

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Волгоградский государственный технический университет»

Факультет Электроники и вычислительной техники
Специальность (направление) Программная инженерия
Кафедра Программного обеспечения автоматизированных систем
Дисциплина Эволюция и сопровождение программного обеспечения

Утверждаю
Зав. кафедрой ПОАС Орлова Ю.А.

«_____» _____ 20 ____ г.

ЗАДАНИЕ
на курсовой проект

Студент Полозов Андрей Сергеевич
(фамилия, имя, отчество)

Группа ПрИн-467

1. Тема: «Реализация и внедрение изменения в программный проект с открытым исходным кодом»

Утверждена приказом от 22.09.2017 № 1247-ст

2. Срок представления проекта к защите «_____» _____ 20 ____ г.

3. Содержание курсового проекта: постановка задачи, спецификация решения, тестовые инструкции, линия времени разработки, перечень коммитов, содержание коммитов

4. Дата выдачи задания «_____» _____ 20 ____ г.

Руководитель работы _____ Сычев О. А.

Задание принял к исполнению _____ Полозов А.С.
«_____» _____ 20 ____ г.

Содержание

1. Постановка задачи.....	3
2. Спецификация решения	3
3. Инструкция по тестированию.....	6
4. Перечень созданных коммитов	7
5. Содержание коммитов.....	8

1. Постановка задачи

Вопрос типа Correct writing настроен на проверку тестов для английского языка. Если создать вопрос данного типа, включить проверку опечаток и в ответе указать большое количество похожих слов, редакционное расстояние Дамерау-Левенштейна между которыми стремится к единице, то система при проверке ответа на такой вопрос будет зависать или совсем откажет. Поэтому необходимо предупреждать пользователя, создающего вопрос о возможных зависаниях.

Задачей является модифицировать модуль Correct Writing для предупреждения пользователя при создании вопроса с включенной проверкой опечаток с ответом, который содержит похожие слова, о возможных зависаниях системы во время проверки теста.

2. Спецификация решения

Для решения данной задачи необходимо определить проблемные ситуации. С помощью ручного тестирования были определены случаи, при которых зависает система. Они сведены в таблицу 1.

В качестве ответа преподавателя использовались ответы, содержащие определенное количество похожих слов с расстоянием Дамерау-Левенштейна равным одному (qnowledge wnowledge ennowledge rnowledge tnowledge ynowledge) и двум (qknowledgeq wknowledgew enowedgee rnowledger tnowledget ynowledgey).

В качестве ответа учащегося использовались ответы преподавателя, но расстояние между каждым словом ответа учащегося и словом из ответа преподавателя было равно одному (aknowledge anowledge aknowledge aknowledge aknowledge aknowledge aknowledge) и двум (aknowledgea anowledgea aknowledgea aknowledgea aknowledgea).

Были использованы длинные слова (9 букв), чтобы протестировать проверку опечаток вопроса с разными порогами прохождения ошибок (lexical error threshold).

Были проведены тесты в которых ответ учащегося был на одно слово короче ответа преподавателя и в которых ответ учащегося был на одно и более слов длиннее ответа преподавателя (в таких случаях в качестве лишних слов использовались так же похожие слова).

Порог (Threshold)	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Случаи отказа системы (Кол-во слов в ответе преподавателя / кол-во слов в ответе учащегося)	7/8 dist = 1	7/8 dist = 1	7/8 dist = 1	7/9 dist = 1	7/9 dist = 1	7/9 dist = 1	7/9 dist = 1

Таблица 1. Результаты проверки зависаний системы

В результате эксперимента было выявлено, что система отказывает при наличии 7 похожих слов в ответе преподавателя и 8 похожих словах в ответе учащегося (наихудший результат).

