

Размещение в Интернете мирового культурного наследия методом краудсорсинга*

Автор приводит определение термина «краудсорсинг». Описаны современные мировые тенденции применения этого метода в различных областях знаний. Представлен анализ эффективности его использования в библиотеках, в частности, применительно к процессу оцифровывания культурного наследия и его размещения в Интернете. Показаны преимущества краудсорсинга перед аутсорсингом для библиотек.

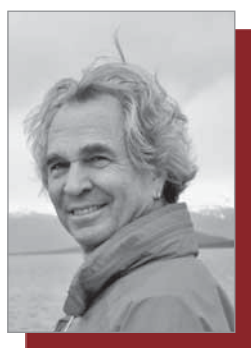
Ключевые слова: краудсорсинг, аутсорсинг, культурное наследие, Интернет, библиотеки, оцифровывание, цифровая коллекция газет, мотивация, оценка проекта.

Понятие «краудсорсинг»

В последнее время часто можно встретить слово «краудсорсинг» (англ. *crowdsourcing*, *crowd* — «толпа», *sourcing* — «использование ресурсов»). Значение слова «толпа» здесь отличается от общепринятого, это любая группа, состоящая из условно-анонимных или не знакомых друг с другом участников.

Термин впервые ввел писатель Дж. Хау (Jeff Howe) в статье «Взлет краудсорсинга» (*The Rise of Crowdsourcing*) [10], в которой он описывает пять принципов, составляющих основу «нового объединения труда»:

1. Толпа рассеяна в пространстве.
2. Ее внимание сосредоточено на конкретном небольшом участке работы.
3. В толпе собирается наибольшее количество специалистов.
4. Она производит в основном некачественную продукцию.
5. Толпа находит самый лучший персонал.



**Фредерик Зарндт
(Frederick Zarndt),**
председатель Секции газет
Международной федерации
библиотечных ассоциаций
и учреждений
(г. Коронадо, США)

* Перевод с английского И.В. Чадновой, ведущего научного сотрудника НИО библиотековедения РГБ.

Как мы увидим далее, некоторые из этих принципов (1 и 5) можно применить к краудсорсингу в процессе оцифровывания культурного наследия, другие (2—4) решительно не подходят к нему.

Интересно, что Хау не упоминает книгу Дж. Суrowицкого (James Surowiecki) «Мудрость толпы» (The Wisdom of Crowds) [19]. В ней описывается несколько экспериментов, в ходе которых было доказано, что процесс выполнения совокупных задач группой (толпой) «нормальных» (normal) людей более эффективен, чем если бы он выполнялся одним специалистом. И хотя Суrowицкий не первым описал этот феномен, он проявляется во многих языках и культурах в качестве пословиц и поговорок (например, «две головы лучше одной»).

Что же означает понятие «краудсорсинг»? Википедия определяет его, как «передачу определенных производственных функций неопределенному кругу лиц. Решение общественно значимых задач силами множества добровольцев, часто координирующих при этом свою деятельность с помощью информационных технологий» [1].

Расширительное значение этого термина можно найти в обзоре литературы по краудсорсингу, сделанном Э. Эстелес Аролас (Enrique Estellés Arolas) и Ф. Гонзалес Ладрон де Гевара (Fernando González Ladrón de Guevara) [5]. Согласно их определению, краудсорсинг является участием в онлайн-деятельности, в ходе которой одно лицо, учреждение, некоммерческая организация или компания предлагают конкретную работу группе лиц, разнородных по числу и опыту, в виде гибкого открытого призыва о решении какой-либо задачи силами волонтеров. Решение задачи различной сложности и модульности, в которой группа должна участвовать посредством своей работы, денег и/или опыта, всегда является взаимовыгодным. Пользователь удовлетворит данный вид потребности, будь это экономическое или социальное признание, самоуважение или развитие собственных навыков, так как лицо, участвующее в краудсорсинге, доведет дело до конца. При этом преимущество, получаемое им, будет зависеть от вида решаемой задачи.

