Дегтяренко И.А., Леонова А.Б. Экспериментальная разработка комплексного подхода к оценке юзабилити интернет-сайтов

-

English version: Degtyarenko I.A., Leonova A.B. Experimental elaboration of the complex approach to websites usability evaluation

Московский государственный университет им. М.В.Ломоносова, Москва, Россия

Сведения об авторах
Литература
Ссылка для цитирования

В эксперименте обоснованы принципы и методическое обеспечение комплексного подхода к оценке юзабилити интернет-сайтов. При этом учитывались показатели пользовательской удовлетворенности и субъективного комфорта в сочетании с объективными индикаторами успешности выполнения рабочих заданий на сайте, эффективности распределения когнитивной нагрузки (по показателям движений глаз) и степени эмоционально-вегетативного напряжения. Продемонстрирована пригодность разработанного подхода для дифференцированной оценки юзабилити в практике проектирования интернет-сайтов.

Ключевые слова: юзабилити, интернет-сайт, удовлетворенность пользователей, эффективность работы, субъективный комфорт, когнитивная нагрузка, движения глаз, эмоционально-вегетативное напряжение

Обращение к ресурсам Интернета приобретает все большее значение в жизни современных людей, охватывая практически все сферы активности человека: трудовую деятельность, образование, поиск справочной информации, общение, отдых, развлечения и пр. Начиная с конца 1980-х годов в научно-прикладных разработках складывается представление о ключевой роли такой характеристики, как *эргономичность* создаваемых программных продуктов, которая определяет их коммерческую привлекательность, масштабность распространения на рынке компьютерных средств, субъективную удовлетворенность пользователей при работе с ними, и связана с эффективностью их использования при решении различных задач в интернет-среде. В современных моделях оценки качества программных продуктов эргономичность обычно описывается с помощью специального термина «юзабилити» (usability), который дословно переводится с английского как «пригодность к использованию» [Костин, 2011].

Классическое определение этого термина дано в стандарте [ISO 9241-11:1998]: «юзабилити – это степень, в которой продукт может быть использован определенной группой пользователей для достижения определенных целей в определенном контексте использования результативно, производительно и с чувством удовлетворенности»[1]. Такая постановка проблемы заостряет внимание на необходимости разработки комплексных методов и средств для оценки качества программных продуктов по критериям удобства и эффективности работы, базирующихся на использовании поведенческих, психологических и психофизиологических показателей.

Методы оценки юзабилити интернет-сайтов

В практике оценки эргономичности программных продуктов (ПП) традиционно применяются такие методы, как экспертная оценка и юзабилити-тестирование [Nielsen, 1993]. Последнее представляет собой специально организованный эксперимент, в ходе которого представители целевой аудитории пользователей ПП выполняют с его помощью ряд заданий, критичных для успешности реализации их деятельности. Проведение такого тестирования на практике чаще всего сводится к внешнему наблюдению за ходом исполнения заданий с целью выявления основных проблем, с которыми сталкиваются участники эксперимента. В тех случаях, когда при проведении тестирования все же фиксируются некоторые количественные параметры, то они обычно ограничиваются оценкой: (1) правильности выполнения заданий (например, по шкале «выполнил самостоятельно»– «выполнил с подсказкой» – «не выполнил») и (2) времени их исполнения. Пример типичной процедуры проведения коммерческих юзабилити-тестирований приведен в работе В.В.Головача [Головач, 2005].

С точки зрения приведенного выше определения юзабилити первая из названных групп количественных показателей является оценкой результативности, а вторая – производительности деятельности пользователя. Кроме того, для оценки производительности иногда применяются и такие частные характеристики моторного поведения, как общее число щелчков «мышью», количество переходов между экранами и т.п. Однако не стоит забывать, что помимо показателей успешности выполнения заданий принципиально важной характеристикой *эффективности деятельности* является цена внутренних затрат, определяемая по величине прикладываемых субъектом когнитивных, эмоциональных и физиологических усилий [Леонова, 2007]. Тем не менее эти показатели почти никогда не используются в практике юзабилити-тестирования, а оценка такого важного показателя, как удовлетворенность пользователя работой с ПП, обычно проводится в форме свободного опроса, то есть без привлечения стандартизованных процедур.

Отметим, что до настоящего времени каких-либо валидизированных и хорошо структурированных тест-опросников пользовательской удовлетворенности в работах отечественных авторов не предлагалось. Более того, и в зарубежной практике исследователи чаще всего пренебрегают использованием стандартизованных опросных методик, несмотря на то что они существуют [Hornbæk, 2006]. Поэтому особую актуальность приобретает решение проблемы по созданию более емких и упорядоченных методов и процедур для оценки юзабилити ПП (в нашем случае интернет-сайтов), с возможностью их объединения в единый методический комплекс.

Комплексный подход к оценке юзабилити. Гипотеза исследования

В выполненном нами исследовании была разработана и апробирована оригинальная схема эксперимента для проведения комплексной оценки юзабилити интернет-сайтов, которая помимо традиционных показателей результативности работы пользователя (успешность и скорость выполнения заданий в интернет-среде) включала также оценки:
– удовлетворенности пользователя интернет-сайтом с помощью валидизированной ранее методики «Оценка пользовательской удовлетворенности сайтом» (ОПУС, см. [Дегтяренко, Леонова, 2012]);
– «внутренней цены» деятельности по показателям (а) субъективного комфорта текущего состояния, (б) эффективности стратегий распределения когнитивной нагрузки и (в) величины затрачиваемых психофизиологических ресурсов.

Цель исследования состояла в экспериментальном обосновании предложенного варианта комплексного подхода к оценке эффективности работы пользователя с интернет-сайтом, пригодного для использования в практике юзабилити-тестирования. Параллельно с этим решались задачи выделения наиболее информативных диагностических показателей по каждой группе использованных методик с учетом возможностей получения на их основе содержательно валидной

интерпретации данных.

