

ИНФОРМАТИКА И УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМАХ

УДК 004.054

И.Б. Зарубин, А.Д. Филинских

МЕТОДИКА ОЦЕНКИ ПОЛНОТЫ РЕГРЕССИОННОГО ТЕСТИРОВАНИЯ С НОРМИРОВКОЙ ПО ВЕСОВЫМ КОЭФФИЦИЕНТАМ

Нижегородский государственный технический университет им. Р.Е. Алексеева

Рассмотрена методика оценки полноты проведенного регрессионного тестирования с применением нормировки использованных принципов отбора тестов по весовым коэффициентам. Описаны преимущества и ограничения данной методики, условия ее успешного использования для оценки влияния внесенных в информационную систему изменений на компоненты, не подвергавшиеся модификации. Представленная методика оценки полноты выполнения РТ позволяет гибко подходить к очередности выполнения и количеству тестов, сэкономив ресурсы, но в то же время непригодна для формирования перечня тестов для РТ и не учитывает возможность включения в разные группы РТ одних и тех же тестов.

Ключевые слова: регрессионное тестирование, коэффициент полноты выполнения регрессионного тестирования, подходы регрессионного тестирования.

Введение

Успешное использование информационных систем (ИС) [1], широко представленных во многих сферах человеческой деятельности, предполагает корректность и безотказность их работы, обеспечиваемые «тестированием» – определенными действиями, нацеленными на поиск ошибок [2]. Существует значительное количество методик поиска ошибок на всех этапах жизненного цикла ИС [3,4]. Регрессионное тестирование (РТ) является одной из самых важных и в то же время сложных в применении методик [5]. Сферой ее применения является обнаружение ошибок в функционале, который не подвергался изменениям непосредственно, но мог быть поврежден или утратил актуальность по причине добавления в разрабатываемую ИС новых возможностей.

Важность РТ обуславливается тем, что этот вид тестирования должен применяться каждый раз, когда в ИС добавляется сколько-нибудь значимый новый функционал или в случае серьезных доработок в существующем функционале, поскольку в сложных многокомпонентных ИС зачастую крайне затруднительно отследить взаимосвязи между компонентами [6]. Помимо этого, РТ призвано не допустить ухудшения качества ИС, которые, как правило, уже находятся в промышленной эксплуатации. Сложность РТ в применении объясняется тем, что для понимания возможных зависимостей компонентов необходимо очень хорошо разбираться в архитектуре ИС, уверенно владеть теорией тестирования и уметь пользоваться различными подходами к отбору проверочных сценариев для РТ. Число специалистов, которые способны оперативно разобраться в архитектуре многокомпонентной системы и уверенно владеют теорией тестирования ИС, включая подходы к отбору проверочных сценариев для РТ, очень невелико, при этом количество разрабатываемых ИС непрерывно растет. Наиболее популярные в настоящее время модели разработки ИС, основанные на Agile [7], предполагают сравнительно короткие итерации (Sprints), длящиеся 2-3 недели, в течение которых выполнить необходимые виды тестирования бывает затруднительно, особенно в случае недостатка ресурсов. РТ необходимо выполнять после успешного

выполнения «дымового», интеграционного, функционального и системного [8] тестирования, в условиях ограниченности сроков.

Соответственно, существует объективная необходимость в удобном и простом инструменте оценки степени полноты РТ, который позволял бы при этом выделять наиболее важные проверочные сценарии, выбранные с помощью различных подходов для каждой конкретной итерации РТ.

Оценка полноты регрессионного тестирования

Оценить собственно полноту выполнения РТ несложно – необходимо вычислить отношение количества выполненных тестов к общему количеству тестов (1):

$$C = \frac{Q_d}{Q_a} \times 100\%, \quad (1)$$

где C – коэффициент полноты выполнения РТ; Q_d – количество выполненных тестов; Q_a – общее количество тестов, отобранных для РТ.

Очевидно, что данная методика оценки полноты выполнения РТ подходит только для ситуаций, когда все проверочные сценарии, отобранные для РТ, равнозначны с точки зрения важности результатов проверки или отобраны с использованием одного из подходов РТ.