В ходе интернет-запроса, сделанного 25 января 2010 г., Википедия предоставила информацию о 34 краудсорсинговых проектах, в июле 2012 г. список вырос до 122, из них пять проектов связаны с оцифровкой книг, журналов, рукописей и архивов:

- проект Национальной библиотеки Австралии «Австралийские исторические газеты» (Australian Historic Newspapers; <http://trove.nla.gov.au/>);

- указатель проекта «Поиск семьи» (Family Search Indexing; <https://familysearch.org/volunteer/indexing>);

- проект «Расшифровывающая Бентама» (Transcribe Bentham; <http://www.transcribe-bentham.da.ulcc.ac.uk/td/>);

- проект «Фотографии семьи Liljenquist времен Гражданской войны в Америке, предоставленные в дар Библиотеке Конгресса США и выложенные на сайте Flickr методом краудсорсинга» (Liljenquist Family American Civil War Photographs Donated to Library of Congress and Crowdsourced on Flickr; http://www.flickr.com/photos/library_of_congress/sets/72157625520211184/);

- сайт «Распределенные корректоры» в проекте «Гутенберг» (Distributed Proofreaders; <http://www.pgdp.net/>);

- проект «Цифровая коллекция газет Калифорнии» использует краудсорсинг в исправлении текста газет с помощью программы OCR (California Digital Newspaper Collection; <http://cdnc.ucr.edu/>).

Виды краудсорсинга

Краудсорсинговые проекты значительно отличаются друг от друга. Википедия распределяет их по семи различным категориям:

- краудголосование (crowdvoting);
- мудрость толпы (wisdom of the crowd);
- краудфинансирование (crowdfunding);
- краудзакупка (crowdpurchasing);
- очень маленький участок работы (micro-work);
- соревнования (competitions);
- скрытый краудсорсинг (implicit crowdsourcing) [3].

В своей диссертации Д. Брэхэм (Daren Brabham) классифицирует краудсорсинговые проекты следующим образом:

- процесс обнаружения знаний и их управление (knowledge discovery and management);
- распространенный поиск (broadcast search);
- создание контролируемой продукции (peer-vetted creative production);
- распределение задач человеческим разумом (distributed human intelligence tasking) [2].

Э. Эстелес Аролас и Ф. Гонзалес Ладрон де Гевара предлагают другую типологию:

- краудраспределение (crowdcasting);
- краудсотрудничество (crowdcollaboration);
- краудподдержка (crowdsupport);
- краудконтент, краудпродукция, краудпоиск, крауданализ (crowdcontent, crowdproduction, crowdsearching, crowdanalyzing);
- краудфинансирование (crowdfunding);
- краудмнение (crowdopinion) [4].

Типичные краудсорсинговые проекты по оцифровке культурного наследия (корректирование текста с помощью программы OCR, транскрипция, тэггирование) в зависимости от ис-

пользуемой типологии подпадают под категории: «работа на очень маленьком участке» (microwork) [3]; «распределение задач человеческим разумом» (distributed human intelligence tasking) [2]; «краудпродукция» (crowdproduction) [4].

Краудсорсинговые проекты

Рассмотрим три краудсорсинговых проекта по оцифровке культурного наследия: два из них являются библиотечными проектами по оцифровке газет и один, развиваемый генеалогической организацией, — по созданию архивных записей.

Проект «Поиск семьи» (Family Search; <http://www.familysearch.org>). Прилагательные «огромный» и «продуктивный» наилучшим образом подходят для характеристики этого проекта, включающего в себя каталог архивных записей о днях рождений, свадьбах, датах кончин, данные переписи населения, военные сведения, информацию о недвижимости и пр., собранные со всех континентов, исключая Антарктиду. Со времени запуска проекта в сентябре 2005 г. волонтеры заархивировали более 1 млрд 500 млн 88 тыс. 700 записей и сделали экспертизу более 832 млн 207 тыс. 800 записей. Сами записи отличаются легкостью поиска и доступа в Интернете.

К июлю 2012 г. проект имел более 780 тыс. зарегистрированных пользователей. В течение 2012 г. каждый месяц число активных пользователей достигло 90 тыс. 215 человек, а экспертов — 9 тыс. 174 человек. Ежемесячный объем составляет 48 млн 935 тыс. 686 заархивированных записей, прошедших экспертизу — 23 млн 891 тыс. 167.

Приведем другие статистические данные о проекте:

- за последние 12 месяцев (июнь 2011 — май 2012 г.) в проекте принимали участие 278 тыс. 906 активных волонтеров;
- в течение 2012 г. в проекте выполняется 792 тыс. 310 экспертных записей в день;
- каждый год в каталоге публикуется более 200 млн фамилий;
- на сегодняшний день в «Поиске семьи» развивается более 150 проектов (каждую неделю добавляются новые, список текущих проектов доступен по адресу: <http://indexing.familysearch.org/>);
- информационный вход и экспертное программное обеспечение в пользовательском интерфейсе доступны на английском, датском, французском, немецком, итальянском, японском, польском, португальском, русском, испанском и шведском языках (в скором времени предполагается добавить и другие языки);
- в индексации проекта используется сделанное по специальному заказу и основанное на языке JavaScript устройство, а также пользовательское программное обеспечение Java для веб-запуска;
- пакеты с изображениями записей скачиваются волонтерами на свои компьютеры, обрабатываются и после завершения работы загружаются на поисковое устройство проекта (каждая запись обрабатывается двумя волонтерами); процесс экспертизы состоит в пересмотре и установлении различий между записями, составленными двумя волонтерами, после устранения этих различий запись готова к публикации.

