

# Лингвистические средства проектирования контента web-сайтов

## Linguistic Design Tools of the Web-Site Content

### Лінгвістичні засоби проектування контенту web-сайтів

*Н. И. Гендина*

*НИИ информационных технологий социальной сферы  
Кемеровского государственного университета культуры и искусств,  
Кемерово, Россия*

*Nataliya Gendina*

*Research Institute of Information Technologies in Social Sphere,  
Kemerovo State University of Culture and Arts, Kemerovo, Russia*

*Н. І. Гендіна*

*НДІ інформаційних технологій соціальної сфери  
Кемеровського державного університету культури і мистецтв, Кемерово, Росія*

Рассматриваются возможности использования лингвистических средств оптимизации разработки и повышения качества контента web-сайтов. Определяются функции и области применения лингвистических средств в технологиях сайтостроения.

Considered are the possibilities of using linguistic tools for the optimization of design and raising the quality of the web-site content. Described are the functions and applications of linguistic tools in site-building technologies.

Розглядаються можливості використання лінгвістичних засобів для оптимізації розробки і підвищення якості контенту web-сайтів. Визначаються функції й області застосування лінгвістичних засобів у технологіях сайтобудівництва.

#### **Постановка проблемы**

В последние годы наблюдается динамичный рост Интернет-ресурсов, важнейшим видом которых являются web-сайты. Создание web-сайтов стало носить массовый характер и обострило проблему их качества. Однако эмпирический характер создания сайтов, отсутствие теоретического обоснования принимаемых при разработке сайтов решений значительно снижает качество создаваемых электронных ресурсов. Проведенные в НИИ информационных технологий социальной сферы Кемеровского государственного университета культуры и искусств (КемГУКИ) исследования контента сайтов образовательных учреждений (вузов культуры и искусств, федеральных образовательных сайтов), сайтов музеев, библиотек, органов исполнительной власти выявил ряд общих недостатков, значительно снижающих эффективность создаваемых Интернет ресурсов (1-5). Выявленные недостатки свидетельствуют о том, что проблема качества сайтов носит комплексный характер и включает как вопросы программно-технического, организационного, так и семантического характера.

Цель данной статьи – анализ возможностей лингвистических средств для оптимизации разработки и повышения качества контента создаваемых web-сайтов. Следует подчеркнуть новизну подобной постановки вопроса: ни в публикациях, посвященных проектированию, разработке и созданию сайтов, ни в микропотоке документов, посвященных лингвистическому обеспечению, обоснование целесообразности и определение областей применения лингвистических средств при создании сайтов, не содержится.

Выявление возможностей использования лингвистических средств для оптимизации разработки и повышения качества контента создаваемых web-сайтов требует ответа на ряд вопросов: какие положения теории лингвистического обеспечения могут найти применение при разработке контента сайтов и на каких этапах; какие именно лингвистические средства и для чего целесообразно использовать при разработке веб-сайтов; какие функции выполняют лингвистические средства в процессе сайтостроительства; каковы преимущества их использования при разработке контента сайтов.

## **Предлагаемый подход**

Для того, чтобы получить целостную картину использования лингвистических средств в ходе сайтостроения, используем метод аналогий, в соответствии с которым проектирование web-сайта рассматривается как частный случай проектирования автоматизированной информационной системы (АИС). В соответствии с таким подходом проектированию сайта присущи все основные стадии и этапы разработки АИС, закрепленные в ГОСТ 34. 601-90 «ИТ. Комплекс стандартов на АС. Автоматизированные системы. Стадии создания». На основе анализа нормативных и учебных изданий, посвященных проектированию АИС, на кафедре технологии автоматизированной обработки информации КемГУКИ создана обобщенная характеристика и детальный перечень стадий и этапов создания сайта, подробно описанная в (6). Нами было проведено исследование содержания работ на каждом из этих стадий и этапов в целях получения системного представления о месте и роли лингвистических средств при разработке и оптимизации контента сайтов.

Как показали результаты исследования, лингвистические средства могут быть использованы на всех трех – предпроектной, проектной и послепроектной- стадиях проектирования сайтов. Однако применительно к этапам проектирования их применение носит более избирательный характер. Рассмотрим, какие именно лингвистические средства и с какой целью могут применяться в ходе проектирования контента сайта.

