

И. И. Гендина

*ИИИ информационных технологий социальной сферы
Кемеровского государственного университета культуры и искусств*

И. И. Колкова, О. И. Алдохина

Кемеровский государственный университет культуры и искусств

Использование формализованных методов при подготовке текста для веб-сайтов учреждений культуры

Рассмотрены типичные недостатки контента веб-сайтов учреждений культуры. Предложено использовать формализованные методы анализа и синтеза информации для построения качественного контента веб-сайта. Раскрыта суть методики формализованного составления текстов для отражения рубрик веб-сайтов учреждений культуры.

В современных условиях веб-сайты учреждений культуры становятся мощным средством обеспечения доступа к культурному наследию и информационным ресурсам широкого круга удаленных пользователей, на практике претворяя в жизнь важнейший принцип информационного общества – всеобщий доступ к информации, образованию и культурным ценностям.

Полноценное выполнение веб-сайтом его сущностных функций прежде всего определяется качеством его информационного наполнения (контента). Проблемы обеспечения высокого качества контента веб-сайтов учреждений культуры в настоящее время являются предметом пристального внимания профессиональной общественности не только на российском, но и на международном уровне. Один из важнейших результатов работы в этом направлении – выработанный в рамках проекта Minerva «Определение потребностей пользователей, содержания и критериев качества веб-сайтов по культуре» документ «Принципы качества веб-сайтов по культуре. Руководство» [1]. В нем представлены десять принципов качества веб-сайтов музеев, библиотек и других учреждений культуры. Важнейшими из них провозглашаются прозрачность, эффективность, ориентация на пользователя. Их реализация предполагает, что информационное наполнение высококачественного веб-сайта должно быть релевантным, простым и доступным для пользователя. Однако ни в отечественной, ни в зарубежной литературе пока нет ответов на вопросы о том, каким образом можно обеспечить дос-

тижение таких потребительских качеств веб-сайта, какова технология подготовки качественного контента сайта.

Создание контента веб-сайта неразрывно связано с подготовкой текстов для наполнения его рубрик и подразбук. Требования, предъявляемые к текстам для веб-сайтов, могут быть подразделены на общие (относящиеся ко всем информационным ресурсам) и специфические (характерные для сетевых информационных ресурсов). К числу общих требований, предъявляемых ко всем текстам, включая тексты для веб-сайтов, относятся полнота, точность, логичность, последовательность, структурированность, четкость и ясность изложения материала, соответствие литературным нормам русского языка. Специфическими требованиями, обусловленными сетевой формой представления информации, являются лаконичность («телеграфный стиль»), порционность и обязательная рубрицированность, диверсификация различных стилей представления информации (официального, научного, публицистического, разговорного) в зависимости от типа сайта и выполняемых им функций, гипертекстовая форма представления информации (преодоление линейности обычных текстов).

Проведенный в НИИ информационных технологий социальной сферы Кемеровского государственного университета культуры и искусств (КемГУКИ) анализ показывает, что недостатки контента существующих веб-сайтов различных учреждений культуры связаны с нарушением требований обеих групп. Типичными недостатками являются неполнота либо избыточность информации; разнородность и неструктурированность информации; несоответствие языка и стиля представления информации статусу, типу сайта и типу рубрик контента сайта; погрешности языка и стиля; затрудненность восприятия информации; монотонность подачи информации; преобладание текста в ущерб иллюстративному материалу; перенос в состав контента веб-сайта типичной для традиционных документов, но не характерной для веб-пространства линейной формы представления информации. Безусловно, что веб-сайты, контент которых обладает такими недостатками, даже при релевантности его содержания запросам пользователей не будут оценены ими как качественные.

Эти недостатки во многом объясняются тем, что созданием контента сайтов зачастую занимаются специалисты, не имеющие профессиональной подготовки в области работы с текстами. Однако и для тех, кто обладает базовыми филологическими, библиотечно-информационными, редакторскими знаниями и умениями, работа над текстами для веб-сайтов представляет собой весьма сложную и трудоемкую задачу, решение которой не обеспечено необходимым теоретическим, технологическим и методическим инструментарием. Ситуация усугубляется хаотичностью лексики сай-

строения, диффузионностью классификации веб-сайтов, несовершенством системы оценки качества веб-сайтов, фрагментарностью описания технологии создания сайта, отсутствием типовых схем контента веб-сайтов, неразработанностью справочной и инструктивно-методической базы для создания веб-сайтов. Не случайно в подготовке их текстов доминируют эмпирически-описательные методы.

Одним из путей повышения качества текстов для веб-сайтов, на наш взгляд, является использование методов формализации, в частности формализованных методов информационного анализа и синтеза информации [2–8]. Формализованные методы вошли в информационную практику как реакция на сверхбольшие объемы информации, явившиеся следствием информационного взрыва XX столетия, и невозможность в массовом технологическом производстве решить проблему аналитико-синтетической переработки информации за счет интеллектуальных методов.