«Программа оцифровки австралийских газет» (Australian Newspapers Digitisation Program, <http://trove.nla.gov.au>). Национальная библиотека Австралии инициировала краудсорсинговый проект по корректированию текста архивных газет с помощью программы OCR. К июню 2012 г. откорректировано 68 млн 908 тыс. 757 газетных строк. Общее число зарегистрированных пользователей составило 201 млн 263 тыс. 553 человека. В течение первых шести месяцев 2012 г. ежемесячно около 3,5 тыс. активных зарегистрированных пользователей откорректировали 245 тыс. газетных строк.

«Цифровая коллекция газет Калифорнии» (California Digital Newspaper Collection (CDNC); <http://cdnc.ucr.edu>). Университет Кали-

форнии в Риверсайде принимает участие в «Национальной программе Библиотеки Конгресса США по оцифровке газет» (the Library of Congress National Digital Newspaper Program) с момента ее начала в 2005 году. К июню 2012 г. в CDNC насчитывалось 55 тыс. 970 оцифрованных выпусков газет, что составило около 500 тыс. страниц. Веб-сайт коллекции предоставляет доступ к постатейному поиску газет и их алфавитному указателю. На сайте Библиотеки Конгресса США «Chronicling America» также размещен алфавитный указатель этих газет с постраничным поиском. В августе 2011 г. CDNC добавила пользовательский модуль для корректирования текста с помощью программы OCR к своему библиотечному программному обеспечению для оцифровки под названием Veridian. К июлю 2012 г. 297 активных зарегистрированных пользователей исправили более 395 тыс. строк текста.

Среди других проектов по оцифровке газет необходимо упомянуть проект «Цифровой раз-мер» (*Digitalkoot*; <http://www.digitalkoot.fi>) Национальной библиотеки Финляндии. Направленный на корректировку текста с помощью OCR, он отличается от других проектов тем, что работа выполняется в форме игры. К июлю 2012 г. игроки потратили 407 тыс. 734 минуты своего времени на исправление текста.

Проект «Распределенные корректоры» (*Distributed Proofreaders*; <http://www.pgdp.net>), входящий в состав проекта «Гутенберг» (*Project Gutenberg*), является самым длительным краудсорсинговым проектом по оцифровке культурного наследия и, возможно, самым длительным действующим проектом. Он был начат Ч. Фрэнксом (Charles Franks) в 2000 г. и официально стал частью проекта («Гутенберг») в 2002 году. По данным на 8 июля 2012 г. волонтеры добавили более 40 тыс. откорректированных текстов, находящихся в общественной собственности.

Целью проекта «Что в меню?» (*What's on the menu?* <http://menus.nypl.org/>) Публичной библиотеки Нью-Йорка является расшифровка блюд в 45 тыс. меню, начиная с 1840 г. до наших дней. К июлю 2012 г. было описано 974 тыс. 329 блюд из 14 тыс. 640 меню.

Проект «Расшифровывающая Бентама» (*Transcribe Bentham*; http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham) является краудсорсинговым проектом, управляемым Университетским колледжем Лондона. Его целью является привлечение широких слоев общественности к онлайн-транскрибированию оригинальных и неизученных рукописей Дж. Бентама (Jeremy Bentham, 1748—1832), великого философа и реформатора».

Управление национальных архивов и исторических записей США недавно приступило к реализации пилотного проекта под названием Citizen

Archivist Dashboard (<http://www.archives.gov/citizen-archivist/>), имеющего целью транскрибировать документы XVIII—XX веков. Документы представляют собой различные письма, президентские записи, избирательные петиции, досье о беглых рабах и классифицируются по уровням сложности: начальному, промежуточному и более усложненному.

Мотивации участия в краудсорсинге

Оцифровка коллекций исторических газет пользуется популярностью у специалистов по генеалогии. Несколько лет назад Национальная библиотека Новой Зеландии провела опрос пользователей «Коллекции старинных документов» (*Papers Past Collection*). Результат показал, что более 50% используют ее в целях поиска семейных архивов.