Для достижения поставленной цели была проведена серия экспериментов моделирующего типа, в которых осуществлялась сравнительная оценка эффективности деятельности испытуемых при выполнении идентичных заданий на двух сайтах, характеризующихся разной степенью эргономичности.

Основной *гипотезой*, проверявшейся в эксперименте, было предположение о том, что при работе испытуемых с более эргономичным сайтом у них будут получены более высокие оценки по всему комплексу диагностических методик, включая показатели: (а) пользовательской удовлетворенности, (б) успешности выполнения заданий на сайте и (в) субъективной комфортности текущего состояния и менее затратных способов актуализации когнитивных и психофизиологических ресурсов.

Методы

Общая характеристика экспериментального исследования

В эксперименте моделировалось выполнение испытуемыми ряда типовых пользовательских задач, связанных с подготовкой и проведением опросов в Интернете, на двух сайтах одной тематической категории, различающихся по степени эргономичности интерфейса. Для этого были выбраны два сайта, находящихся в открытом доступе для пользователей Интернета.

Выбор этих сайтов определялся следующими причинами:
– эти сайты достаточно широко используются представителями социономических профессий (психологами, социологами, маркетологами, преподавателями, сотрудниками кадровых служб и т.д.);
– деятельность по проведению онлайн-опросов соответствует содержанию программы профессиональной подготовки студентов психологических вузов, которые являлись участниками нашего исследования;
– по данным предварительной экспертной оценки, проведенной авторами исследования, эти сайты обладали разной степенью эргономичности – сайт 1 в целом характеризуется более простой структурой и оптимальностью графического оформления интерфейса (далее – Э-сайт) по сравнению с информационно «зашумленным» сайтом 2 (далее – Н-сайт);
– с технической точки зрения эти сайты являются достаточно надежными инструментами и позволяют решать сложные задачи на протяжении длительного времени, что позволяет моделировать «экологически валидные» ситуации работы пользователя в интернет-среде.

Все участники эксперимента выполняли идентичный набор заданий на каждом из сайтов. В ходе эксперимента проводился сбор данных по всему комплексу диагностических методик согласно с подготовленным планом и схемой проведения эксперимента (см. ниже).

Испытуемые

В исследовании приняли участие 26 испытуемых – студентов факультетов психологии Московского государственного университета имени М.В.Ломоносова и Московского государственного лингвистического университета (16 женщин и 10 мужчин в возрасте от 19 до 22 лет). Все испытуемые имеют достаточный опыт работы с Интернетом (не менее двух лет) и пользуются им не реже одного раза в неделю, а большинство – ежедневно.

Схема эксперимента

Эксперимент был построен по плану интраиндивидуального дизайна. Каждый испытуемый в течение одного экспериментального сеанса выполнял набор одинаковых заданий сначала на одном (*серия 1*), а затем – на другом сайте (*серия 2*). Порядок работы испытуемых с разными сайтами варьировался: половина испытуемых работала сначала с Э-сайтом, а потом с Н-сайтом, другая половина – наоборот. Перед началом эксперимента с участниками проводилось *вводное занятие*.

До начала и после окончания каждой экспериментальной серии проводились *встроенные диагностические замеры*, включавшие заполнение тест-опросников на самооценку состояния и пользовательской удовлетворенности, а также регистрацию показателей электрокардиограммы (ЭКГ). Кроме этого, по ходу выполнения заданий на каждом сайте проводилась синхронизированная видеозапись моторной активности испытуемого и бесконтактная регистрация движения глаз. Общая схема проведения эксперимента с указанием содержания и времени проведения диагностических замеров в схематичном виде представлена на рис. 1.