Оценка полноты выполнения регрессионного тестирования в случае отбора тестов с использованием двух методик

На практике отбор проверочных сценариев для РТ не ограничивается использованием одного подхода. Рассмотрим, как можно оценить полноту выполнения РТ, тесты для которого были отобраны с помощью двух подходов – отбор тестов высокой важности и тестов, которые были неуспешно пройдены в процессе предыдущего РТ. В случае равной важности тестов, отобранных с помощью разных методик, методику оценки полноты выполнения регрессионного тестирования можно выразить как отношение суммы *выполненных тестов*, отобранных обоими подходами к *общей сумме тестов*, отобранных обоими подходами (2):

$$C = \frac{Q_{d1} + Q_{d2}}{Q_{a1} + Q_{a2}} \times 100\%, \quad (2)$$

где C – коэффициент полноты выполнения РТ; Q_{d1} – количество выполненных тестов, отобранных при помощи подхода 1; Q_{d2} – количество выполненных тестов, отобранных при помощи подхода 2; Q_{a1} – общее количество тестов, отобранных при помощи подхода 1 для РТ; Q_{a2} – общее количество тестов, отобранных при помощи подхода 2 для РТ.

Для учета важности тестов необходимо ввести нормировку для каждого из подходов, с помощью которых отбирались проверочные сценарии для РТ.

Оценка полноты выполнения регрессионного тестирования в случае отбора тестов с использованием двух методик и нормировкой методик отбора

Для учета важности проверочных сценариев, отобранных различными подходами для РТ, предлагается оценивать полноту выполнения РТ с нормировкой подходов отбора с помощью весовых коэффициентов (ВК), т.е., необходимо вычислить полноту выполнения тестов, отобранных для РТ, используя формулу (1) отдельно для каждого из подходов отбора, а затем выполнить нормировку, используя ВК.

Например, предположим, что тесты для РТ были отобраны с помощью двух подходов – отбор тестов высокой важности и тестов, которые были пройдены неуспешно в процессе предыдущего РТ. Полноту выполнения РТ в данном случае можно рассчитать следующим образом (3):

$$C_a = \frac{w_1 \times C_1 + w_2 \times C_2}{w_1 + w_2}. \quad (3)$$

где C_a – общий коэффициент выполнения РТ; w_1 – ВК для тестов высокой важности; C_1 – коэффициент выполнения тестов высокой важности, рассчитанный по формуле (1); w_2 – ВК для тестов, проваленных в предыдущий запуск РТ; C_2 – коэффициент выполнения тестов, проваленных в предыдущий запуск РТ, рассчитанный по формуле (1). Для удобства использования и приближенных оценок значимости подходов отбора удобно использовать такие значения весовых коэффициентов, которые в сумме давали бы 1 (единицу). В этом случае формула (3) примет следующий вид (4):

$$C_a = (w_1 \times C_1 + w_2 \times C_2). \quad (4)$$

Предположим, что для РТ были отобраны 40 тестов высокого приоритета и 50 тестов, которые были пройдены неуспешно в предыдущей итерации РТ. При этом, исходя из нужд и требований проекта на разработку данной ИС, запуск тестов высокого приоритета более важен с точки зрения успешности поиска ошибок и, как следствие, более важен с точки зрения коммерческого успеха ИС. Соответственно, ВК у тестов высокого приоритета положим равным 0,75, а ВК тестов, которые были пройдены неуспешно в предыдущей итерации РТ – 0,25. Таким образом, получаем формулу 4 вида (5):

$$C_a = (0,75 \times C_1 + 0,25 \times C_2). \quad (5)$$

В условиях недостатка ресурсов для выполнения тестирования возможны ситуации, когда отобранные для РТ не могут быть выполнены полностью за отведенный для РТ срок или же выполнение РТ вскрыло низкую эффективность регрессионных тестов с точки зрения поиска ошибок в ИС: например, при выполнении 50 % отобранных тестов не выявлено ни одной ошибки. В этом случае можно определить порог полноты выполнения РТ, отличный от 100 %, по достижению которого инженеры-тестировщики смогут переключиться на другие виды тестирования, необходимые для полноценной проверки качества ИС. Предположим, что в вышеприведенном примере необходимый порог полноты выполнения РТ принят за 90 %. Исходя из этого, можно рассчитать, что для прохождения порога полноты выполнения РТ (порог полноты выполнения также можно назвать критерием завершенности [9]) достаточно выполнить все тесты из первой группы и 60 % тестов (или, в абсолютных числах, 30 тестов из 50) из группы 2 (6):