В отчете за 2010 г. о пользователях проекта «Находка» (*Trove*) Национальная библиотека Австралии предоставила данные о том, что 50% пользователей являются исследователями семейных архивов, возраст более половины из них достигает 55 лет и старше [14].

Отчет о пользователях коллекции «Цифровые газеты Юты» (*Utah Digital Newspapers*) за март-апрель 2012 г. показал, что около 70% из них используют эту коллекцию в генеалогических целях [7].

Б. Гейгер (Brian Geiger) провел неформальное исследование пользователей «Коллекции цифровых газет Калифорнии» (*the California Digital Newspapers Collection*). Свои результаты он озвучил на заседании Группы по интересам газет (*the Newspaper Interest Group*) на ежегодной Конференции Американской библиотечной ассоциации (ALA) в 2012 г.: из 136 опрошенных более 75% имеют возраст 50 лет и старше, а около половины — интересуются семейными архивами [6].

Алфавитный указатель проекта «Поиск семьи» востребован большей частью в генеалогических целях, и хотя руководство проекта до сих пор не опубликовало отчет о его пользователях, подобно отчетам коллекций Калифорнии, Новой Зеландии и Австралии, можно утверждать, что они тоже используются специалистами по генеалогии и семейной истории.

Что же говорят сами пользователи о процессе исправления текста? В отчете Национальной библиотеки Австралии о корректорах текста проекта «Находка» под названием «Много рук делают легкую работу» (*Many hands make light work*), представленном Р. Холли (Rose Holley), говорится: «Я наслаждаюсь процессом корректирования — это лучший путь узнать больше об истории и интересных фактах. Кроме того, корректируя текст, я оказываю услугу сообществу»; «Мы устали от домашней работы и выполняем этот процесс, так как он очень привлекателен и помогает нам» [9].

Два корректора текста, работающих в проекте «Цифровые коллекции газет Калифорнии» высказываются следующим образом: «Мне очень интересно принимать участие в оцифровке различных историй. Я уверен, что генеалогия стала моим хобби на многие годы. Я корректирую текст в этой коллекции, так как считаю ее очень полезным проектом»; «Я корректирую текст статей местного значения, а не государственного, национального или международного уровня. Главная цель — помочь исследователям в их поисках людей, мест, организаций, информации о разных событиях, используя онлайн-доступ в коллекции».

В книге «Когнитивный избыток: креативность и щедрость в эру связи» (Cognitive Surplus: Creativity and Generosity in a Connected Age) К. Ширки (Clay Shirky) отмечает, что сегодня люди учатся использовать свое свободное время для создания чего-либо вместо потребления. С помощью некоторых приблизительных расчетов автор вычислил, что общее число познавательных усилий, затраченных людьми при создании всех Википедий на каждом языке, составляет около миллиона часов. Более того, американцы каждый год тратят два миллиарда часов на просмотр ТВ, а этого было бы достаточно для создания 2 тыс. проектов, подобных Википедии [18].

Однако только один из пяти принципов Дж. Хау, лежащих в основе «нового объединения труда», можно применить к корректировке текста в проектах «Находка» и «Цифровые коллекции газет Калифорнии»: толпа (корректоры текста) рассеяна в пространстве. Второй принцип Хау (ее внимание сосредоточено на конкретном небольшом участке работы) применим только к некоторым корректорам текста, кто проверяет небольшое количество строк и никогда более не возвращается к этому процессу. Безусловно, этот принцип совершенно не подходит тем, кто проверяет ежемесячно по 1 тыс. строк. Автор полагает, что для неоплачиваемых волонтеров наиболее сильной мотивацией является тот факт, что краудсорсинг прост и интересен, дает ощущение причастности, расширяет круг общения (не только с волонтерами).

Редактор статьи под названием «Повышение мотивации» (Increase Motivation) [11] перечисляет мотивационные факторы, многие из которых общеизвестны:

1. Ожидаемые последствия. Никогда не применяйте угрозы. Они повернут людей против вас. Однако предупреждение о негативных последствиях в случае отсутствия результатов (для каждого участника) может дать хороший эффект. Этот фактор имеет большое значение и для собственной мотивации. Если вы не можете добиться результата от совместной работы, то сможете ли вы вообще чего-нибудь добиться?

2. Приятное поощрение. Это старая приманка в технике «кнута и пряника». Поощрения вызывают у людей желание работать.