## **Результаты анализа использования лингвистических средств на предпроектной стадии**

Предпроектная стадия разработки сайта включает обследование предметной области в рамках которой создается сайт; разработку его концептуальной модели; разработку технического задания. На данном этапе лингвистические средства могут быть использованы как инструмент формализации, позволяющий определить семантические (отраслевые, тематические, предметные, проблемные) и формальные (хронологические, территориальные, языковые, типологические др.) границы контента сайта. Результатом этого этапа является построение логико-понятийной схемы предметной области, для которой создается сайт (например, «Дополнительное образование детей», «Информационно-коммуникационные технологии в образовании» и т. п.) или объекта сайтостроения (этнографический музей, вузовская библиотека, университет культуры и искусств, администрация области, района, города и т. п.). Решение этой задачи наиболее целесообразно вести с помощью различных иерархических классификационных ИПЯ: ББК, УДК, ДКД, ГРНТИ, отраслевых и проблемных классификаторов, рубрикаторов. Именно эта группа ИПЯ позволяет очертить семантические границы разрабатываемого контента сайта, определить важнейшие логические отношения типа «род-вид», «целое часть», установить взаимосвязи и иерархические отношения в пределах создаваемого контента. Наряду с иерархическими классификационными ИПЯ на данном этапе могут найти применение дескрипторные ИПЯ тезаурусного типа, позволяющие устранить синонимию в используемом понятийно-терминологическом аппарате и определить вышестоящие и нижестоящие термины ключевых понятий, раскрывающих создаваемый контент сайтов. Результатом использования лингвистических средств на данном этапе является разработка логико-понятийной схемы предметной области в виде иерархически упорядоченного перечня рубрик, иерархического древа (графа), таблицы и/или дескрипторные статьи ключевых понятий, отражающих контент создаваемого сайта.

Для формализации представления состава категорий потенциальных пользователей разрабатываемого сайта может быть использована такая группа Общероссийских классификаторов, как классификаторы информации о населении и кадрах: Общероссийский классификатор профессий рабочих, должностей служащих и тарифных разрядов; Общероссийский классификатор информации о населении, Общероссийский классификатор информации по социальной защите населения и др.

Одним из важнейших видов работ в ходе предпроектного обследования является определение состава и анализ имеющихся сайтов в заданной предметной области. Анализ контента уже имеющихся сайтов, аналогичных типу и функциональному назначению разрабатываемого сайта, представляет чрезвычайно важную задачу, т. к. позволяет выявить достоинства и недостатки существующих web-ресурсов и на этой основе разработать собственную концептуальную модель контента сайта.

Как показали проведенные нами исследования, анализ имеющихся сайтов в заданной предметной области наиболее рационально проводить, используя разработанную в теории лингвистического обеспечения технологию создания лексико-семантической основы ИПЯ. Как известно, важнейшими этапами создания любого ИПЯ являются отбор лексических единиц; нормализация лексики; систематизация лексики; оформление организационной структуры ИПЯ. Последовательное выполнение этих этапов позволяет осуществить анализ сайтов-аналогов. Методика анализа предусматривает выявление рубрик, отражающих контент обследуемых сайтов; устранение синонимии и многозначности в формулировке рубрик; ранжирование выявленных и нормализованных рубрик по частоте их встречаемости; определение на основе частоты встречаемости «типовой модели» контента анализируемых сайтов.

Важнейшим этапом, завершающим предпроектную стадию разработки контента сайта, является создание информационного образа объекта сайтостроения. Термин «информационный образ объекта сайтостроения» был введен нами с целью формализации разработки контента сайта и преодоления господствующего в современной практике сайтостроения эмпирического подхода. Информационный образ – это максимально полный, упорядоченный с использованием фасетного и/или иерархического принципов, перечень характеристик (атрибутов и аспектов), которые дают целостное представление об объекте сайтостроения и позволяют моделировать контент сайта в зависимости от его общих и специфических функций. Как следует из данного определения, моделирование контента сайта опирается на такие лингвистические средства, как иерархические и/или фасетные классификации, выступающие в роли логического «каркаса», позволяющего обозначить границы контента сайта, определить его объем, содержание и структуру, установить его связи с другими понятиями. Наряду с уточнением границ контента сайта, информационный образ позволяет обеспечивать решение целого ряда других важных задач: определение состава рубрик, отражающих контент сайта; распределение рубрик по уровням вложения; установление взаимосвязей рубрик.