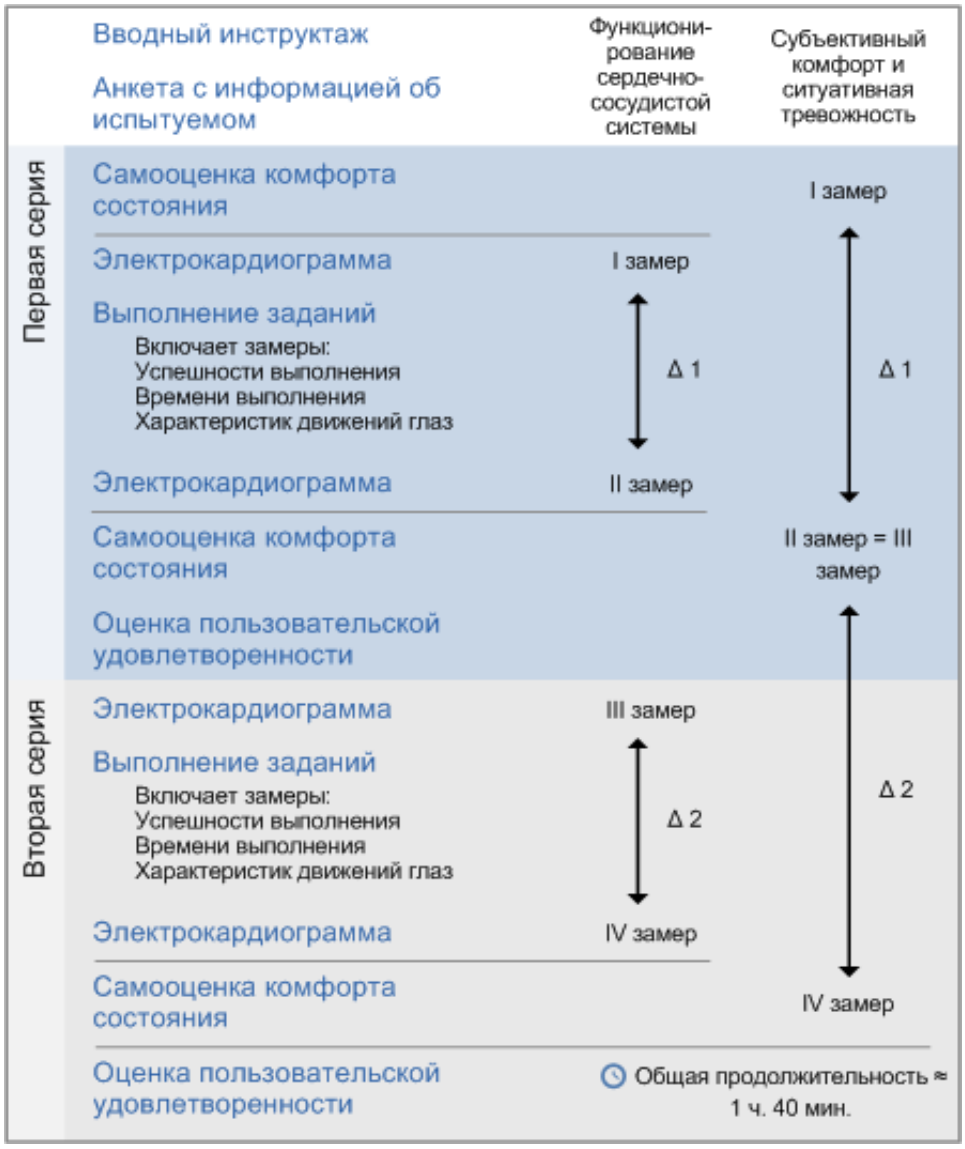

Рис. 1. Общая схема проведения и содержание основных этапов эксперимента.
Примечания. Символы Δ1 и Δ2 обозначают величину изменения соответствующей группы показателей за время соответствующей серии.

Экспериментальный стенд

Техническая реализация эксперимента осуществлялась на базе установки, включавшей два

компьютера (для испытуемого и экспериментатора), устройства бесконтактной регистрации движений глаз и устройства записи и спектрального анализа ЭКГ (см. рис. 2). На компьютерах испытуемого и экспериментатора использовались операционные системы Windows 7 и Windows XP, а также интернет-браузер Google Chrome, программа видеорегистрации Techsmith Morae, секундомер TimeLeft и программа для удаленного управления с компьютера экспериментатора TeamViewer.

Для бесконтактной регистрации движений глаз использовалось высокоскоростное дистанционное устройство EyeLink 1000 компании SR Research (разрешающая способность – 500 Гц без фиксации головы). Для регистрации и спектрального анализа показателей ЭКГ использовался прибор ВНС-Спектр компании Нейрософт. Запись производилась с трех отведений на предплечьях испытуемого в течение трех минут.

Рис. 2. Общий вид экспериментальной установки.
Примечания. 1 – веб-камера для аудио- и видеофиксации действий испытуемого, 2 – компьютер испытуемого, 3 – устройство дистанционной регистрации движений глаз EyeLink 1000, 4 – контроллер устройства регистрации движений глаз, 5 – компьютер экспериментатора, 6 – вегетотестер ВНС-Спектр, применявшийся для записи ЭКГ.

Вводное занятие

Перед началом основой части эксперимента со всеми испытуемыми (группы по 6–12 чел.) проводился вводный инструктаж, в ходе которого демонстрировались и практически осваивались основные приемы создания анкет в сервисе Google Documents. При этом испытуемые пробовали самостоятельно выполнять задания, по содержанию сходные с экспериментальными.

Экспериментальные задания

В *основной части* эксперимента испытуемым предлагалось для выполнения 7 заданий, содержание которых было идентичным для работы на каждом из двух сайтов:

1. Ознакомиться с сайтом, его возможностями и структурой (ознакомительное задание).
2. Создать в системе новый опрос.
3. Ввести приветственный текст, который будет отображаться на экране в начале опроса.
4. Создать вопрос с двумя вариантами ответа, сделать его обязательным для ответа.
5. Создать еще один вопрос, который будет размещен на отдельной странице анкеты.
6. Разместить на второй странице анкеты две фотографии, загрузив их с «рабочего стола» компьютера.
7. Настроить условное предъявление вопросов таким образом, чтобы вторая страница отображалась или не отображалась в зависимости от ответа респондента на первый вопрос.

Успешность и время выполнения заданий фиксировались в протоколе экспериментатора. Успешность оценивалась по 4-балльной шкале:
4 – самостоятельное выполнение задания в соответствии со всеми заданными критериями;
3 – выполнение задания в соответствии с заданными критериями, но при использовании инструкции-подсказки;
2 – выполнение задания соответствовало реализации требуемых критериев, но до истечения времени работы над ним испытуемый не отмечал, что выполнил его;
1 – задание не выполнялось.

Процедура

Основные этапы эксперимента осуществлялись в полностью компьютеризованном режиме. Все инструкции, тексты заданий и материалы опросных методик предъявлялись испытуемому в электронном виде (с помощью приложения LimeSurvey 1.90). Во вводной инструкции кратко сообщалось, что будет происходить во время эксперимента, но специально не обсуждались научные цели исследования. После этого испытуемые заполняли анкеты с индивидуальными биографическими данными.

Перед началом каждой экспериментальной серии экспериментатор со своего компьютера открывал на компьютере испытуемого соответствующий сайт. Участник должен был вслух прочитать текст задания, после чего экспериментатор запускал на компьютере испытуемого секундомер, а участник приступал к выполнению задания. Когда испытуемый считал задание выполненным, то вслух сообщал об этом. Экспериментатор оценивал качество выполнения задания по заранее определенному набору критериев. Если все критерии были выполнены, то работа над заданием прекращалась и фиксировалась продолжительность его исполнения. Если же соответствие какому-либо критерию не было достигнуто, то экспериментатор сообщал причину, по которой задание не могло считаться выполненным успешно.