Повышение качества подготовки вторичных документов за счет снижения субъективизма библиотечно-информационных специалистов, призванных каждодневно решать сложную интеллектуальную задачу выявления в тексте первичного документа основного смыслового содержания и отделения существенной информации от несущественной; обеспечение возможности четкой ориентации пользователей в текстах вторичных документов, оперативной узнаваемости их сущности; массовая воспроизводимость результатов формализованного свертывания; уменьшение временных, интеллектуальных и иных трудозатрат на аналитико-синтетическую переработку информации – все эти причины обусловили популярность и востребованность формализованных методов в библиотечно-информационной практике.

В библиотечно-информационной деятельности формализованные методы работы с информацией, как правило, ассоциируются с процессами свертывания информации и созданием различных вторичных документов. Традиционными областями применения формализованных методов, в которых убедительно доказана их эффективность, являются технологии свертывания информации: индексирования, аннотирования и реферирования, составления обзоров, успешно реализуемые на массивах как научно-технических, так и публицистических текстов.

Вместе с тем формализованные методы способны обеспечить существенное повышение эффективности работы с информацией и при использовании их в технологиях развертывания информации, связанных с подготовкой научных статей, докладов, диссертаций и других первичных документов. Использование метода формализации олицетворяет осознанное стремление подняться над эмпирически-описательными методами конструирования

ния текстов и на этой основе обеспечить повышение их качества. Суть методики формализованного составления текста заключается в ориентации на заранее выделенный перечень относительно самостоятельных смысловых фрагментов – аспектов содержания. При этом для каждого аспекта устанавливается определенная совокупность только ему присущих формальных признаков (маркеров), однозначно определяющих его содержание. Маркеры могут быть выражены устойчивыми словами или словосочетаниями фрагментами фраз.

Термин *формализация* в контексте рассматриваемых задач может трактоваться как аспектно-маркерный метод представления содержания определенной тематической рубрики или подрубрики контента веб-сайта с использованием перечня заранее заданных для нее аспектов и соответствующих им маркеров. Использование аспектно-маркерного метода представления содержания рубрики или подрубрики контента веб-сайта предполагает предварительное определение вида рубрики. Как правило, для описания контента сайта используются рубрики двух видов: справочно-информационного характера («Новости», «Мероприятия», «Проекты», «Персоны» и т. п.) и коммуникативного характера («Вопрос – ответ», «Наиболее часто задаваемые вопросы», «Заказ» и др.). Следует заметить, что рубрики справочно-информационного характера в составе контента веб-сайта могут отражать объекты как единичного («История музея», «История библиотеки», «История филармонии», «История театра» и т. п.), так и собирательного характера («Новости», «Персоны», «Проекты» и т. д.). При этом отражение объектов как единичного, так и собирательного характера в составе контента веб-сайта с помощью аспектно-маркерного метода сводится к выполнению несложного алгоритмического предписания, предусматривающего ознакомление с унифицированным описанием конкретной рубрики (или подрубрики), включающим максимально полный перечень всех аспектов ее содержания; выбор из имеющегося перечня аспектов содержания, подлежащих отражению в составе текста той или иной рубрики (подрубрики), конструирование в соответствии с заданным перечнем аспектов содержания текста рубрики (подрубрики) и соответствующих им маркеров текста рубрики (подрубрики) веб-сайта.

Составление унифицированных описаний может быть обеспечено с помощью полного заимствования готовых схем описаний объектов, разработки собственных схем описаний объектов, сочетания заимствованных и самостоятельно разработанных схем описания объектов (комбинированный способ). Использование комбинированных способов структурирования описаний объектов осуществляется в тех случаях, когда набор представ-

ленных в заимствованных описаниях характеристик объекта оказывается недостаточным для его отражения на веб-сайте.

Основой для составления такой унифицированной схемы с помощью метода заимствования могут стать методические рекомендации открытой энциклопедии «Рубрикана», размещенной на веб-сайте «Рубрикон» (www.rubricon.com). Этот сетевой информационный ресурс включает значительный массив энциклопедических описаний различных объектов культуры и искусства: театров, библиотек, цирков, кинотеатров, архивов и т. п. Наряду с энциклопедическими описаниями открытой энциклопедии «Рубрикана» в качестве источников для выявления аспектов, раскрывающих содержание тех или иных рубрик, могут также использоваться нормативные, справочные, учебные и научные документы и ИПЯ (УДК, ББК, тезаурусы и т. п.).

Разработка собственных схем описаний объектов предполагает формирование массива веб-сайтов, содержащих однотипные рубрики, например рубрику «Новости». Затем тексты, отражающие данную рубрику на всех анализируемых веб-сайтах, подлежат формализованному анализу с целью выявления максимально полного перечня аспектов содержания. Выявленные аспекты упорядочиваются в ранжированный ряд в зависимости от частоты встречаемости. Наиболее часто встречаемые аспекты упорядочиваются в соответствии с семантической формулой. Например, для получения унифицированной схемы описания рубрики «Новости» нами проанализированы по десять новостных сообщений, представленных на тридцати веб-сайтах учреждений культуры (театры, музеи, библиотеки). Таким образом, формализованному анализу подлежало 300 новостных сообщений.