Проведенные исследования показали, что в основу создания информационного образа также может быть положена технология создания лексико-семантической основы ИПЯ, которая в этом случае будет включать:

1. Выявление и отбор лексических единиц, выражающих основные характеристики объекта сайтостроения. Именно от качества выполнения этого этапа во многом зависит качество создаваемого веб-ресурса. Основными источниками отбора лексики при этом являются: уже существующие, т. е. ранее разработанные ИПЯ, располагающие фиксированным словарем: классификационные ИПЯ (УДК, ББК, ДКД, ГРНТИ, классификаторы), дескрипторные ИПЯ (дескрипторные словари, информационно-поисковые тезаурусы); методические рекомендации открытой энциклопедии «Рубрикана», размещенной на web-сайте «Рубикон»; массив нормативных, справочных, учебных, научных документов по тематике создаваемого сайта. Принципиально важно подчеркнуть, что все эти источники отбора лексики должны использоваться одновременно, в комплексе.

2. Нормализация выявленных лексических единиц реализуется посредством двух операций: 1) представлением лексических единиц в единообразной грамматической форме (морфологический уровень; 2) устранением синонимии и полисемии (семантический уровень).

3. Систематизация лексических единиц заключается в их группировке, упорядочении и установлении между ними смысловых связей (парадигматических отношений).

4. Интеграция лексических единиц в информационный образ объекта сайтостроения соответствует по своей сути этапу, именуемому в технологии создания лексико-семантической основы ИПЯ как организационное оформление лексики ИПЯ. Оно реализуется в графическом, знаковом воплощении информационного образа контента сайта, которое зависит от типа используемых лингвистических средств: нормализованные и систематизированные лексические единицы могут быть представлены в информационном образе либо как комплекс предметных рубрик, либо дескрипторных статей: либо фрагмента иерархической классификации, иерархического дерева, таблицы, матрицы, фасетов и т. п.

Предлагаемая технология разработки информационного образа может быть использована при разработке контента конкретных сайтов; оптимизации контента ранее созданных сайтов; подготовке и переподготовке кадров информационно-библиотечного профиля.

## **Результаты анализа использования лингвистических средств на проектной стадии**

Наибольшее применение лингвистические средства находят на этапе логического проектирования. В данном случае используются уже не столько конкретные ИПЯ, сколько методы формализованного анализа текста, разработанные в теории индексирования как средства снижения субъективности действий индексаторов. В частности, как показали исследования, весьма эффективным является аспектный метод. Его целесообразно применять для разработки аспектных структур наполнения каждой рубрики, характеризующей контент сайта. Так, например, при разработке контента сайта музея важнейшими рубриками являются такие, как «Направления деятельности музея», «Фонды», «Экспозиции», «Выставки», «Ресурсы музея», «Филиалы музея», «История музея» и др. Каждая из этих рубрик должна, в свою очередь, располагать четким перечнем аспектов, раскрывающих ее содержание. Например: «Направления деятельности музея» раскрываются через такие аспекты, как фондовая, экспозиционная, научно-просветительная, рекламно-издательская деятельность; а рубрика «История музея» предполагает включение таких аспектов, как основные этапы истории, основатели музея, деятели общенационального значения, работавшие в музее и др. Результатом использования аспектного метода является разработка перечня аспектов рассмотрения материала в составе каждой рубрики и подрубрики, характеризующих контент сайта. Аспектный метод обеспечивает единообразие и сопоставимость данных, представленных на сайте.

## **Результаты анализа использования лингвистических средств на послепроектной стадии**

При подготовке опытного образца сайта для итоговой системной композиции рубрик сайта используются, как правило, принципы фасетных и/или иерархических классификационных систем, позволяющие на главной странице сайта дать целостное представление о содержании сайта и показать состав и схему взаимосвязей рубрик сайта (уровни вложения). Следует подчеркнуть, что на практике несоблюдение принципа иерархии при выделении рубрик на главной странице сайта является весьма распространенной погрешностью, существенно затрудняющей пользователям ориентацию в содержании сетевого ресурса.