При наличии 7 похожих слов в ответе преподавателя и 7 похожих словах в ответе учащегося система зависает. Расстояние ДЛ между похожими словами ответа преподавателя — 1. Расстояние ДЛ между похожими словами ответа учащегося — 1. Расстояние между соответствующими словами ответа преподавателя и словами ответа учащегося — 1. Были так же протестированы ответы со словами, длина которых варьируется от 2 до 10 букв. Результаты тестирования совпадают с предыдущими. Из этого следует, что необходимо запрещать сохранять вопросы и выводить сообщение об ошибке при сохранении вопроса с ответом из 7 и более похожих слов (на случай если учащийся введет не 7, а 8 похожих слов) при условии, что порог прохождения ошибки среди похожих слов (отношение количества опечаток к длине слова) в ответе преподавателя выше, чем выбран в настройках вопроса (чтобы дать возможность сохранить ответ при выбранной низкой чувствительности к

опечаткам, так как в этом случае система пропустит совершенные опечатки и не будет зависеть).

В модуле Correct Writing содержатся файлы, в которых находятся все выдаваемые системой сообщения.

В файлы MOODLE_DIR/question/type/correctwriting/lang/en и MOODLE_DIR/question/type/correctwriting/lang/ru были соответственно добавлены следующие строки:

```
$string['systemmayfreeze'] = 'System may freeze while processing typos if answer contains more than 6 very similar words. Turn off typos analysis or lower the error threshold in symbols!';
```

```
$string['systemmayfreeze'] = 'Система может зависнуть при обработке опечаток, если ответ содержит более 6 очень похожих слов. Выключите анализ опечаток или опустите порог ошибки в символах!';
```

которые позволяют получить определенное сообщение по ключу „systemmayfreeze“.

В файл MOODLE_DIR/question/type/correctwriting/edit_correct_writing_form, который отвечает за редактирование формы при создании вопроса, в функцию проверки правильности заполнения вопроса был добавлен код, который проверяет наличие в ответе учащегося большого количества похожих слов. Была реализована функция has_many_similar_words, которая проверяет в массиве токенов присутствие большого количества похожих токенов. Она возвращает истинное значение, когда в массиве токенов содержится более 6 токенов, редакционное расстояние которых равно одному или двум. Изменения вступили в силу после сброса кэша системы Moodle, так как Moodle кэширует сообщения.

В результате проделанной работы мы получили во время редактирования вопроса вывод сообщения о том, что при включенной проверке опечаток и наличии более 6 очень похожих слов в ответе

преподавателя возможны зависания системы при тестировании и невозможность сохранить вопрос (рис 1 и рис 2).

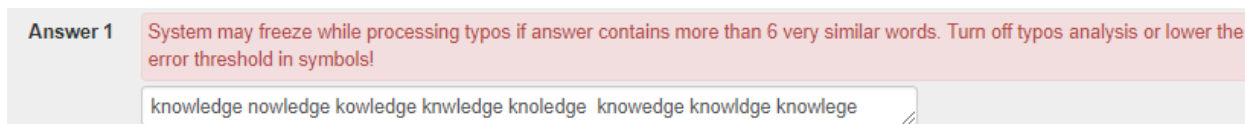


Рис 1. Сообщение о возможных зависаниях системы (на английском)

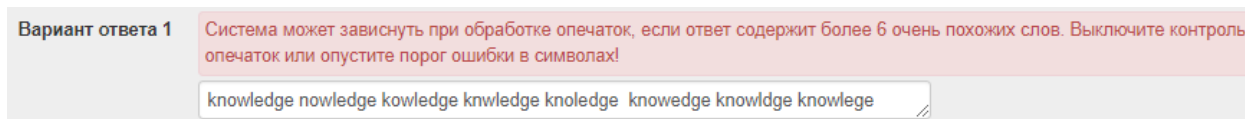


Рис 2. Сообщение о возможных зависаниях системы (на русском)