3. Создание системы стимулирования. Этот призыв для людей с эгоистичной натурой. Дайте им возможность больше получать для себя путем большей отдачи для вас.

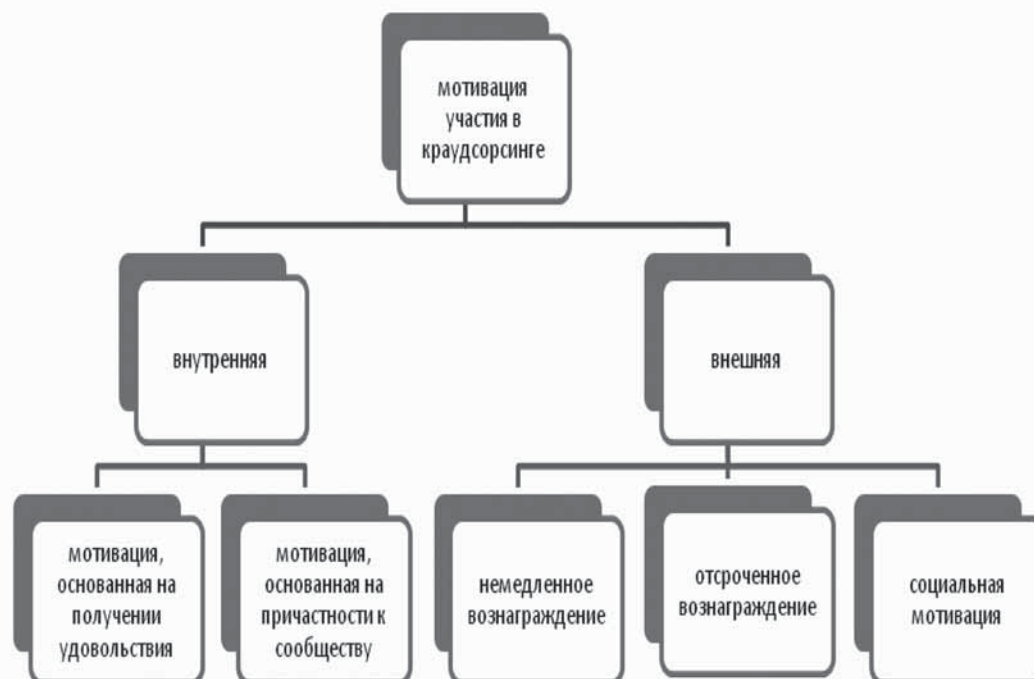
4. Детальный инструктаж. Если вы хотите добиться конкретного результата, то дайте и конкретные инструкции к действию. Люди лучше работают, если они знают, что можно ожидать от их работ.

5. Краткосрочные и долгосрочные цели. Ставьте цели как на короткий, так и на длительный срок, чтобы управлять процессом и создать целостную философию.

6. Доброта. Привлеките людей на свою сторону, и они захотят помочь вам.

7. Конечные сроки. Многие люди работают с большей отдачей, если перед ними четко обозначить конечные сроки, они лучше концентрируют свое внимание, если перед ними все время «маячит» конечный срок. Используйте это в своих целях для получения результата.

8. Командный дух. Создайте обстановку товарищества. Люди работают более эффективно, если они ощущают себя частью команды.



9. Оцените результат. Оценивайте каждый результат в отдельности, а также общий в группе. Людям нравится, когда их работа не игнорируется.

10. Личная потребность. Принимайте во внимание личную потребность каждого. Что им необходимо? Только поняв это, вы будете способны заставить людей работать продуктивнее.

11. Концентрация на результатах. Никто не любит работать за других. Сделайте акцент на результатах — разьясните что вы хотите от людей и добейтесь, чтобы они работали по своему усмотрению.

12. Доверие и уважение. Уважайте людей, как они того заслуживают, и они ответят вам тем же.

13. Ставьте сложные задачи. Люди счастливы, когда они поступательно движутся к цели. Дайте им возможность столкнуться с новой трудной задачей, и у них прибавится энтузиазма.

14. Позвольте людям созидать. Не ожидайте от каждого, что он будет идти вашим путем, создайте для него креативную обстановку, это придаст людям оптимизма и может привести к возникновению новых идей.

15. Конструктивная критика. Очень часто люди не понимают, что делают что-то неправильно. Скажите им об этом. Большинство людей хотят улучшить свой результат и сделают над собой усилие.