Вторым важным направлением применения лингвистических средств на данной стадии является представление в рамках разработанного информационного образа исходных текстовых, графических, фото и др. материалов для сайта. Как и на проектной стадии, в данном случае актуальным является использование методов свертывания и трансформации исходных текстов в соответствии с требованиями сетевого электронного представления информации. Строго говоря, в классической теории лингвистического обеспечения библиотечно-информационной технологии разработка этого направления пока отсутствует, поскольку технология подготовки «сетевых текстов» получила свое развитие весьма недавно как результат преодоления таких распространенных недостатков, как избыточность, разнородность, неструктурированность, затрудненность восприятия информации на сайте. Суть проблемы состоит в том, что имеющиеся традиционные тексты чаще всего оказываются не пригодными для размещения на сайте и требуют обязательной предварительной переработки. В результате анализа специальной литературы нами были выявлены основные требования, которым должен удовлетворять текст, размещенный в сети Интернет.

Несмотря на то, что большинство правил одинаково применимы, как для печатных публикаций, так и для сетевых, при подготовке текстов для размещения в сети нужно учитывать и ряд специфических особенностей, обусловленных наличием такой нетривиальной единицы измерения текста, как «один экран»: краткость; информативность; простота и лаконичность предложений; структурированность и порционность текста; нелинейность. Эти требования также во многом определяются особенностью сканирующего» характера чтения текстов, размещенных в глобальной сети, предполагающем лишь беглый просмотр заголовков, а также большие временные затраты, чем на чтение печатных текстов.

Для решения проблемы неоднородности представления информации об однотипных объектах, характеризующихся в рамках одного и того же сайта, рационально использовать уже рассмотренный выше аспектный метод. Так, например, для раскрытия содержания часто встречающейся в контенте сайтов учреждений культуры рубрики «Проекты», целесообразно использовать перечень

таких аспектов, как наименование проекта, хронологический период, руководителя, участников проекта; цель проекта, краткое содержание проекта; полученные или ожидаемые результаты.

Использование лингвистических средств происходит и на таких этапах послепроектной стадии, как эксплуатация сайта и модернизация сайта. Эксплуатация сайта неразрывно связана с ведением и актуализацией используемого лингвистического обеспечения сайта. Прежде всего, речь идет о сохранении и поддержании в рабочем состоянии семантической модели контента сайта, целостность которой задается разработанным на предпроектной стадии информационным образом. Несогласованность, произвольность включения новых рубрик на главной странице сайта без связи их с созданным ранее информационным образом, ведет к нарушению системности представления данных и затруднительности ориентации для пользователя. Модернизация сайта также предполагает разработку информационного образа, базирующегося на комплексном использовании лингвистических средств.

## **Выводы**

1. Лингвистические средства могут быть использованы на всех стадиях и на основных этапах проектирования и создания контента сайтов как важнейшие инструменты формализации представления информации и структурирования данных с целью облегчения доступа пользователей к создаваемым веб-ресурсам..

2. Наиболее «лингвистически емким» является предпроектная стадия проектирования сайта, предполагающая создание информационного образа объекта сайтостроения. Использование комплекса лингвистических средств при разработке информационного образа позволяет обеспечить полноту отражения всех характеристик, присущих конкретному объекту сайтостроения; дает возможность корректировки (сужения либо расширения) состава используемых характеристик при описании конкретного объекта в соответствии с целями, задачами, особенностями проектируемого контента сайта, а также потребностями заказчика.

3. Основными функциями лингвистических средств при разработке контента сайтов являются обеспечение доступа потребителей информации к информационным web-ресурсам и обеспечение качества создаваемого контента. Реализация этих функций может обеспечиваться за счет применения иерархических классификационных и/ или фасетных ИПЯ, позволяющих четко структурировать информационный ресурс; реализации возможностей дескрипторных ИПЯ и тезаурусного способа представления данных для обеспечения нормализации и достижения однозначности используемого в рамках конкретного контента сайта понятийно-терминологического аппарата; использования технологии создания лексико-семантической основы ИПЯ как основы создания информационного образа; привлечения аспектного метода представления информации с целью достижения однородности и сопоставимости представляемых на сайте сведений; применения методов свертывания информации, обеспечивающих достижение лаконичности, порционности, структурированности создаваемых сетевых текстов для облегчения восприятия информации.