3. Инструкция по тестированию

1. Откройте главную страницу Moodle
2. Войдите в учетную запись в качестве
3. Перейдите на вкладку «Домашняя страница»
4. Откройте выпадающее меню с настройками и выберите пункт «Банк вопросов»
5. Создайте вопрос типа «Correct writing»
6. Открыть страницу настройки вопроса Correct Writing
7. В настройках вопроса включите проверку опечаток
8. В одном из ответов в поле ответа на вопрос введите «knowledge wnowledge enowledge knowledge knowledge ynowledge»;
9. Система выведет сообщение «System may freeze while processing typos if answer contains more than 6 very similar words. Turn off typos analysis or lower the error threshold in symbols!» («Система может зависнуть при обработке опечаток, если ответ содержит более 6 очень похожих слов.

Выключите анализ опечаток или опустите порог ошибки в символах!») и не даст сохранить вопрос.

10. Если отключить проверку ошибок или ввести менее 7 слов, то система не будет выводить такое сообщение и даст сохранить вопрос.

11. В одном из ответов в поле ответа на вопрос введите «qnowledgea wknowledgea ennowledgea rnowledgea tnowledgea ynowledgea»;

12. Система выведет сообщение «System may freeze while processing typos if answer contains more than 6 very similar words. Turn off typos analysis Turn off typos analysis or lower the error threshold in symbols!!» («Система может зависнуть при обработке опечаток, если ответ содержит более 6 очень похожих слов. Выключите контроль опечаток или опустите порог ошибки в символах!») и не даст сохранить вопрос

13. Если отключить проверку ошибок или ввести менее 7 слов, то система не будет выводить такое сообщение и даст сохранить вопрос

14. В одном из ответов в поле ответа на вопрос введите «qknowledgeaq wknowledgeaw enknowledgeae rknowledgear tknowledgeat ynowledgeay»;

15. Система не выведет предупреждающее сообщение (так как слова слишком разные) и разрешит сохранить вопрос.

4. Перечень созданных коммитов

Хэш	Сообщение
d017f6c	Добавил сообщение о том, что система может зависнуть при ключенном контроле опечаток, если слова очень похожи
ec35849	Добавил функцию has_many_similar_words - проверка ответа на множество похожих слов
8d3f549	Добавил вывод сообщения с предупреждением о возможном зависании контроля опечаток

5. Содержание коммитов

d017f6c

question/type/correctwriting/lang/en/qtype_correctwriting.php

Двустороннее сравнение

Просмотреть файл

Комментарий

ИЗМЕНЁН

...

...

104 104 \$string['syntaxanalyzerisincompatiblewithlang'] = 'Tree coverage analysis requires language, that can be parsed. Please, try another lan

105 105 \$string['syntax_analyzer'] = 'Tree coverage analysis';

106 106 \$string['syntax_analyzer_help'] = 'Tree coverage analysis can be used if language supports parsing for reducing amount of mistakes, show

107 107 +\$string['systemmayfreeze'] = 'System may freeze while processing typos if answer contains more than 6 very similar words. Turn off typos

107 108 \$string['questioneditingheading'] = 'Question editing settings';

108 109 \$string['tokens'] = 'Tokens';

109 110 \$string['toobigfloatvalue'] = 'This value should be no more than {\$a}';

...

question/type/correctwriting/lang/ru/qtype_correctwriting.php

Двустороннее сравнение

Просмотреть файл

Комментарий

ИЗМЕНЁН

...

...

102 102 \$string['syntaxanalyzerisincompatiblewithlang'] = 'Анализ покрытия синтаксического дерева требует, чтобы по языку можно было построить т

103 103 \$string['syntax_analyzer'] = 'Анализ покрытия синтаксического дерева';

104 104 \$string['syntax_analyzer_help'] = 'Анализ покрытия синтаксического дерева позволяет уменьшить количество ошибок выдаваемых студенту за с

105 105 +\$string['systemmayfreeze'] = 'Система может зависнуть при обработке опечаток, если ответ содержит более 6 очень похожих слов. Выключите

105 106 \$string['tokens'] = 'Лексемы';

106 107 \$string['toobigfloatvalue'] = 'Значение поля не должно превышать {\$a}';

107 108 \$string['toosmallfloatvalue'] = 'Значение поля не должно быть менее {\$a}';

...

ec35849

question/type/correctwriting/edit_correctwriting_form.php

Двустороннее сравнение

Просмотреть файл

Комментарий

ИЗМЕНЁН

...