В ходе анализа в тексте каждого новостного сообщения выделялись все представленные в нем аспекты содержания. В результате выявлены семь аспектов содержания, которые использовались в большинстве (более 50%) проанализированных сообщений. Эти аспекты в обобщенном виде могут быть представлены семантической формулой: «Когда? + Где? + О чем? + Кто? + Что сделано (Что представлено)? + Как? + Поэтому». В соответствии с этой формулой унифицированная схема текста рубрики «Новости» любого веб-сайта учреждения культуры должна включать следующие аспекты: предмет рассмотрения (событие, тема, проблема); место происходящего события; время происходящего события; характеристика события; лица, принимающие участие в событии; результаты, выводы; оценка события.

В качестве примера ниже приведена аспектная структура новостного сообщения, построенного в соответствии с данной выше семантической формулой. Основой для ее построения был один из фрагментов рубрики «Новости», представленный на сайте одного из объектов нашего исследования.

Аспект содержания	Пример
Время происходящего события	27 апреля 2007 г.
Место происходящего события	Новгородская областная научная библиотека
Предмет рассмотрения (событие, тема, проблема)	Презентация электронного альбома «Особо охраняемые природные территории Новгородской области»
Лица, принимающие участие в событии	В презентации приняли участие студенты третьего курса Новгородского государственного университета
Характеристика события	После презентации прошло обсуждение вопросов, связанных с возможностью использования электронного альбома для развития туристической деятельности в Новгороде
Результаты, выводы	Известные в Новгороде туристические фирмы предложили новые маршруты, связанные с охраняемыми территориями
Оценка события	Презентация получила высокую оценку экспертов в области туризма и охраны природы, а также специалистов Комитета культуры и туризма Новгородской области

В виде связного текста эти сведения выглядят так:

«27 апреля 2007 г. в Новгородской областной научной библиотеке состоялась презентация электронного альбома «Особо охраняемые природные территории Новгородской области». В презентации приняли участие студенты третьего курса Новгородского государственного университета. После презентации прошло обсуждение вопросов, связанных с возможностью использования электронного альбома для развития туристической деятельности в Новгороде. Известные в Новгороде туристические фирмы предложили новые маршруты, связанные с охраняемыми территориями. Презентация получила высокую оценку экспертов в области туризма и охраны природы, а также специалистов Комитета культуры и туризма Новгородской области».

Использование аспектно-маркерного метода как средства формализации подготовки текстов для веб-сайтов позволяет обеспечить построение качественного контента веб-сайта, удовлетворяющего требованиям целостности и полноты его содержания. Вместе с тем наличие унифицированных в пределах веб-сайта схем описания объектов позволяет четко структурировать материал, делает его «прозрачным», тем самым значительно облегчает восприятие информации для пользователя за счет логичности, последовательности, четкости и ясности изложения описания объектов сайтостроения. Аспекты содержания, явно выраженные в тексте той или иной рубрики веб-сайта, могут служить основой для формирования системы гиперссылок и тем самым для преодоления линейности исходных текстов.

Паряду с использованием для построения рубрик контента аспектно-маркерный метод может применяться и для совершенствования разработки схемы поиска информации, представленной в составе контента веб-сайта. При этом создание контента веб-сайта становится не только менее сложным, но и менее дорогостоящим делом.

СПИСОК ИСТОЧНИКОВ

1. **Принципы качества веб-сайтов по культуре.** Руководство / ред. М. Т. Н. Темпера, А. Темпера – М., 2006. – 61 с.
2. **Гендина, Н. И.** Выявление аспектов содержания документов с помощью формальных текстовых признаков // НТИ. Сер. 2. Информационные процессы и системы. – 1979. – № 4. – С. 7–12.
3. **Гендина, Н. И.** Реферирование // Справочник библиографа / науч. ред. А. Н. Ванеев, Н. А. Минкина. – 2-е изд., перераб. и доп. – СПб., Профессия, 2003. – С. 375–387.
4. **Гендина, Н. И.** Аннотирование / Н. И. Гендина, В. П. Леонов // Справочник библиографа / науч. ред. А. Н. Ванеев, В. А. Минкина. – 2-е изд., перераб. и доп. – СПб., Профессия, 2003. – С. 362–374.
5. **Гендина, Н. И.** Формализация свертывания краеведческой информации как способ повышения качества информационных продуктов и услуг: результаты исследований научно-исследовательского института информационных технологий социальной сферы Кемеровского государственного университета культуры и искусств [Электронный ресурс] / Н. И. Гендина, О. И. Алдохина // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. – Электрон. дан. – М. : ГПНТБ России, 2004. – 1 электрон. опт. диск (CD-ROM). – Систем. требования: IBM PC, Windows 95 и выше. – ISBN 5-85638-091-6. – № гос. регистрации 0320400576. – 1500 экз.