16. Требуйте постоянного совершенства. Не позволяйте людям стоять на месте, каждый раз кто-то поднимает планку все выше (особенно для себя самого).

17. Сделайте процесс веселым. Работа становится приятной, когда она вовсе не ощущается как работа. Создайте людям веселую обстановку, и позитивное окружение даст лучшие результаты.

18. Создавайте возможности. Предоставьте людям возможность двигаться вперед. Дайте им знать, что тяжелая работа окупится сполна.

19. Общение. Держите каналы общения открытыми. Во избежание потенциальных проблем вы можете предупредить их заранее до возникновения каких-либо конфликтов.

20. Стимулирование. Добавьте разнообразия в процесс. Не заставляйте людей выполнять все время одни и те же скучные задачи. Стимулирование окружения прибавляет энтузиазма и создает возможность мыслить масштабно.

П. Органичек (Peter Organisciak), соискатель степени доктора библиотечно-информационных наук в Университете Иллинойса, исследующий проблемы краудсорсинга, упоминает о точно таких же мотивациях: деньги, увлечение, скука, завершенность, благотворительность, ученое окружение, участие, польза для себя, ускоренный темп, интерес [16].

Последние теоретические разработки о мотивациях процесса краудсорсинга сделаны Н. Кауфманом (Nicholas Kaufmann) и соавторами [12] (см. схему). Согласно теории мотивации, мотивы человеческого поведения разделяются на внутренние и внешние. Внутренней «называется мотивация, порождаемая интересом или получением удовольствия от выполнения задачи как таковой, она существует скорее внутри личности, чем порождается извне», с другой стороны, внешняя мотивация «относится к деятельности, имеющей целью достижение результата» [15]. Мы полагаем, что именно внутренняя мотивация является доминирующей в краудсорсинговых проектах по оцифровке культурного наследия.

Краудсорсинг, как и все, имеет стоимость и цену. Некоторые его аспекты, например такие, как количество проверенных строк, число активных и зарегистрированных пользователей, количество посещений веб-сайта, легко измерить. Другие аспекты, особенно те, которые не так просто оценить, являются самыми сложными. Рассмотрим более легкие вопросы, касающиеся работы с персоналом.

Расходы, которых можно избежать. Очевидная польза от краудсорсинговых проектов по корректированию OCR-текста состоит в улучшенном поиске. Это особенно важно для коллекций оцифрованных газет, так как точность OCR-текста часто бывает очень плохой. Например, Э. Кильджин (Edwin Kiljin) сообщает, что неточность в газетах начала XX в. достигает 68% [13].

В образце, состоящем из 45 страниц оцифрованных газет периода 1803—1954 гг. проекта «Находка», она варьируется от 71 до 98% [8].

Конечно, процесс исправления текста можно доверить аутсорсинговым агентствам, как это сделали в Австралии, Новой Зеландии, Сингапуре и Калифорнии. Но так как сам процесс является затратным, то все корректирование было ограничено газетными заголовками, или же, как например в проекте «Находка», к ним были добавлены первые четыре строчки некоторых статей.

Сделаем примерный расчет, используя число откорректированных строчек газетного текста в проектах CDNC и «Находка». В зависимости от года выпуска газеты, числа столбцов, размера шрифта и макета в каждой строке газетного столбца насчитывается от 25 до 50 знаков. Предположим, что мы имеем 35 знаков в каждой строке. В зависимости от стоимости услуг агентства, цена аутсорсинга процесса исправления текста с точностью до 99,5% варьируется от 0,35 долл. США до 1 долл. США за 1 тыс. знаков (для примерного расчета возьмем 0,5 долл. США за 1 тыс. знаков). По данным на 12 июля 2012 г. волонтеры коллекции CDNC откорректировали 394 тыс. 365 строчек текста. Можно определить, что стоимость труда волонтеров этой коллекции составляет:

$394\,365 \text{ строчек} \times 35 \text{ знаков} : 1000 \text{ знаков} \times 0,5 \text{ долл. США} = 6\,901 \text{ долл. США.}$

Подобный расчет для проекта «Находка» НБ Австралии, где волонтеры откорректировали 69 млн 918 тыс. 892 строчки текста (по данным на 12 июля 2012 г.), показал результат оценки их труда в 1 млн 223 тыс. 581 долл. США. Очень значительные суммы!

Как мы увидим далее, оценка стоимости труда волонтеров не является самым большим преимуществом краудсорсинга.