...

25 25 defined('MOODLE_INTERNAL') || die();

26 26 require_once(\$CFG->dirroot . '/question/type/shortanswer/edit_shortanswer_form.php');

27 27 require_once(\$CFG->dirroot . '/blocks/formal_langs/block_formal_langs.php');

28 28 +require_once(\$CFG->dirroot . '/blocks/formal_langs/classes/comparing_options.php');

29 29 +require_once(\$CFG->dirroot . '/blocks/formal_langs/tokens_base.php');

28 30

29 31 //require_once(\$CFG->dirroot . '/question/type/correctwriting/enumeditor_form/correctwriting_text_and_button.php');

30 32 /**


```

921 923
922 924      /**
925 +      * Function returns threshold whether array of tokens has many similar tokens or not
926 +      * @param array $tokens Array of block_formal_langs_token_base objects
927 +      * @param boolean $usecase Boolean flag is true, if case is important
928 +      * @return float minus one, if array of tokens doesn't have many similar tokens, or threshold
929 +      * (the ratio of similar words' length to Damerau-Levenshtein distance between similar words),
930 +      * if it does.
931 +      */
932 +      public function has_many_similar_words($tokens, $usecase) {
933 +          // Is case important or not.
934 +          $options = new block_formal_langs_comparing_options();
935 +          $options->usecase = $usecase;
936 +          // For each token in array.
937 +          for ($i = 0; $i < count($tokens); $i++) {
938 +              $short_distance_word_counter = 0;
939 +              for ($j = 0; $j < count($tokens); $j++) {
940 +                  $distance = block_formal_langs_token_base::damerau_levenshtein($tokens[$i]->value(), $tokens[$j]->value(), $options);
941 +                  // If tokens are very similar.
942 +                  if ($distance == 1) {
943 +                      $short_distance_word_counter++;
944 +                  }
945 +              }
946 +              // If there are 7 and more similar tokens in array.
947 +              if ($short_distance_word_counter >= 6) {
948 +                  return 1/core_text::strlen($tokens[$i]->value());
949 +              }
950 +          }
951 +          return -1;
952 +      }
953 +
954 +      /**
955 +      * Returns value for case when processed string returns value

```

8d3f549

question/type/correctwriting/edit_correctwriting_form.php

ИЗМЕНЁН

Двустороннее сравнение

Просмотреть файл

Комментарий

...

```

900 900            if (count($tokens) > count($descriptions)) {
901 901                $mesg = get_string('writemoredescriptions', 'qtype_correctwriting');
902 902            }
903 +            $threshold = $this->has_many_similar_words($tokens, $data['usecase']);
904 +            // If lexical analyzer is enabled and threshold in correct answer is lower than lexical error threshold.
905 +            if ($islexicalanalyzereenabled && $threshold > 0 && $threshold < $data['lexicalerrorthreshold']) {
906 +                $mesg = get_string('systemmayfreeze', 'qtype_correctwriting');
907 +            }
908
903 908            if (count($tokens) < count($descriptions)) {
904 909                $mesg = get_string('writelessdescriptions', 'qtype_correctwriting');
905 910            }

```

...

```

921 926            return $errors;
922 927        }
923 928
924 928        - /**
929 +        /**
930
925 930        * Function returns threshold whether array of tokens has many similar tokens or not
926 931        * @param array $tokens Array of block_formal_langs_token_base objects
927 932        * @param boolean $usecase Boolean flag is true, if case is important

```