выдаваемое сообщение |
| Во входном графе ребро вместо дуги | Ввод некорректен: во входном графе ребро вместо дуги (строка X) |
| Во входном графе дуга связана с необъявленной ранее вершиной | Ввод некорректен: во входном графе дуга связана с необъявленной ранее вершиной (строка X) |
| Во входном графе более чем одна дуга в строке | Ввод некорректен: во входном графе более чем одна дуга в строке(строка X) |
| Во входном графе объявление вершины следует за любой дугой | Ввод некорректен: во входном графе объявление вершины следует за дугой (строка X) |
| Во входном графе отсутствует “}” | Ввод некорректен: во входном графе отсутствует “}” |
| Во входном графе отсутствует “{” | Ввод некорректен: во входном графе отсутствует “{” |
| Отсутствует “;” в конце строки | Ввод некорректен: отсутствует “;” в конце строки X |
| Отсутствует символ “[“ в строке с дугой или переменной | Ввод некорректен: отсутствует “[“ в строке X |
| Отсутствует символ “]“ в строке с дугой или переменной | Ввод некорректен: отсутствует “]“ в строке X |
| Отсутствует символ открывающейся двойной кавычки после “label=” | Ввод некорректен: неверное оформление спецификаторов в строке X(отсутствует символ открывающейся двойной кавычки после) |

Продолжение таблицы 1

|  |  |
| --- | --- |
| Ошибка | Выдаваемое сообщение |
| Отсутствует символ закрывающейся двойной кавычки после “label=” | Ввод некорректен: неверное оформление спецификаторов в строке X(отсутствует символ закрывающейся двойной кавычки после) |
| На одной строке более одного объявления вершины или ребра | Ввод некорректен: на строке X более одного объявления вершины или ребра |
| Отсутствует файл ключевых слов | Ошибка: ключевые слова не найдены |

## 4.3 Требования к составу и параметрам технических средств

В состав технических средств должен входить персональный компьютер, включающий в себя:

1) процессор Intel Core i3 с тактовой частотой 2 ГГц, не менее;

2) оперативную память объёмом 4 Гб, не менее;

3) операционная система Windows 7 или выше.

## 4.4 Требования к информационной и программной совместимости

На вход должны быть поданы следующие файлы:

1) файл xml, содержащий разбор текста, сделанный в «Stanford CoreNLP» (см. Приложение 3);

2) текстовый файл, содержащий текст в формате html (допустимо использование только тегов <span style="color:#0000ff">, </span>, <p> и </p>) с помеченной в нём прямой речью (прямая речь выделяется тегом по шаблону <span style="color:#0000ff"> Прямая речь </span>);

3) файл dot, написанный по правилам, изложенным в приложении 2.

Выходные файлы располагаются в папке Result, которая находится в корневой папке программы. При запуске программы в папке Result создаётся папка, имя которой соответствует времени и дате (без точек) начала работы программы. Файлы имеют формат html. В файле с названием “Result” находится текст, полученный на вход, с выделенной в нём прямой речью (каждому персонажу соответствует свой цвет текста). Цвета, обозначающие различных персонажей распределены так, что при количестве персонажей до 100 различимы на глаз. Если речь может принадлежать нескольким персонажам, то участок прямой речи подчёркнут, а за ним следует запись всех возможных субъектов в скобках, цвет прямой речи при этом не изменяется от изначального. Если имя субъекта определить не удалось, цвет косвенной речи не меняется. В файле “Key” находятся имена персонажей. Каждое имя окрашено цветом, соответствующим цвету его речи в тексте. Каждое имя начинается с новой строки с заглавной буквы.

Исходные коды программы должны быть реализованы на языке программирования C++, используются библиотеки Qt.

Программа использует данные разбора «Stanford CoreNLP».

# 5 Требования к программной документации

Состав программной документации должен включать в себя следующие документы в печатной форме:

1) техническое задание;

2) технический проект в виде пояснительной записки;

3) программа и методика испытаний;

4) руководство программиста.

Состав программной документации должен включать в себя следующие документы в электронной форме:

1) копии всех документов бумажной формы;

2) рабочая документация;

3) текст программы.

# 6 Стадии и этапы разработки

Разработка должна быть проведена в четыре этапа:

1. техническое задание — 7 марта 2018 года;

2. технический проект — 4 апреля 2018 года;

3. рабочий проект — 8 мая 2018 года;

4. внедрение — 30 мая 2018 года.

# 

# 7 Порядок контроля и приемки

Программа сдается для проверки преподавателю не позднее чем за 3 недели до конца семестра (8 мая 2018 г.).

При обнаружении в программе ошибок и недостатков исполнитель устраняет их в недельный срок и предоставляет программу на повторную проверку.

# Приложение 1

Список ключевых глаголов

|  |  |
| --- | --- |
| acknowledge | moan |
| admit | notify |
| affirm | observe |
| allege | plead |
| announce | postulate |
| argue | pray |
| ask | preach |
| assert | predict |
| assure | promise |
| avow | prophesy |
| boast | propose |
| brag | protest |
| certify | reassure |
| claim | recommend |
| comment | remark |
| complain | remind |
| concede | report |
| confess | request |
| confirm | reveal |
| declare | say |
| decree | speak |
| disclose | specify |
| emphasize | stipulate |
| explain | suggest |
| foretell | swear |
| grumble | teach |
| inform | testify |
| insist | tell |
| joke | threaten |
| lament | vow |
| maintain | warn |

# Приложение 2

Описание формата ввода графа на языке dot, совместимого с программой

Граф должен быть описан на языке dot по следующим правилам.

1) Связи в графе могут являться только дугами и не могут являться рёбрами.

2) Каждая дуга графа имеет подпись по шаблону [label="Субъективные спецификаторы через запятую"].

3) Каждая дуга описывается на новой строке

4) Каждая вершина объявляется до первой дуги в записи.

5) Каждая вершина имеет подпись по шаблону [label="Объективные спецификаторы через запятую"].

6) Каждая вершина объявляется с новой строки.

7) В подписи вершины первый объективный спецификатор – имя.

8) Запись содержит только название графа и символ «{» на первой строке.

9) На последней строке находится только символ «}».

10) Не допускаются пустые строки.

11) Дуга графа показывает как персонаж, обозначенный как вершина-родитель может обращаться к персонажу, обозначенному как вершина-наследник.

12) Допускается использование только символа -> и команды [label = ]. Иные команды не поддерживаются.

Шаблон верно составленного графа на языке dot:

digraph G {

X [label="X, A"];

Y [label="Y, B"];

X -> Y [label="C"];

Y -> X [label="D"];

}

Где X и Y – имена персонажей; A и B – перечень объективных спецификаторов X и Y соответственно; С и D – субъективные спецификаторы Y для X и X для Y соответственно.

# Приложение 3

Минимальные данные файла xml, необходимые для работы программы

Для корректного функционирования программы минимально необходимо сохранить содержимое следующих тегов внутри тега < token id=""> </token>, расположенных в теге :

1) <word></word>;

2) <lemma></lemma>;

3) <POS></POS>.

Также необходимо сохранить теги <sentence id=""> </sentence> внутри тега <sentences></sentences>.

Для функционирования сужения списка по местоимениям необходимо также сохранить в xml разбор coreference. Если данные этого разбора отсутствуют, то сужение списка не происходит. Для работы функции сужения списка персонажей по связи необходимо всё содержимое тега <coreference> </coreference>.