$$90\% = (0,75 \times 100\% + 0,25 \times 60\%). \quad (6)$$

Таким образом, для достижения необходимой полноты выполнения РТ достаточно выполнить в сумме 70 тестов. Очевидно, что в случае, когда запуск любого из тестов занимает приблизительно одинаковое время, наиболее быстрый путь достижения необходимого порога выполнения РТ – это выполнение тестов из группы с большим ВК (рис. 1).

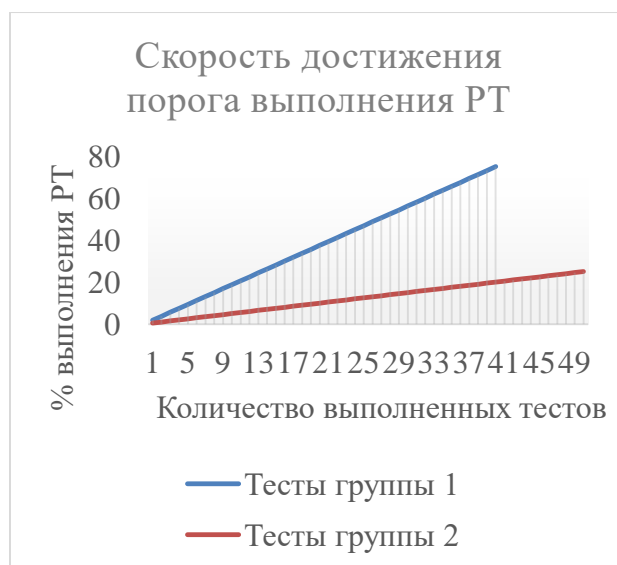


Рис. 1. Скорость достижения порога выполнения РТ

Исходя из практики, тесты, особенно отобранные с использованием различных подходов, не являются равнозначными – одни из них требуют больше времени на исполнение, другие – сложной предварительной настройки или дополнительных устройств; в некоторых необходимо выполнить большее количество действий. Именно в таких ситуациях проявляется удобство приведенной выше методики – можно применять гибкий подход к отбору выполняемых тестов (выполнив, например, все тесты из группы 2, исключив из выполнения 4 теста из группы 1, так как выполнение одно теста из группы 1 дает прирост выполнения РТ в 2,5 %), который позволит достичь необходимого порога полноты выполнения с использованием наименьших ресурсов.

Оценка полноты выполнения регрессионного тестирования в случае отбора тестов с использованием произвольного количества подходов отбора и нормировкой методик отбора

Существует достаточно большое количество подходов отбора тестов для РТ [10]. На практике все они одновременно не применяются, но использование 3-4 подходов встречается часто. Исходя из этого, необходимо модифицировать формулу (3) для произвольного количества подходов отбора тестов для РТ (7):

$$C_a = \frac{w_1 \times C_1 + w_2 \times C_2 + w_n \times C_n}{w_1 + w_2 + \dots + w_n}, \quad (7)$$

где C_a – общий коэффициент выполнения РТ; w_1 – ВК для тестов высокой важности; C_1 – коэффициент выполнения группы тестов 1, рассчитанный по формуле (1); w_2 – ВК для тестов группы 2; C_2 – коэффициент выполнения группы тестов 2, рассчитанный по формуле (1); w_n – ВК для тестов группы n ; C_n – коэффициент выполнения тестов, проваленных в предыдущий запуск РТ, рассчитанный по формуле (1).