Время, отведенное на выполнение заданий, было ограничено. Для каждого из них было выделено 2 временных интервала, длительность которых варьировалась в зависимости от экспертной оценки его сложности, выполненной на основании предварительно проведенных пилотажных серий. После истечения первого интервала испытуемый получал право воспользоваться инструкцией-подсказкой. После истечения второго интервала задание считалось невыполненным. Экспериментатор сам заканчивал выполнение задания, кратко поясняя свои действия испытуемому.

Диагностические методики опросного типа

Для характеристики субъективного отношения испытуемого к работе на конкретном сайте и самооценке текущего состояния с помощью специализированных тест-опросников определялись:

1) степень *удовлетворенности пользователя интернет-сайтом* – с помощью валидизированной авторской методики ОПУС [Дегтяренко, Леонова, 2012], дополненной шкалой «Общая удовлетворенность»;

2) *простота выполнения заданий на сайте* – по однополярной шкале от 1 до 5 баллов, заполняемой после выполнения каждого отдельного задания (за исключением первого – ознакомительного – задания, после которого по той же шкале оценивалось «первое впечатление от сайта»);
3) *субъективная оценка текущего состояния* – с помощью известных стандартизованных методик:
– опросник «Шкала состояний» [Леонова, Капица, 2003];
– сокращенная версия шкалы Спилбергера «Ситуативная тревожность» [Леонова, Наумова, 2009].

Диагностические показатели

На первом этапе обработки данных, полученных в эксперименте, были рассчитаны значения информативных показателей по каждой из включенных в диагностический комплекс методик.

А. Для оценки *успешности выполнения заданий* на каждом сайте отдельно подсчитывалось общее количество заданий, выполненных с тем или иным уровнем качества, то есть с оценкой успешности, соответственно, не менее 4, 3 и 2 баллов.

Использовались также скоростные характеристики работы на том или ином сайте – суммарное время выполнения всех заданий (в мин/сек) и время, затрачиваемое на ознакомительное задание.

Б. Степень *удовлетворенности пользователя работой с сайтом* определялась по показателям основных шкал методики ОПУС: (1) «эффективность», (2) «простота, (3) «полезность», (4) «эмоциональная привлекательность». Первичные оценки, полученные по этим показателям, были переведены в сопоставимые Z-значения, нормализованные в соответствии с данными общей выборки стандартизации (подробнее см. [Дегтяренко, Леонова, 2012]). Аналогичным образом была рассчитана обобщающая оценка по пунктам дополнительной шкалы «Общая удовлетворенность».

Также подсчитаны усредненные оценки сайтов по шкале «Простота выполнения заданий» по всем заданиям.

В. *Субъективные оценки текущего состояния* оценивались по величине сдвигов в значениях общепринятых показателей использованных тест-опросников от начала к концу работы с тем или иным сайтом:– опросник «Шкала состояний» – по разнице значений «Индекса субъективного комфорта» (ИСК);
– шкала «Ситуативная тревожность» – по разнице сводных показателей «Ситуативная тревожность» (СТ).

Г. В качестве показателей *эффективности распределения когнитивных нагрузок* использовался ряд усредненных параметров глазодвигательной активности, наиболее часто рассматриваемых в специальной литературе [Величковский и др., 2010; Ahlstrom, Friedman-Berg, 2006; Grootjen et al., 2007]: (1) длительности фиксаций, (2) частоты фиксаций, (3) диаметра зрачка, (4) частоты морганий и (5) пиковой скорости саккад за все время работы с каждым из сайтов.

Д. Оценка *уровня психофизиологических затрат и наличия признаков дисбаланса* в деятельности сердечно-сосудистой системы проводилась с опорой на стандартный набор показателей, получаемый при спектральном анализе ЭКГ [Дмитриева, Глазачев, 2000; Heart Rate Variability, 1996]. Для интегрированной оценки этих первичных данных была использована разработанная нами ранее 4-факторная модель, которая позволяет рассчитывать значения индексов по четырем обобщенным показателям [Леонова, 2011]: (1) вегетативной напряженности, (2) симпатической мобилизации, (3) парасимпатического дисбаланса и (4) вариабельности ЧСС. С содержательной точки зрения они характеризуют доминирующий тип эмоционально-вегетативного реагирования испытуемых на ситуацию.

Методы анализа данных

На основе расчета значений всех перечисленных диагностических показателей была составлена матрица данных для дальнейшей статистической обработки результатов, реализованной на базе стандартного пакета программ SPSS 19.0. Для выявления значимых различий между оценками показателей, полученными при работе испытуемых с разными сайтами, был использован парный Т-критерий Вилкоксона для связанных выборок, дополненный в ряде случаев процедурами корреляционного (по Спирмену) и кластерного анализа.

Анализ результатов исследования

В сводном виде результаты сравнения оценок работы на двух сайтах по всему комплексу использованных показателей представлены в таблице 1. С помощью критерия Т-Вилкоксона было выявлено большое число статистически достоверных различий между ними практически по всем группам диагностических методик.

Таблица 1

Средние значения диагностических показателей при выполнении экспериментальных заданий на разных сайтах и достоверность различий между ними (по критерию Т-Вилкоксона)

Показатели	Средние оценки по сайтам		T
	Э-сайт	Н-сайт	
Методика «Оценка пользовательской удовлетворенности сайтом» (ОПУС)			
Эффективность	0,24	–0,63	4,08***
Простота	–0,33	–1,75	7,42***
Полезность	0,07	–0,64	5,71***
Эмоциональная привлекательность	–0,11	–0,73	3,58***
Дополнительные показатели субъективной удовлетворенности			
Общая удовлетворенность	–0,13	–0,95	4,53***
Простота выполнения заданий	3,90	2,90	6,08***
Первое впечатление от сайта	3,58	3,50	0,35
Успешность и время выполнения заданий			
Показатель успешности = 4	4,08	1,92	7,25***
Показатель успешности ≥ 3	5,46	3,73	5,55***
Показатель успешности ≥ 2	5,77	4,15	6,34***
Общее время выполнения	18,26	28,29	–8,87***
Время на ознакомительное задание	1,35	1,45	–1,31°
Сдвиги переживания субъективной комфортности текущего состояния			
Индекс субъективного комфорта (ИСК)	+0,54	–2,38	2,72**
Индекс ситуативной тревожности (СТ)	–0,50	+0,38	–0,90
Характеристики глазодвигательной активности			
Длительность фиксации (мс)	262	254	2,37*
Кол-во фиксаций (в мин.)	180,5	175,2	1,02
Кол-во морганий (в мин.)	13,3	12,1	2,44*
Диаметр зрачка	276	261	1,73°
Пиковая скорость саккад	255	248	0,72
Сдвиги в структуре показателей вегетативного напряжения			
Вегетативная напряженность	–0,64	–1,29	1,28

