

Международная организация по стандартизации: Подготовлен проект технического отчета «Статистика и показатели качества для веб-архивации»

http://rusrim.blogspot.com/2012/12/blog-post_1528.html

December 12, 2012

Международная организации по стандартизации работает на своей первой публикации по теме веб-архивации.

Как сообщил 20 ноября 2011 года чешский блог по электронной сохранности «Digital Preservation CZ», техническим подкомитетом TC 46/SC 8 подготовлен проект технического отчета ISO/DTR 14873 «Информация и документация - Статистика и показатели качества для веб-архивации» (Information and documentation — Statistics and Quality Indicators for Web Archiving) объёмом 62 стр., который в настоящее время доступен по адресу



http://netpreserve.org/sites/default/files/resources/SO_TR_14873_E_2012-10-02_DRAFT.pdf .

Как отметили чешские коллеги, «Интересен уже вводный раздел, который определяет основные термины в области веб-архивов».

В предисловии в документе, в частности, сказано:

«Настоящий технический отчет был разработан в ответ на существующую во всем мире потребность в руководствах по управлению и оценке деятельности и продуктов в сфере веб-архивации.»

Под веб-архивацией понимается деятельность по отбору, сбору/захвату, хранению, обеспечению долговременной сохранности и управлению доступом во времени к снимкам интернет-ресурсов. Она началась в конце 1990-х годов, исходя из убеждения в том, что в будущем архив интернет-ресурсов может стать важнейшим источником документированной информации для научных исследований, для коммерции и государственного управления. Интернет-ресурсы рассматриваются как часть культурно-исторического наследия и, как следствие, их сохранность обеспечивается так же, как и сохранность относимых к такому наследию печатных изданий. Многие учреждения, занимающиеся веб-архивацией, рассматривают её как расширение своей традиционной миссии по обеспечению сохранности национального культурно-исторического наследия, и во многих странах такое отношение одобряется и получает правовую основу посредством использования таких правовых механизмов, как законодательство об обязательных экземплярах (legal deposit).

В Интернете доступен широкий спектр ресурсов, включая тексты, графические изображения, видео- и звукозаписи и другие мультимедийные форматы. Помимо взаимосвязанных веб-страниц, имеются также новостийные группы, новостийные рассылки, блоги и интерактивные

услуги, такие как игры, доступные с использованием различных транспортно-коммуникационных протоколов. Веб-архивы накапливают копии интернет-ресурсов, регулярно автоматически собираемых с помощью специального программного обеспечения (*harvesting software*). Целью является воспроизведение ресурсов, включая внутренние взаимосвязи (например, в виде гипертекстовых ссылок), как можно ближе к тому, какими они были в оригинальной среде. Основной задачей веб-архивации является обеспечение сохранности во времени документированной информации об Интернете, как можно ближе к её первоначальной форме, для различных целей - научных, профессиональных и личных.

Веб-архивация – это новый, но расширяющийся вид деятельности, постоянно требующий новых подходов и инструментов для того, чтобы идти в ногу с быстро развивающимися веб-технологиями. Для архивации интернет-ресурсов используются различные подходы, - начиная от захвата отдельных веб-страниц и до захвата целиком доменов верхнего уровня, - выбираемые исходя из стратегической важности, придаваемой этой работе архивирующим учреждением, из имеющихся ресурсов, и иногда и из законодательно-нормативных требований. С точки зрения использования организациями, веб-архивация может быть на различных уровнях зрелости. Если в одних организациях она стала элементом обычной деловой деятельности, то другие только лишь начинают экспериментальные программы по изучению данной проблемы.

В зависимости от масштабов и целей веб-архивации, можно выделить две широкие категории стратегий веб-архивации: массовую (*bulk harvesting*) и селективную (*selective harvesting*). Крупномасштабный массовый захват (например, всего национального домена) решает задачу захвата снимка всего домена (или части его). Селективный захват выполняется в гораздо меньших масштабах, является более целенаправленным и проводится чаще, часто основываясь на таких критериях, как тема, событие, формат (например, захват аудио- или видео-файлов) или на соглашениях с владельцами контента. Главное различие между этими двумя стратегиями заключается в уровне контроля качества, - в проведении оценки захваченных веб-сайтов на соответствие заранее установленным стандартам качества. Масштабность работы по захвату домена делает невозможным выполнение какого-либо «ручного» визуального сравнения реального ресурса и его захваченной версии, что является распространенным методом обеспечения качества при селективном захвате.

Настоящий технический отчет призван показать, как деятельность веб-архивов, являющихся частью более обширных коллекций культурно-исторического наследия, можно оценивать и как ею можно аналогичным образом управлять, в соответствии с установленными требованиями, на основе применения традиционных библиотечных рабочих процессов. В техническом отчете рассматриваются вопросы развития коллекций, определения характеристик, описания, обеспечения долговременной сохранности, использования, а также организационной структуры; и показывается, что для веб-архивации остаётся, в принципе, в силе большинство аспектов традиционных процессов управления коллекциями, - хотя на практике требуется их определенная «подстройка».

К настоящему техническому отчету следует относиться как к продолжающейся работе. Предполагается, что часть его содержания в

будущем войдёт в стандарты ISO 2789 «Информация и документация - Международная библиотечная статистика» (Information and documentation – International library statistics) и ISO 11620 «Информация и документация - Показатели эффективности работы библиотек» (Library performance indicators).

Хотя настоящий технический отчет содержит обзор текущего состояния веб-архивации, он в основном сосредоточен на вопросах определения и использования статистики и показателей качества веб-архивов. Получение определенных статистических данных опирается на возможности используемого для захвата, индексации или просмотра программного обеспечения, и иной выбор программного обеспечения может привести к отличающимся результатам. Технический отчет, однако, не одобряет и не рекомендует какое-либо конкретное программное обеспечение. В нем предложен набор показателей, помогающих оценить эффективность и качество веб-архивов в целом.»

Содержание технического отчета следующее:

Предисловие

Введение

1 Область применения

2 Нормативные ссылки

3 Термины и определения

4 Методы и задачи веб-архивации

4.1 Методы захвата

4.2 Методы доступа и описания

4.3 Методы обеспечения долговременной сохранности

4.4 Правовая основа веб-архивации

4.5 Дополнительные мотивы для веб-архивации

5 Статистика

5.1 Общие положения

5.2 Использование статистики для развития коллекций

5.3 Характеризация коллекции

5.4 Использование коллекции

5.5 Обеспечение долговременной сохранности веб-архива

5.6 Оценка затрат на веб-архивацию

6 Качество

6.1 Общие положения

6.2 Ограничения

6.3 Описание

7 Использование и получаемая отдача

7.1 Общие положения

7.2 Целевое использование и пользователи

7.3 Отдача для групп пользователей

7.4 Использование предлагаемой статистики группами пользователей

7.5 Процесс веб-архивации и соответствующие показатели эффективности

Библиография

Источник: блог «Digital Preservation CZ»

<http://digital-preservation-cz.blogspot.ru/2012/11/indikatory-kvality-pro-webove-archivy.html>