Методика распределения весовых коэффициентов для отобранных регрессионных тестов в многокомпонентных информационных системах

Корректно выбранные ВК способствуют выработке наиболее выгодной, с точки зрения экономии ресурсов, последовательности выполнения регрессионных тестов. Исходя из этого, очевидно, что выбор корректных ВК приобретает весьма важное значение. Для стандартных подходов отбора тестов для РТ, ВК могут определяться в проектной документации, описывающей процесс проверки качества и поиска ошибок ИС. Они могут быть или постоянными величинами, подобранными на ранних этапах реализации проекта, или некоторым образом варьироваться в зависимости от нужд конкретного запуска РТ.

В настоящее время все большую популярность приобретает реализации ИС с использованием микро-сервисной архитектуры [11]. ИС здесь состоит из нескольких (а иногда и из большого числа) взаимодействующих между собой модулей; добавление новых возможностей в систему представляет собой, как правило, доработку одного-двух модулей, при этом остальные модули остаются неизменными. Необходимо отметить, что один модуль, в который были внесены изменения, может взаимодействовать или, так или иначе, оказывать влияние на другие модули, которые не были изменены. Поэтому вопрос о необходимости проведения РТ и правильного выбора тестов для ИС, реализованных в рамках микро-сервисной архитектуры, очень важен.

Рассмотрим упрощенную схему абстрактной ИС, состоящей из нескольких компонентов (рис. 2).

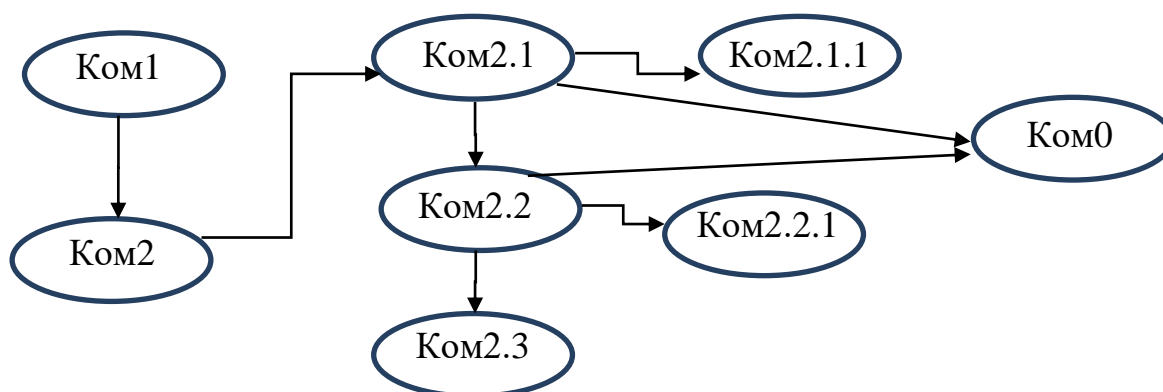


Рис. 2. Пример взаимодействия компонентов в многомодульной ИС

Предположим, что в процессе добавления в ИС новой возможности были внесены изменения в компонент Ком2. Исходя из этого, в соответствии с архитектурой системы, представленной на рис. 2, в рамках РТ необходимо в первую очередь выполнить тесты, проверяющие наличие регресса в компонентах Ком1 и Ком2.1 – т.е., выбрать для тестов по компонентам Ком1 и Ком2.1 более высокий ВК. При этом, из рис. 2 видно, что внесение изменений в компонент Ком2.1 требует проведения более детального РТ для компонентов Ком2, Ком2.1.1, Ком0 и Ком2.2, не отменяя, при этом, необходимости проведения РТ с более низкими ВК для других компонентов системы, так как, как мы видим из архитектуры ИС, практически все компоненты системы взаимосвязаны.

Таким образом, очевидно, что специалист, производящий отбор тестов для РТ (как правило, это руководитель отдела тестирования – QA/Test lead) должен обладать знаниями по архитектуре системы практически на уровне архитектора ИС. Иными словами, перечень и порядок проведения РТ и принципы отбора тестов для него должны быть согласованы с ответственными лицами в отделе разработки и архитектором системы. Кроме того, особняком стоит вопрос о методах оформления тестов компонентов и системы в целом таким образом, чтобы иметь возможность оперативно отобрать тесты «покомпонентно» или даже во взаимодействии со связанными модулями. Стандартных для любого теста полей для указанных целей уже явно недостаточно. В качестве быстрой и наименее трудозатратной меры можно предложить, например, введение в тест дополнительных полей с указанием названия компонента, к которому относится данный тест, и, возможно, названия непосредственно связанных с данным компонентом модулей.