Симпатическая мобилизация	–0,02	–0,121	0,37
Парасимпатический дисбаланс	–0,11	+0,019	–0,48
Вариабельность сердечного ритма	+0,02	+0,235	–0,95
ЧСС	–5,96	–8,62	1,20

Примечания. *** $p \le 0,001$; ** $p \le 0,01$; * $p \le 0,05$; ° $p \le 0,1$

Ниже представлен детализированный анализ обнаруженных различий. Напомним, что более эргономичный сайт обозначается как *Э-сайт*, а менее эргономичный – как *Н-сайт*.

Удовлетворенность пользователя работой с разными сайтами

По абсолютному большинству показателей пользовательской удовлетворенности более эргономичный *Э-сайт* получил достоверно более высокие оценки, чем менее эргономичный *Н-сайт* с максимально высоким уровнем значимости ($p < 0,001$). Различия зафиксированы по всем основным шкалам методики ОПУС. С этим хорошо согласуются данные, полученные по дополнительным шкалам «Общая удовлетворенность» и «Простота выполнения заданий», оценки по которым также выше для *Э-сайта* ($p < 0,001$). Эти данные свидетельствуют о пригодности использованных психодиагностических инструментов для достаточно четкой и дифференцированной оценки различных аспектов юзабилити интернет-сайтов. Исключением является только шкала «Первое впечатление от сайта» (оценки по которой даются сразу после ознакомительного задания), что говорит о постепенном формировании субъективных оценок пользовательской удовлетворенности в процессе накопления опыта работы с конкретным сайтом.

Успешность и временные характеристики выполнения заданий

Значения по большинству показателей успешности и временных затрат при выполнении заданий на каждом из сравниваемых сайтов также существенно различаются ($p < 0,001$). При этом показатели количества и качества решения задач, а также скорости их исполнения были существенно выше при работе с *Э-сайтом*. Менее выражены различия только по показателю времени, затрачиваемого на ознакомительное задание. Однако и в этом случае имела место тенденция к его сокращению при работе с *Э-сайтом* ($p < 0,1$) – по-видимому, как следствие более компактного дизайна и меньшей информационной насыщенности его интерфейса по сравнению с *Н-сайтом*.

Динамика субъективных оценок текущего состояния

При сравнении сдвигов в самооценке текущего состояния на протяжении всего периода работы на каждом из сайтов обнаружены достоверные различия по показателю «Индекс субъективного комфорта» (ИСК, $p < 0,01$). Они отражают не только изменения в общем уровне ИСК, но и разнонаправленность в динамике переживаний комфортности текущего состояния при работе на разных сайтах. Так, оценки ИСК после завершения работы на *Э-сайте* возрастают по сравнению с фоном и находятся в диапазоне оптимальных значений. Напротив, после работы с *Н-сайтом* они резко ухудшаются и смещаются в зону выраженного неблагополучия. Достоверных различий в динамике показателя «Индекс ситуативной тревожности» (СТ) при выполнении работы на разных сайтах выявлено не было, хотя по абсолютным оценкам уровень ситуативной тревожности после завершения работы на *Н-сайте* нарастает и смещается в диапазон повышенных значений.

Показатели глазодвигательной активности

Были обнаружены достоверные различия по ряду показателей глазодвигательной активности, регистрируемых в процессе работы с разными сайтами. При работе с *Э-сайтом* в среднем увеличивалась длительность фиксаций ($p < 0,05$), частота морганий в минуту ($p < 0,05$) и наблюдалась тенденция к увеличению среднего диаметра зрачка ($p < 0,1$).

Согласно литературным данным, нарастание этих показателей может быть связано с повышением уровня когнитивной нагрузки, но это совсем не обязательно говорит об использовании менее эффективных стратегий выполнения умственных действий [Величковский и др., 2010; Ahlstrom, Friedman-Berg, 2006; Grootjen et al., 2007]. В частности, более длительные фиксации (около 300 мс и более) обычно связаны с увеличением диаметра зрачка, а также предваряются и завершаются сравнительно небольшими по амплитуде саккадами. Согласно мнению Б.М.Величковского [2006], такой паттерн реализации длительных фиксаций отражает актуализацию процессов фокальной обработки информации, направленных на прицельную идентификацию предметов и событий. Напротив, преобладание менее продолжительных фиксаций говорит о доминировании процессов пространственного распределения внимания в поле восприятия и меньшей упорядоченности стратегий зрительного поиска.

В соответствии с этой трактовкой можно содержательно проинтерпретировать обнаруженные различия в паттернах глазодвигательных показателей при работе с разными сайтами. Для работы с *Э-сайтом* характерно доминирование процессов фокальной обработки информации, направленных на упорядоченное выделение в зрительном поле зон для последующей когнитивной обработки выделенного фрагмента интерфейса. В случае выполнения заданий на *Н-сайте* использование таких упорядоченных стратегий затруднено вследствие необходимости частого переключения внимания на поиск релевантной информации в зашумленном зрительном поле.