4. Лингвистическое обеспечение сайтостроения является весьма перспективным научно-технологическим направлением, требующим дальнейшего изучения и развития, поскольку позволяет обеспечить реализацию принципов технологичности, системности и целостности в условиях массового создания сайтов. Лингвистические средства, рассматриваемые как инструмент формализации моделирования контента сайта позволяют обеспечить принятие обоснованных решений при определении состава рубрик, отражающих смысловое наполнение создаваемого веб-ресурса; позволяют перейти к аргументированному принятию решений по структуре и наполнению контента конкретного сайта; открывает возможности для снижения субъективизма разработчиков сайтов; уменьшения интеллектуальных, временных и стоимостных затрат на разработку сайтов.

5. Развитие лингвистического обеспечения сайтостроения предполагает решение ряда проблем: включая необходимость создания узко-отраслевых, проблемных классификаторов и рубрикаторов, поскольку прямое использование таких иерархических классификационных систем, как ББК, УДК, ГРНТИ, ДКД и т. п. в качестве средств структурирования информации для отдельных типов сайтов не целесообразно в силу того, что все эти ИПЯ были созданы для решения иных информационных

задач. Самостоятельную проблему представляет преодоление господствующего эмпирического подхода при разработке контента сайтов, при котором не учитываются возможности использования лингвистических средств.

### Литература

1. Гендина, Н. И. Оценка качества сайтов: методика и результаты пилотажного исследования [Электронный ресурс] / Гендина Н. И., Колкова Н. И., Алдохина О. И. // 7-ая Международная Конференция «EVA 2004 Москва «Информация для всех: культура и технологии информационного общества»: материалы конф. – Электрон. дан. – М.: Центр ПИК, 2004. – 1 электрон. опт. диск (CD-ROM).
2. Колкова, Н. И. Оптимизация проектирования контента сайтов учреждений культуры и искусства: результаты исследований НИИ информационных технологий социальной сферы Кемеровского государственного университета культуры и искусств [Текст] / Н. И. Колкова, О. И. Алдохина, И. Л. Скипор // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы 11-й Международной конференции «Крым 2004» – Электрон. дан. – М.: ГПНТБ России, 2004. – 1 электрон. опт. диск (CD-ROM).
3. Гендина, Н. И. Моделирование контента сайта органа исполнительной власти [Текст] / Н. И. Гендина, Н. И. Колкова, О. И. Алдохина, С. В. Сорокопуд // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы 12-й Международной конференции «Крым 2005» – Электрон. дан. – М.: ГПНТБ России, 2005. – 1 электрон. опт. диск (CD-ROM).
4. Гендина, Н. И. Создание исследовательского и методического инструментария разработки сайтов библиотек и музеев: результаты разработок НИИ информационных технологий социальной сферы Кемеровского государственного университета культуры и искусств [Текст] / Н. И. Гендина, Н. И. Колкова, О. И. Алдохина, И. Л. Скипор // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы 12-й Международной конференции «Крым 2005» – Электрон. дан. – М.: ГПНТБ России, 2005. – 1 электрон. опт. диск (CD-ROM).
5. Гендина, Н. И. Информационный образ как основа разработки контента сайтов учреждений культуры [Электронный ресурс] / Н. И. Гендина, Н. И. Колкова, О. И. Алдохина, И. Л. Скипор // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы 13-й Международной конференции «Крым 2006» – Электрон. дан. – М.: ГПНТБ России, 2006. – 1 электрон. опт. диск (CD-ROM).
6. Колкова, Н. И. Прикладная информатика: технология дипломного и курсового проектирования [Текст]: учебное пособие специальностей «Прикладная информатика» (в информационной сфере), «Прикладная информатика» (в социальной сфере), «Прикладная информатика» (в социально-культурной сфере) / Н. И. Колкова, И. Л. Скипор. – Кемерово: КемГУКИ, 2007. – 434 с.