Обоснование достаточной величины полноты выполнения РТ

На первый взгляд, значение полноты выполнения РТ, отличной от 100 %, выглядит ошибочным: нет гарантии, что при определенной в начале РТ величине в 90 % в невыполненных 10 % не окажется ошибок, критичных для корректного функционирования системы. Возможна ситуация, когда в РТ выбрали 10 тестов, 9 (90 % всего РТ) выполнили успешно, а в невыполненном 1 тесте оказалась ошибка, которая блокирует какую-либо критическую функциональность ИС. Необходимо понимать, что критически важные возможности ИС должны проверяться перед запуском РТ другими видами тестирования – модульным, «дымовым», интеграционным и функциональным. Вероятность того, что критически важная ошибка избежит нахождения вышеупомянутыми тестами, которые корректно сформулированы и достаточно полно проверяют свои уровни ИС, очень невелика. Обнаружение таких ошибок в процессе выполнения РТ свидетельствует о погрешностях в организации процессов тестирования и разработки и поддержания тестовой документации.

Таким образом, можно сказать, что основным результатом запуска РТ являются некоторое количество найденных ошибок средней важности – видимых заказчиком/конечным пользователем, но не блокирующих функционирование ИС в целом и наиболее значимых ее компонентов в частности.

Грамотное РТ состоит из этапа планирования с выделением тестируемых областей ИС, определения не только важности, но и приоритета тестов, отобранных для РТ, выбора правильных весов для подхода отбора тестов и групп тестов, далее – в своевременном запуске РТ (например, не имеет смысла начинать выполнение РТ раньше функционального тестирования: новые возможности, добавленные в систему, могут содержать такое количество ошибок, что заблокируют возможность передачи новой версии ИС заказчику и, следовательно, до исправления всех важных ошибок запуск РТ не имеет смысла), непрерывном анализе результатов РТ и своевременно принятого решения о прекращении выполнения РТ. Причинами последнего могут быть достижение желаемых результатов или признание невозможности продолжения РТ в силу его нецелесообразности.

С точки зрения обоснования величины полноты выполнения РТ, анализ результатов имеет важное значение. Нельзя однозначно утверждать, что количество найденных в процессе РТ ошибок линейно зависит от количества запущенных тестов, отобранных для РТ в случаях использования нескольких подходов для отбора. Тем не менее, разделив выбранные тесты на несколько однородных по составу групп (например, тесты одного уровня важности для какого-либо конкретного компонента) и выполнив 50 % случайно выбранных тестов в каждой группе, можно сделать предварительный вывод об обнаружении ошибок при выполнении оставшихся 50 % тестов в каждой из групп и акцентировать внимание на выполнении именно тех групп тестов, в которых обнаружено большее количество ошибок – в соответствии с принципом «Вероятность того, что в некоторой части программы остались необнаруженные ошибки, прямо пропорциональна количеству уже обнаруженных там ошибок» [12].