Отдельно следует обсудить выявленные различия по показателю частоты морганий, который имеет более высокие оценки при работе с *Э-сайтом* ($p < 0,05$). Отметим, что по поводу интерпретации этого показателя в литературе приводятся противоречивые мнения. Так, в исследованиях Дж.Б.Брукингса и др. (цит. по [Di Nocera et al., 2007]) указывается на наличие отрицательной корреляционной связи этого показателя с уровнем когнитивной нагрузки. Напротив, в работах К.Такахаси и соавт. [Takahashi et al., 2000] отмечается, что частота морганий возрастает при повышении нагрузки. Увеличение частоты морганий интерпретируется также и как следствие развития утомления, и как индикатор усиления целенаправленности при реализации сенсорных действий. В последнем случае акт моргания фиксирует момент завершения исполнения одного целостного фрагмента действия и перехода к следующему [Kahneman, 1973].

Относительная непродолжительность работы испытуемых в нашем эксперименте вряд ли позволяет говорить о развитии выраженного утомления. Поэтому можно предположить, что увеличение числа морганий при работе с *Э-сайтом* говорит о наличии более четко структурированных переходов от выполнения одного фрагмента умственных действий к следующему, что указывает на большую эффективность используемых стратегий распределения когнитивных нагрузок в этом случае по сравнению с *Н-сайтом*.

Эмоционально-вегетативное реагирование при выполнении заданий

Усреднение данных по всей выборке испытуемых не позволило выявить достоверных различий между сайтами ни по одному из обобщенных показателей вегетативного напряжения (см. таблицу 1). В то же время причиной этого могла быть высокая индивидуальная вариативность в работе психофизиологических механизмов, на что часто указывается в специальной литературе [Дмитриева, Глазачев, 2000; Леонова, 2011; Thayer, Lane, 2000]. Для учета влияния этого фактора был проведен кластерный анализ внутри этой категории показателей по всей выборке испытуемых (по методу Quick Cluster Analysis). В результате были выделены две подгруппы испытуемых, образовавшие кластеры 1 и 2, которые различались по типу эмоционально-вегетативного

реагирования на ситуацию.

Психофизиологический статус представителей кластера 1 (18 чел.) характеризовался повышенным уровнем симпатической мобилизации без признаков парасимпатического дисбаланса. Для представителей кластера 2 (8 чел.) был типичен несколько сниженный уровень общей вегетативной напряженности с выраженными признаками парасимпатического дисбаланса. Наличие значимых различий между обобщенными кардиологическими показателями при выполнении заданий на разных сайтах проводилось отдельно по каждому кластеру (сравнение по Т-критерию Вилкоксона). В результате были выявлены характерные паттерны изменений в типах эмоционально-вегетативного реагирования на ситуацию, специфичные для разных кластеров.

Кластер 1 (см. рис. 3). У представителей этой подгруппы при работе с *Э-сайтом* наблюдалось нарастание показателя симпатической мобилизации ($p < 0,05$), свидетельствующее о более выраженном влиянии центральных механизмов активизации энергетических ресурсов, и снижение показателя вариабельности сердечного ритма ($p < 0,01$), что говорит о стабильном функционировании механизмов вегетативной регуляции деятельности. В отличие от этого при работе с *Н-сайтом* было отмечена тенденция к снижению уровня симпатической мобилизации ($p < 0,1$) с одновременным нарастанием вариабельности ЧСС ($p < 0,05$). Иными словами, у представителей этой подгруппы при работе с *Э-сайтом* активизация психофизиологических ресурсов происходила по оптимальному типу, близкому к состоянию «операциональной напряженности»[2], тогда как работа с *Н-сайтом* вызывала признаки дисбаланса в механизмах мобилизации энергетических ресурсов, характерных для состояний повышенной эмоциональной напряженности «импульсивного типа».

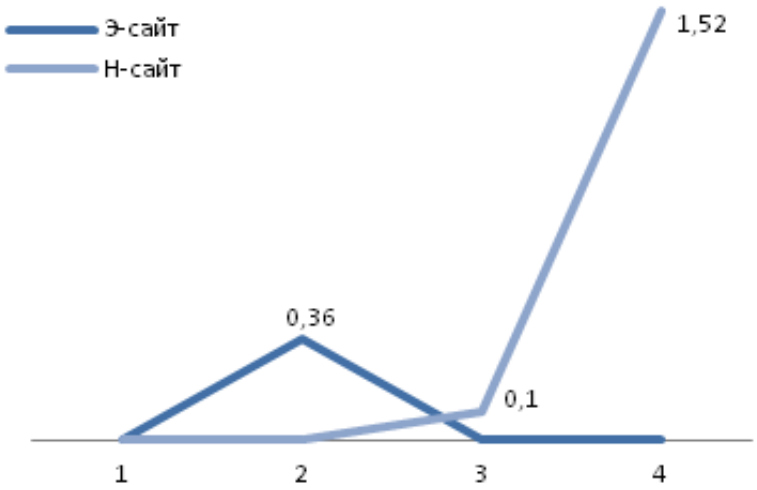

Рис. 3. Кластер 1 (N = 18). Сдвиги показателей эмоционально-вегетативного реагирования за время выполнения заданий на каждом из сайтов.
Примечания. 1 – Вегетативная напряженность, 2 – Симпатическая мобилизация, 3 – Парасимпатический дисбаланс, 4 – Вариабельность ЧСС.

Кластер 2 (см. рис. 4). У представителей этой подгруппы при выполнении заданий на *Э-сайте* наблюдалась тенденция к снижению признаков парасимпатического дисбаланса и уменьшению вариабельности ЧСС ($p < 0,1$), что соответствует развитию оптимального состояния «операциональной напряженности». В то же время при работе с *Н-сайтом* усиливались признаки парасимпатического дисбаланса, приводящие к заторможенности и скованности реакций, а также резкое нарастание вариабельности ЧСС ($p < 0,05$). Такой вариант неэффективного вегетативного реагирования на ситуацию относится к классу состояния повышенной эмоциональной напряженности «тормозного типа».