Неизбежные расходы. Что включается в оценку краудсорсингового проекта? Это, прежде всего, оценка самого программного обеспечения, программно-технической инфраструктуры, ее поддержки и администрирования, Интернета и коммуникаций. Не менее важными являются: служба технической поддержки, состоящая из волонтеров, отвечающих на часто задаваемые вопросы; маркетинг и распространение информации о поиске волонтеров для проекта; применение руководств и передовых методик.

Такая оценка в значительной степени зависит от масштаба проекта и заработной платы в нем. Например, проект «Поиск семьи» является большим подразделением еще большей организации, имеющей бюджет в миллионы долларов. С другой стороны, поддержка программного обеспечения цифровой библиотеки проекта CDNC занимает менее 20% времени системного администратора, а сам проект имеет годовой бюджет менее 20 тыс. долл. США.

Затраты на программное обеспечение в краудсорсинге легко рассчитать для коммерческих коробочных программных продуктов (COTS) и сложно — для программного обеспечения (ПО), создаваемого на заказ.

Именно таковым является в настоящее время ПО для оцифровки культурного наследия методом краудсорсинга.

По самым приблизительным подсчетам мы получили следующее: неограниченная лицензия на ПО цифровой библиотеки Veridian, а также ПО для корректирования OCR-текста стоит 33 тыс. долл. США в год. Если включить в эту сумму очевидные и менее очевидные расходы, то получается ежегодная стоимость в 49 тыс. 500 долл. США. Однако настоящая цена может быть значительно меньшей для маленьких проектов с ограниченной лицензией и значительно большей — для ПО, создаваемого на заказ для больших проектов. В этом случае более высокими будут затраты на его поддержку.

Повышенная точность. Необходимо понимать, что точность необработанного текста сильно варьируется и часто бывает очень низкой. Предположим, что примерная точность знаков достигает 90%. Средняя длина слова в английском языке — пять знаков. Это означает, что слова имеют среднюю точность в 59% или то, что только шесть слов из десяти в необработанном тексте будут откорректированы. А по самому оптимистичному предположению 95-процентная точность текста на самом деле оказывается только 77-процентной. Так как средняя длина предполагает и такие слова, как артикли, то средняя длина самых «интересных» слов для поиска, например, персоналии, названия мест и организаций, будет больше, а значит точность — меньше.

В информационном поиске слово «точность» (precision) означает часть искомого объекта, соответствующего поиску, а слово «отзыв» (recall) — часть соответствующего найденного объекта. Самый лучший счет для этих двух слов является 1.0. Превосходная точность означает, что ничего не соответствующего не найдено, а превосходный отзыв — все соответствующее найдено.

Насколько известно, еще никто не измерил точность откорректированных OCR-текстов (или расшифрованных рукописей) в краудсорсинговых проектах. Однако операторы аутсорсинговых агентств обычно исправляют текст с точностью до 99,5%, поэтому вполне резонно предположить, что эта же точность возможна и для наших проектов. По крайней мере, специалисты по генеалогии, составляющие большинство волонтеров — корректоров текста, будут более тщательно подходить к исправлению фамилий. Точность в 99,5% означает точность английского слова средней длины в 97,5%.

Реальная польза от краудсорсинга. В своем блоге Т. Оунс (Trevor Owens) выдвигает версию о том, что реальная польза от корректирования или расшифровки текста в краудсорсинге заключается не в улучшенном поиске или поисках с повышенной точностью и отзывом. По его мнению, она

состоит в осмысленной деятельности и возможности для волонтеров внести свой вклад [17].

Краудсорсинг лучше использовать в цифровых коллекциях, чем в отображении их на экранах. Он предоставляет главную возможность, которую не даст ни одна цифровая платформа, — возможность для человека сделать что-то большее, чем просто потреблять информацию.

Этот процесс предлагает личности осмысленные пути для вовлечения в общественную жизнь.

В последнее время у организаций, хранящих культурное наследие, возрос интерес к методу краудсорсинга. Например, этот метод используется компанией Netflix в целях улучшения алгоритмов ее видеорекомендаций; Канадской горной компанией — в поисках золота; компанией Kickstarter — в целях изыскания средств на инновационные проекты (включая и библиотечные!), а также Galaxy Zoo — для классификации миллиона галактик на аппаратуре Sloan Digital Sky Survey. Кто же мог рассчитывать, что организации, хранящие культурное наследие, будут использовать краудсорсинг для повышения своей значимости в глазах местных сообществ?

Мы полагаем, что технологии Web 2.0 и метод краудсорсинга предоставляют библиотекам интересные возможности, позволяющие вовлечь в библиотечный процесс пользователей.

