

УДК [005.2:004] (076.5)

**Методология создания текстовых документов  
на основе лексикологического синтеза**

*Б.В. Черников*

*Национальный исследовательский университет*

*«Высшая школа экономики»*

*Профессор кафедры «Архитектура программных систем»*

*e-mail: bor-cher@yandex.ru*

Современная деятельность организации любого профиля неразрывно связана с реализацией управленческих решений, процесс выработки которых опирается на информационное обеспечение. В связи с этим чрезвычайно важен обмен своевременной и точной информацией, который чаще всего выполняется с помощью предварительно подготовленных документов, сопровождающих производственные процессы.

Функции управления организацией выражаются посредством соответствующих документов. В условиях неуклонного усложнения деятельности особое внимание обращает на себя возрастание требований полноты и своевременности, предъявляемых к управленческой информации. Офисные функции, определяющие подготовку необходимых документов для выработки управленческого решения, реализуются в организациях в рамках документационного обеспечения управления, включающего в себя процессы документирования и организации работы с документами, причем если задачи второго процесса автоматизированы специализированными системами, то вопросы автоматизации создания, оформления и изготовления документов развиты недостаточно и требуют особого рассмотрения. В структуру процесса подготовки документов входят операции составления документа. Характерно, что при автоматизации процесса подготовки документов руководители, как правило, не чувствуют преимуществ внедрения новых технологий. Особую значимость этот вопрос приобретает в случаях создания документов непосредственно его авторами, что чаще всего имеет место. Пока же в организациях и на предприятиях многие документы, в основном, создаются с использованием возможностей текстовых процессоров.

Передача документов по каналам связи и их хранение сопряжены со значительностью объема и недостаточной защищенностью. Выход из создавшегося положения может быть найден только в пересмотре традиционных процедур подготовки документов. На основе результатов проведенного анализа выявлено, что значительное число документов, сопровождающих производственные процессы, можно отнести к категории слабоформализуемых.

Учитывая значимость роли процесса документирования в совершенствовании документационного обеспечения управления деятельностью организаций, разработана методология информатизации документационного обеспечения и подготовки слабоформализуемых документов, циркулирующих в системе управления и непосредственно относящихся к проведению основной деятельности организаций и предприятий, а также прикладных аспектов ее реализации.

На современном этапе процесс документирования информации, поддерживающей производственные процессы организаций, должен отвечать следующим требованиям:

- максимально возможная формализация;
- минимальное время создания документов;
- автоматизация процесса формирования конкретного экземпляра документа в условиях слабой формализации.

Данные требования соответствуют как отечественным нормативным документам, так и положениям европейского стандарта MoReq2. Кроме указанных требований, MoReq2 большое внимание уделяет простоте пользования и производительности, что гармонирует с требованиями, определяемыми стандартом открытых систем.

Принцип лексикологического синтеза предполагает проведение предварительного анализа совокупности документов определенного вида, на основе которого выделяется устойчивый набор формулировок.

Каждой формулировке документа ставится в соответствие опорное слово, выбор которого однозначно определяет наличие конкретной формулировки в документе. Если фрагмент текста документа содержит значительное

количество строк и всегда присутствует в документе в строго определенной последовательности построения предложений, то данный фрагмент текста может определяться одним опорным словом.

Формирование лексикологической схемы и лексикологического дерева проводится на основе анализа связи опорных слов, образуя маршрут (траекторию) их выбора при формировании документа. Лексикологическая схема позволяет установить взаимосвязи опорных слов, учитывая, что в различных случаях может наблюдаться существенное изменение маршрута формирования документов, что и определяет возможную вариативность отдельно взятых экземпляров.

На этапе автоматизированного формирования документов (период эксплуатации систем) используется база данных комплекса формируемых документов, содержащая реквизиты, формы, лексикологические деревья и комплекты опорных слов документов. При формировании конкретного экземпляра документа после выбора первого опорного слова в документ внедряется фрагмент, соответствующий этому опорному слову. Процесс выбора опорных слов повторяется по всему маршруту формирования документа с учетом взаимосвязей опорных слов. По завершении процесса выбора опорных слов в унифицированную форму вставляются все положенные реквизиты и оформляется конкретный экземпляр документа.