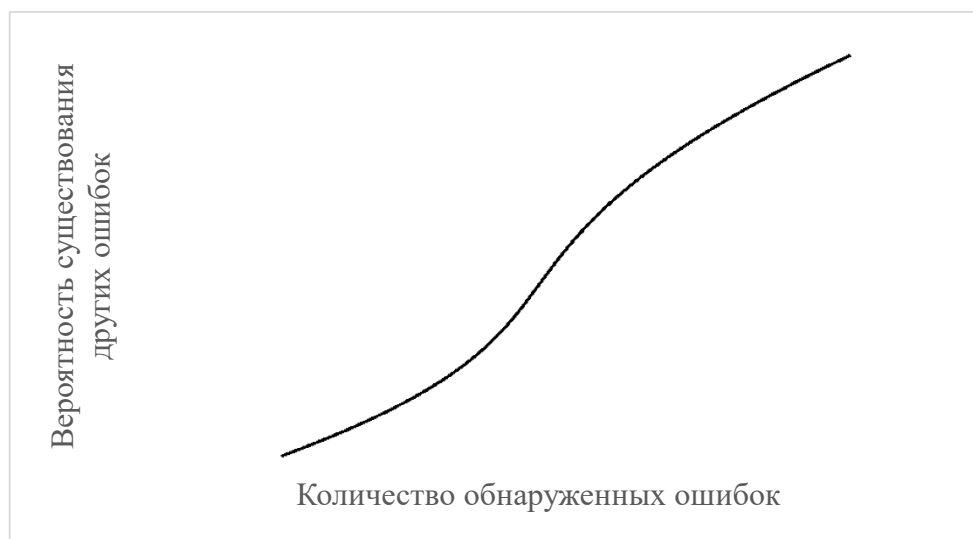


Рис. 3. Соотношение между количеством обнаруженных и необнаруженных ошибок

Очевидно, что применение указанного на рис. 3 принципа оценки количества необнаруженных ошибок в незапущенных тестах некоей однородной группы тестов возможно только в случае запуска некоего значимого количества тестов внутри этой группы – 1-2 тестов для вывода будет недостаточно. Например, для регрессионного тестирования отобрана группа из 20 тестов на модуль вычисления KQI. Все тесты имеют равную важность, сходное количество шагов, взаимодействуют с одинаковым количеством смежных модулей и т.д. Выполнив из этого набора 10 случайно выбранных тестов и обнаружив с помощью них одну ошибку, можно с уверенностью утверждать, что, выполнив оставшиеся 10 тестов, 10 ошибок

обнаружено не будет. И, наоборот, выполнив 10 тестов и не обнаружив ни одной ошибки, предпочтительно сосредоточить усилия по тестированию на других группах тестов, отобранных для РТ, при выполнении которых были найдены ошибки. С другой стороны, наличие значительного количества регрессионных тестов, в которых не обнаруживаются ошибки, может сигнализировать о некорректных подходах к отбору тестов для РТ или недобросовестности инженеров по тестированию.

При составлении проектной документации [13] на разрабатываемую ИС в ней зачастую предусматривается наличие некоторого незначительного количества некритических ошибок при передаче ИС в промышленную эксплуатацию (конечно, при условии дальнейшей поддержки этой ИС и, следовательно, исправления известных или найденных в процессе эксплуатации ошибок). Если при выполнении РТ будет очевидно, что в случае продолжения выполнения РТ заложенный в проектной документации порог количества ошибок в ИС превышен не будет, имеет смысл переключиться на другие виды тестирования.

Исходя из вышеизложенного, можно предложить корректировать ВК в процессе выполнения РТ тех групп тестов, в которых обнаружены ошибки в сторону увеличения, уменьшая при этом (или вовсе – обнуляя) ВК групп тестов, выполнение которых не приводит к обнаружению ошибок.

Вывод

Предлагается следующий принцип использования методики оценки полноты выполнения РТ:

- 1) первый этап: выполнение, в соответствии с приоритетами и весовыми коэффициентами 50 % тестов из каждой отдельно взятой однородной группы регрессионных тестов;
- 2) второй этап: анализ полученных результатов, перенормировка групп тестов с повышением ВК у групп тестов, при запуске которых было выявлено более количество ошибок;
- 3) третий этап: запуск оставшихся отобранных для РТ тестов до момента достижения необходимого порога полноты выполнения РТ;
- 4) четвертый этап: анализ результатов полного цикла РТ, планирование ВК для следующих циклов РТ в соответствии с количеством найденных ошибок в каждой группе тестов.

Представленная методика оценки полноты выполнения РТ позволяет оперативно оценить степень завершенности текущей итерации РТ и выделить тесты, выполнение которых окажет большее влияние на достижение порога выполнения РТ, а также предоставляет возможность гибко подходить к очередности выполнения и количеству тестов. Это позволяет сократить время выполнения РТ, сэкономив при этом другие ресурсы. При этом данная методика непригодна для формирования перечня тестов для РТ, что является очень актуальным для проектов, в которых разработано значительное количество тестов различными методами. Кроме того, данная методика не учитывает возможность включения в разные группы РТ одних и тех же тестов.