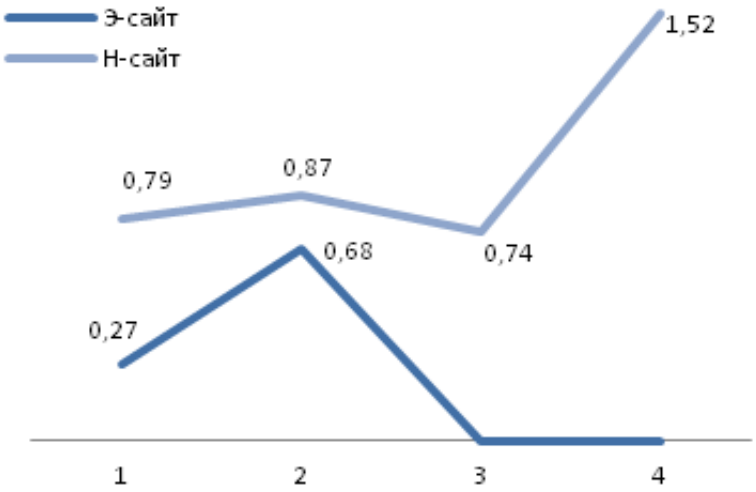

Рис. 4. Кластер 2 (N = 8). Сдвиги показателей эмоционально-вегетативного реагирования за время выполнения заданий на каждом из сайтов.
Примечания. 1 – Вегетативная напряженность, 2 – Симпатическая мобилизация, 3 – Парасимпатический дисбаланс, 4 – Вариабельность ЧСС.

Сопоставление данных об индивидуализированных типах эмоционально-вегетативного реагирования указывает на то, что при работе с *более эргономичным сайтом* у представителей обеих подгрупп преобладают адекватные способы активизации психофизиологических ресурсов, соответствующие оптимальному рабочему состоянию «операциональной напряженности». Выполнение заданий на *менее эргономичном сайте* приводило к деструктивным изменениям в протекании регуляторных процессов по типу нарастания эмоционально-вегетативного напряжения. Несмотря на наличие индивидуальных различий в специфике типа реагирования на затруднения в ситуации («импульсивного» или «тормозного»), они указывают на увеличение затрат внутренних ресурсов, что является одним из важных проявлений снижения эффективности деятельности [Леонова, 2007].

Выводы и заключение

Результаты экспериментального исследования указывают на систематический характер различий, обнаруживаемых по всему комплексу показателей при выполнении серии идентичных заданий на интернет-сайтах с разной степенью эргономичности. Согласно логике разрабатываемого нами комплексного подхода, эти различия соответствуют следующим критериям, необходимым для дифференцированной оценки юзабилити соответствующего класса программных продуктов:

1. Удовлетворенность пользователя работой с сайтом (*мотивационная составляющая*), диагностируемая по основным шкалам методики «Оценка пользовательской удовлетворенности сайтом» и дополненная оценками по шкалам «Общая удовлетворенность» и «Простота работы с сайтом». По всем этим показателям оценки более эргономичного сайта были значительно выше менее эргономичного сайта.

2. Продуктивность и скоростные характеристики работы с сайтом (*результативная составляющая*), рассчитываемые по объективным показателям успешности работы (количество и качество выполненных заданий) и временным затратам на исполнение. Успешность работы на более эргономичном сайте в среднем была почти на 40% выше (причем по числу максимально

качественно выполненных заданий – на 53%), чем на менее эргономичном сайте, а временные затраты сократись более чем на 35%.

3. Внутренняя «цена деятельности» (*ресурсная составляющая*), определяемая с точки зрения:
– субъективных оценок текущего состояния. При работе с более эргономичным сайтом в течение всего периода выполнения заданий сохранялся стабильно благополучный уровень переживаемой комфортности состояния;
– объективных показателей эффективности распределения когнитивных нагрузок. По данным бесконтактной регистрации движений глаз, выполнение заданий на более эргономичном сайте осуществлялось с помощью целенаправленных и упорядоченных стратегий обработки информации – в отличие от преобладания стратегий свободного зрительного поиска при работе с менее эргономичным сайтом;
– психофизиологических показателей, определяющих тип эмоционально-вегетативного реагирования на ситуацию и обобщающих данные спектрального анализа ЭКГ. При выполнении заданий на более эргономичном сайте происходит адекватная мобилизация физиологических ресурсов и формирование состояния оптимального рабочего напряжения. В случае работы с менее эргономичным сайтом развиваются деструктивные состояния эмоциональной напряженности импульсивного или тормозного типа.

Полученные данные подтверждают конструктивность предложенного комплексного подхода к оценке юзабилити интернет-сайтов и перспективность его дальнейшего развития. Его реализация позволяет выйти за границы используемых в практике юзабилити-инженерии узкоспециализированных приемов и процедур, направленных на оценку только некоторых показателей успешности работы и/или общей удовлетворенности пользователя сайтом. В разрабатываемом подходе заложена возможность интеграции разноуровневых показателей, позволяющих одновременно оценивать как продуктивность решения компьютеризованных задач, так и степень адекватности привлекаемых для этого внутренних ресурсов, что является одним из главных критериев оценки *эффективности* деятельности.

Сочетание субъективных и объективных методов сбора и анализа данных позволяет операционализировать представления о структуре умственных и исполнительных действий, что позволяет выявить «зоны дефицита» при работе с разными интернет-сайтами. Именно это, по нашему мнению, и является главным объектом оценки юзабилити программных продуктов. Быстрое развитие современных технологий и появление компактной аппаратуры открывают новые возможности для создания более совершенных средств, которые позволят обеспечить одновременное применение методов регистрации объективных и субъективных показателей с выходом на содержательный анализ многомерных данных. В этой связи полученные в нашем исследовании результаты будут полезны для проведения углубленных научных разработок и их реализации в практико-ориентированных проектах по оценке юзабилити интернет-сайтов.