Полный перечень опорных слов с учетом их взаимосвязей образует лексикологическое дерево документа, «прохождение» по ветвям которого обеспечит выбор формулировок, используемых в документе.

Сформирована модель разработки автоматизированной технологии создания слабоформализуемых документов на основе лексикологического синтеза. В соответствии с решаемыми на каждом этапе задачами в структуре модели предусмотрены четыре стадии разработки: аналитическая, унификационная, информационная и процедурная. На аналитической стадии выделяется комплекс документов организации, для которых будет разрабатываться технология автоматизированного формирования документов. Процесс приведе-

ния документов к единообразию по форме и содержанию реализуется на унификационной стадии, значение которой велико, поскольку унификация обеспечивает возможность совместного и многократного использования конструкций различных документов и интероперабельность, т.е. независимость от технической и программной платформы их создания и обработки, что создает предпосылки к экономии времени и материальных затрат. Информационная стадия разработки технологии лексикологического синтеза слабоформализуемых документов предназначена для построения своеобразной инфологической модели данных. На процедурной стадии проводится подготовка программных компонентов, предназначенных для непосредственной реализации процесса создания документов и интеграции системы формирования с информационной системой предприятия.

Традиционный подход, который применяется в настоящее время при использовании типизации и трафаретизации документов, предлагает деление информации на постоянную и переменную. Однако при разработке процесса автоматизированного формирования документов на основе лексикологического синтеза целесообразно использовать более глубокую классификацию информации, а именно:

- унифицированная постоянная информация, подготовленная заранее и хранимая в базе данных или содержащаяся в тексте программы, которая автоматически внедряется в формируемый документ программными средствами;
- унифицированная переменная информация, содержащая стандартизированные и формализованные данные, хранимая в базе данных и вводимая при формировании документа путем выбора требуемых формулировок;
- переменная вводимая информация, подчиненная определенным требованиям по способу представления данных и содержащая конкретизирующие сведения, как правило, для конкретного экземпляра документа и вводимая с клавиатуры непосредственно при подготовке документа;
- неунифицированная информация, содержащая свободные формулировки и вводимая при необходимости прямым набором текста с клавиатуры.

Лексикологический синтез слабоформализуемых документов позволяет существенно сократить нагрузку на системы хранения информации. При составлении документа формируется индексная последовательность, подвергаемая дополнительному сжатию для создания сохраняемого информационного пакета, а при воссоздании документа производится расшифровка этого пакета с использованием организованного цикла восстановления фрагментов на основе согласованной базы данных лексикологического дерева. Анализ эффективности предлагаемого метода обработки слабоформализуемых документов при организации их хранения показывает возможность значительного (в десятки и даже сотни раз) сокращения объемов хранимой информации по сравнению с прямым сохранением или сжатием информации.

Внедрение автоматизированного способа лексикологического синтеза в практику формирования слабоформализуемых документов обеспечивает смещение групповых центров аккумуляции документарных компонентов из областей неунифицированной информации в сторону унифицированных элементов, что предопределяет возможность автоматизированного выбора унифицированных формулировок в автоматизированной схеме. Обращает на себя внимание значительность роста унифицированной информации по сравнению с традиционной схемой формирования документов (в 5 раз в организационно-распорядительных документах, в 8 раз в производственно-технологической документации и документации высших учебных заведений, в 9 раз в документации лечебных учреждений) и значительное снижение объемов переменной вводимой и неунифицированной информации.

Анализ трудоемкости при изменении условий формирования документов демонстрирует в среднем более чем двукратное снижение трудозатрат при использовании предлагаемых технологий в организациях различного профиля деятельности.

В целом разработка и исследование возможностей лексикологического синтеза текстовых документов показали высокую эффективность и перспективность этой технологии создания документов.