Библиографический список

1. **Кузнецов, С.Д.** Большая Российская энциклопедия / С.Д. Кузнецов. – Режим доступа: [https:// bigenc.ru/technology_and_technique/text/3444940](https://bigenc.ru/technology_and_technique/text/3444940).
2. **Майерс, Г.** Искусство тестирования программ / Г. Майерс, Т. Баджетт, К. Сандлер. – М., 2019. – 272 с.
3. **Савин, Р.** Тестирование Дот Ком или пособие по жестокому обращению с багами в интернет-стартапах / Р. Савин. – М., 2017. – 312 с.

4. Национальный стандарт Российской Федерации ГОСТ Р 56136-2014 Управление жизненным циклом продукции военного назначения. Термины и определения. Переиздание 11.2016. – 20 с.
5. **Старолетов, С.М.** Основы тестирования и верификации программного обеспечения / С.М. Старолетов. – М.: Лань, 2018. – 344 с.
6. **Зарубин, И.Б.** КОГРАФ-2019: сборник материалов 29-й Всероссийской научно-практической конференции по графическим информационным технологиям и системам / И.Б. Зарубин, А.Д. Филинских. – Н. Новгород: НГТУ им. Р.Е. Алексеева, 2019. – 218 с.
7. **McHugh, M.** An Agile Implementation within a Medical Device Software Organisation. Software Process Improvement and Capability Determination / M. McHugh, F. McCaffery, G. Coady // Communications in Computer and Information Science, 2014. – P. 190-201.
8. **Савин, Р.** Тестирование Дот Ком или пособие по жестокому обращению с багами в интернет-стартапах / Р. Савин. – М., 2017. – 312 с.
9. **Майерс, Г.** Искусство тестирования программ / Г. Майерс, Т. Баджетт, К. Сандлер. – М., 2019. – 272 с.
10. **Yoo, S.** Regression testing minimization, selection and prioritization: A survey. King`s College London / Yoo S., Harman M, 2007. – 60 с.
11. **Balalaie, A.** Microservices Architecture Enables DevOps: Migration to a Cloud-Native Architecture / A. Balalaie, A. Heydarnoori, P. Jamshidi // IEEE Software: journal. – 2016. – V. 33. – №. 3. – P. 42-52.
12. **Майерс, Г.** Искусство тестирования программ / Г. Майерс, Т. Баджетт, К. Сандлер. 2019. – 272 с.
13. **Филинских, А.Д.** Управление проектными данными / А.Д. Филинских, И.Н. Мерзляков. – Н. Новгород: НГТУ им. Р.Е. Алексеева, 2017. – 93 с.

*Дата поступления
в редакцию: 11.10.2019*

I.B. Zarubin, A.D. Filinskih

METHOD OF ESTIMATION OF COMPLETENESS OF REGRESSION TESTING WITH NORMALIZATION BY WEIGHT COEFFICIENTS

Nizhny Novgorod state technical university n.a. R.E. Alekseev

Purpose: The article is devoted to methods of assessing completeness of regression testing using normalization principles used for the selection of tests on weights.

Design/ methodology/ approach: Introduction of weight as a value of significance of tests of the chosen approach of selection of tests for regression testing in assessing the completeness of regression testing.

Findings: The presented method of assessing the completeness of regression testing execution allows us to quickly assess the degree of completion of the current iteration of regression testing and identify tests, the implementation of which will have a greater impact on achieving the threshold of regression testing execution.

Research limitations/implications: This technique is not suitable for the formation of a list of tests for regression testing, which is a very topical issue for projects in which a significant number of tests have been developed using various methods. In addition, this technique does not take into account the possibility of including the same tests in different groups of regression testing.

Originality/ value: This technique provides flexibility to the order of execution and the number of tests, which will reduce the execution time of the test, while saving other resources.

Key words: regression testing, coefficient of completeness of regression testing, regression testing approaches.