Финансирование
Исследование выполнено при поддержке Российского фонда фундаментальных исследований, проект 11-06-00463-а.

Литература

Величковский Б.Б., Злоказова Т.А., Капица М.С. Эффективность обработки прерываний в условиях свободных и вынужденных переключений // Экспериментальная психология. 2010. No. 2. С. 45–47.

Величковский Б.М. Когнитивная наука: основы психологии познания: в 2 т. М.: Смысл: Академия, 2006.

Головач В.В. Юзабилити-тестирование по дешевке // Юзетикс: дизайн интерфейсов и юзабилити-тестирование. М., 2005. http://usethics.ru/blog/lib/testing_by_the_cheap/

Дмитриева Н.В., Глазачев О.С. Индивидуальное здоровье и полипараметрическая диагностика функционального состояния человека. М.: Горизонт, 2000.

Дегтяренко И.А., Леонова А.Б. Оценка удовлетворенности пользователя работой с интернет-сайтом // Национальный психологический журнал. 2012. No. 1. С. 95–103.

Костин А.Н. Круглый стол «Юзабилити как новое направление исследований в инженерной психологии» // Психологический журнал. 2011. Т. 32(6). С. 113–124.

Леонова А.Б. Психическая надежность профессионала и современные технологии управления стрессом // Вестник Московского университета. Сер. 14, Психология. 2007. No. 3. С. 69–81.

Леонова А.Б. Стресс-резистентность и структурная организация процессов саморегуляции деятельности // Психология саморегуляции в XXI веке / ред. В.И.Моросанова. М.: Нестор-История, 2011. С. 354–375.

Леонова А.Б., Капица М.С. Методы субъективной оценки функциональных состояний человека // Практикум по инженерной психологии и эргономике / ред. Ю.К.Стрелков. М.: Академия, 2003. С. 134–166.

Леонова А.Б., Наумова Н.Н. Комплексная психофизиологическая оценка индивидуальной устойчивости к стрессу: метод. рекомендации / НИИ нормальной физиологии им. П.К.Анохина. М., 2009. (Труды Научного совета РАМН по экспериментальной и прикладной физиологии).

Наенко Н.И. Психическая напряженность. М.: Изд-во Моск. ун-та, 1976.

Ahlstrom U., Friedman-Berg F.J. Using eye movement activity as a correlate of cognitive workload // International Journal of Industrial Ergonomics. 2006. Vol. 36(7). P. 623–636.

Di Nocera F., Camilli M., Terenzi M.A. Random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload // Journal of Cognitive Engineering and Decision Making. 2007. Vol. 1(3). P. 271–285.

Grootjen M., Neerincx M.A., van Weert J.C.M., Truong K.P. Measuring cognitive task load on a naval ship: Implications of a real world environment // D.D.Schmorrow, L.M.Reeves (Eds.). Augmented Cognition. Pittsburgh, PA: HCII, 2007. P. 147–156.

Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology // Circulation. 1996. Vol. 93(5). P. 1043–1065.

Hornbæk K. Current practice in measuring usability: Challenges to usability studies and research // International Journal of Man-Machine Studies. 2006. Vol. 64(2). P. 79–102.

ISO 9241-11:1998, Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.

Kahneman D. Attention and effort. Ann Arbor, MI: Prentice-Hall, 1973.

Nielsen J. Usability engineering. San Francisco: Morgan Kaufmann, 1993.

Takahashi K., Nakayama M., Shimizu Y. The response of eye-movement and pupil size to audio instruction while viewing a moving target // Proceedings of the 2000 symposium on Eye tracking research and applications (ETRA '00). New York, NY: ACM, 2000. P. 131–138.

Thayer J.F., Lane R.D. A model of neurovisceral integration in emotion regulation and dysregulation // Journal of Affect Disorders. 2000. Vol. 61(3). P. 201–216.

Примечания

[1] Следует отметить, что термин «юзабилити» во многом является синонимичным понятию «эргономичность» с той лишь разницей, что акцент в его применении ставится на оценке умственных усилий и удобства использования определенного инструмента (в нашем случае – программного продукта), а не только внешнего вида и эстетического дизайна оборудования.

[2] Здесь и далее при описании разных типов состояний психической напряженности используется терминология, предложенная в классической работе Н.И.Наенко [Наенко, 1976].

Поступила в редакцию 12 ноября 2011 г. Дата публикации: 25 апреля 2012 г.

Сведения об авторах

Дегтяренко Иван Александрович. Аспирант кафедры психологии труда и инженерной психологии, факультет психологии, Московский государственный университет им. М.В.Ломоносова, ул. Моховая, д. 11, стр. 9, 125009 Москва, Россия.
E-mail: ia.degtyarenko@gmail.com

Леонова Анна Борисовна. Доктор психологических наук, профессор, заведующая лабораторией психологии труда, факультет психологии, Московский государственный университет им. М.В.Ломоносова.
E-mail: ableonova@gmail.com

Ссылка для цитирования

Стиль psystudy.ru
Дегтяренко И.А., Леонова А.Б. Экспериментальная разработка комплексного подхода к оценке юзабилити интернет-сайтов. Психологические исследования, 2012, No. 2(22), 6. http://psystudy.ru. 0421200116/0018.

ГОСТ 2008
Дегтяренко И.А., Леонова А.Б. Экспериментальная разработка комплексного подхода к оценке юзабилити интернет-сайтов // Психологические исследования. 2012. № 2(22). С. 6. URL: http://psystudy.ru (дата обращения: чч.мм.гггг). 0421200116/0018.

[Последние цифры – номер госрегистрации статьи в Реестре электронных научных изданий ФГУП НТЦ "Информрегистр". Описание соответствует ГОСТ Р 7.0.5-2008 "Библиографическая ссылка". Дата обращения в формате "число-месяц-год = чч.мм.гггг" – дата, когда читатель обращался к документу и он был доступен.]

К началу страницы >>