СПИСОК ИСТОЧНИКОВ

1. Краудсорсинг // Википедия [Электронный ресурс]. — Режим доступа: <http://ru.wikipedia.org/Краудсорсинг>
2. *Brabham D.C.* Crowdsourcing as a Model for Problem Solving : Leveraging the Collective Intelligence of Online Communities for Public Good : PhD Dissertation / D.C. Brabham. — University of Utah, 2010.
3. Crowdsourcing // Wikipedia. — URL: <http://en.wikipedia.org/wiki/Crowdsourcing>
4. *Estellés Arolas E.* Clasificación de iniciativas de crowdsourcing basada en tareas / E. Estellés Arolas, F. González Ladrón de Guevara // El profesional de la información. — 2012. — Vol. 21. — № 3. — P. 283—291.
5. *Idem.* Towards an integrated crowdsourcing definition / E. Estellés Arolas, F. González Ladrón de Guevara // Journal of Information Science. — 2012. — Vol. 38. — № 2. — P. 189—200.
6. *Geiger B.* Improving the California Digital Newspaper Collection Software [Monkey survey] // 2012 ALA Annual Conference (Anaheim, CA, USA, June 21—26, 2012).
7. *Herbert J.* Small Town Papers : Still Delivering the News / J. Herbert, R. Olsen // 2012 IFLA General Conference (Helsinki, Finland, August 11—17, 2012).
8. *Holley R.* How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs // D-Lib Magazine. — 2009. — March/April. — URL: <http://www.dlib.org/dlib/march09/holley/03holley.html>

9. *Idem*. Many Hands Make Light Work. — URL: http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf
10. *Howe J.* The Rise of Crowdsourcing // *Wired*. — 2006. — № 14.06. — URL: <http://www.wired.com/wired/archive/14.06/crowds.html>
11. Increase Motivation. — URL: <http://www.pickthebrain.com/blog/21-proven-motivation-tactics>
12. *Kaufmann N.* More than Fun and Money : Worker Motivation in Crowdsourcing : A Study on Mechanical Turk / N. Kaufmann, T. Schulze, D. Veit // Proceedings of the Seventeenth Americas Conference on Information Systems (Detroit, MI, August 4—7, 2011). — URL: http://schader.bwl.uni-mannheim.de/fileadmin/files/publikationen/Kaufmann_Schulze_Veit_2011-More_than_fun_and_monney_Worker_motivation_in_Crowdsourcing-A_Study_on_Mechanical_Turk_AM-CIS_2011.pdf
13. *Kiljin E.* The Current State-of-Art in Newspaper Digitization // *D-Lib Magazine*. — 2008. — January / February. — URL: <http://www.dlib.org/dlib/january08/klijn/01klijn.html>
14. Libraries Australia Advisory Committee : Trove Report 2010. — URL: <http://www.nla.gov.au/librariesaustralia/files/2011/11/laac-paper-2010-2-08-trove-report.pdf>
15. Motivation // *Wikipedia*. — URL: <http://en.wikipedia.org/wiki/Motivation>
16. *Organisciak P.* Motivation of Corwds : The Incentives that Make Crowdsourcing Work. — URL: <http://crowdstorming.wordpress.com/2008/01/31/motivation-of-crowds-the-incentives-that-make-crowdsourcing-work/>
17. *Owens T.* Crowdsourcing Cultural Heritage : The Objectives Are Upside Down. — URL: <http://www.trevorowens.org/2012/03/crowdsourcing-cultural-heritage-theobjectives-are-upside-down/>
18. *Shirky C.* Cognitive Surplus : Creativity and Generosity in a Connected Age / C. Shirky. — New York : Penguin Press, 2010.
19. *Surowiecki J.* The Wisdom of Crowds / J. Surowiecki. — New York : Random House, 2004.

Анонс

ПРЕДСТОЯЩИЕ КОНГРЕССЫ ИФЛА

• **Всемирный библиотечный и информационный конгресс — 79-я Генеральная конференция и Ассамблея ИФЛА** на тему: «Библиотеки будущего: безграничные возможности» (Future Libraries: Infinite Possibilities) состоится 17—23 августа 2013 г. в Сингапуре.

Определены регионы предстоящих в 2014—2019 гг. конгрессов ИФЛА:

- 2014 — г. Лион, Франция
- 2015 — Африка
- 2016 — Северная Америка
- 2017 — Европа
- 2018 — Латинская Америка или страны Карибского бассейна
- 2019 — Европа