

РАЗРАБОТКА КРИТЕРИЕВ ДЛЯ ОЦЕНКИ КАЧЕСТВА МОБИЛЬНЫХ ПРИЛОЖЕНИЙ

Ёрш А. О., Дейнеко П. В., Байданов А. М.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Дик С. К. – к.ф.м.н., доцент

Целью данной научной работы было разработать надёжный, многомерный критерий для испытания, классификации и оценки качества мобильных приложений. Литературный поиск проводился для выявления статей, содержащих явные критерии оценки качества сайтов или приложений и опубликованных между январём 2000-го и 2019-го годов. Существующие критерии для определения качества приложения были распределены по категориям для разработки новых уменьшенных шкал оценки мобильного приложения. Были выявлены 5 обширных категорий критериев, включая 4 объективных шкалы качества: активность пользователей, функциональность, эстетика и качество информации; и одна субъективная шкала качества.

Учитывая быстрое распространение приложений для смартфонов, пользователям становится всё труднее выявлять и оценивать высококачественные приложения. О качестве приложений доступно немного информации, кроме оценок звёздами, которые опубликованы на веб-страницах продавцов, а обзоры приложений являются субъективными по своему характеру и могут быть получены из подозрительных источников. Выбор приложений на основе популярности даёт мало либо не значимую информацию о качестве приложения.

Большинство опубликованной литературы концентрируется на технических аспектах и представлены в форме инструкций, которые не оценивают качество этих функций. Качество может быть описано как функция [1]:

- 1) содержимого;
- 2) внешнего вида и мультимедиа;
- 3) навигации;
- 4) структуры и дизайна;
- 5) уникальности.

Проведённый синтез критериев оценки включает 165 критериев, сгруппированных по 13 группам (например, дизайн и эстетика, простота использования). Однако, 33 критерия не могли быть сгруппированы и маркированы как “прочее”, подчёркивая сложность задачи.

Надёжный и объективный инструмент необходим для того, чтобы рассчитать степень, в которой мобильные услуги удовлетворяют критериям качества. Шкала должна быть простой для понимания и использования с минимальным обучением.

Для выявления статей, содержащих явные критерии оценки качества, связанные с сайтом или приложением, проводился многогранный литературный поиск. Англоязычные работы с января 2000-го до января 2019-го были получены из PsycINFO, ProQuest, EBSCOhost, IEEE Xplore, Web of Science и ScienceDirect. Условиями поиска были: “mobile” и “app*” или “web*” в папе с “quality” или “criteria” или “assess*” или “evaluat*”. Профессиональные учебные пособия, неопубликованные манускрипты и материалы конференций были также исследованы для дополнительных критериев качества. После изначального контроля заголовка и аннотации, были включены только те исследования, в которых сообщается о критериях оценки качества для приложений или веб-содержимого.

В процессе исследования был создан список исчерпывающий мобильных приложений. Критерии включения приложений:

- 1) русский язык;
- 2) бесплатность;
- 3) доступность в популярных магазинах мобильных приложений;
- 4) из категорий “Здоровье и фитнес”, “Стиль жизни”, “Медицина”, “Производительность”, “Музыка”, “Образование” и “Утилиты”.

Критерии включения категории были основаны на внимательном изучении заголовков и типов приложений, представленных в этих категориях.

60 приложений были случайно выбраны, используя веб-сайт рандомизации. Первые 10 были использованы для обучения и разработки опытного образца. Было 2 эксперта-оценщика, которые тестировали каждое из первых 10-ти приложений в течение как минимум 10 минут и затем независимо оценили их качество, используя критерии оценки мобильного приложения. Оценщики собрались для сопоставления оценок и устранения двусмысленности в содержании шкалы до достижения консенсуса. Критерии были пересмотрены, основываясь на этом опыте, а оставшиеся 50 приложений были опробованы и независимо оценены. Минимальный размер выборки из 41 необходим для определения того, что истинная согласованность заключений различных исследователей лежит в пределах 0,15 выборочного наблюдения 0,80, с гарантией в 87% (основана на 10 тысячах запусков) [2]. Размер выборки из 50, следовательно, обеспечивает существенную

уверенность в оценке согласованности заключений различных исследователей в конкретном исследовании.

Стратегия поиска принесла 25 публикаций, включая рецензируемые статьи журнала ($n = 14$), материалы конференций ($n = 8$) и интернет-ресурсы ($n = 3$), содержащие явные критерии качества, связанные с сетью или мобильными. Всего было извлечено 427 критериев, 56 удалено как дубликаты и 22 сочтены неприменимыми для приложения. Оставшиеся 349 критерия были сгруппированы экспертной группой в 6 категорий: одна, относящиеся к классификации приложения, 4 категории – по объективным качествам приложения (активность пользователей, функциональность, эстетика и качество информации) и одна – по субъективному качеству приложения (таблицу 1), с помощью итеративного подхода.

Таблица 1. Количество критериев для оценки качества приложений, выявленных в литературном поиске.

Категория критерия	Количество	%
Классификация, конфиденциальность, безопасность, регистрация, сообщество, принадлежность приложения	12	3.4
Эстетика, графика, разметка, внешняя привлекательность	52	14.8
Активность пользователей, развлечение, кастомизация, интерактивность, соответствие целевой группе и пр.	66	18.9
Функциональность, производительность, навигация, строгий дизайн, простота использования	90	25.8
Информация, качество, количество, визуальная информация, достоверность, цель, описание	113	32.4
Субъективное качество, стоит рекомендовать, стимулирует повторное использование, общий рейтинг удовлетворённости	19	4.6

Категория классификации содержит описательную информацию о приложении (например, цену, платформу, рейтинг), а также его технические аспекты (например, авторизация, защита паролем, возможность совместного использования). Дополнительные разделы содержат информацию о целевой возрастной группе приложения (если это необходимо), а также информацию о том, на какие аспекты нацелено приложение. Эта сфера может быть адаптирована для включения / исключения конкретных областей содержимого, если это необходимо.

Критерии качества приложения были сгруппированы в рамках активности пользователей, функциональности, эстетики, качества информации и субъективных категорий качества для разработки 23-х подкатегорий. Каждый элемент использовал 5-бальную шкалу:

- 1) ненадлежащий;
- 2) скудный;
- 3) приемлемый;
- 4) хороший;
- 5) превосходный.

Показатели для этих оценочных привязок были написаны для каждого элемента. В случае, когда элемент не мог быть применим для всех приложений, был включен вариант «Не применим». Экспертная группа внимательно изучила элементы, и терминология оценочных показателей была использована по всей шкале для уточнения соответствия и состава языка.

Вычисление средних оценок активности пользователей, функциональности, эстетики, качества информации объективной подшкалы и общая средняя оценка качества приложения – это то, как формируются критерии. Средняя оценка использовалась вместо общей, потому что элемент может быть оценён как «Не применимый». Кроме того, средняя оценка использовалась для предоставления соответствия оценки качества близкому формату оценки звёздами.

Разработанные критерии предоставляют многомерную, надёжную и гибкую шкалу оценки качества приложения для исследователей, разработчиков и пользователей. Текущие результаты показывают, что критерии являются надёжным показателем качества приложения, при условии, что оценщики проходят достаточную и надлежащую подготовку.

Список использованных источников:

1. Moustakis V. Website quality assessment criteria. 2004 Nov Presented at: 9th international conference of information quality / V. Moustakis, C. Litos, A. Dalivigas, L. Tsironis – 2004 Nov 5-7 – Boston, MA, USA p. 59-73.
2. Z ou G.Y. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med – 2012 Dec 20.