УДК 004.91:005.92

АНАЛИЗ ИНФОРМАЦИИ ТЕКСТОВЫХ ДОКУМЕНТОВ ПРЕДПРИЯТИЙ МИКРОЭЛЕКТРОННОЙ ПРОМЫШЛЕННОСТИ

^{1,2}Черников Б.В., ³Кремер Е.А.¹ООО «Газпром ВНИИГАЗ», Москва, e-mail: bor-cher@yandex.ru;²Российский экономический университет им. Г.В. Плеханова, Москва;³Национальный исследовательский университет «МИЭТ», Москва, e-mail: kremerea@gmail.com

Управленческая и конструкторская деятельность в микроэлектронной промышленности осуществляется с помощью документов. Документооборот в большинстве случаев осуществляется в бумажном виде. Однако в связи с ростом числа и объема накопленных документов возникает проблема хранения и подготовки документов. В связи с этим особый интерес начинают представлять автоматизированные процедуры формирования документов, которые позволяют не только сократить время на создание документа, но и значительно уменьшить его объем хранения в электронном виде. На предприятиях микроэлектронной промышленности используются различные системы документации. Стандартизация осуществляется за счет нормативных документов. Вследствие разнородности документов, используемых на предприятии, требуется их анализ с целью выявления их информационного состава, определения соотношения постоянной и переменной информации в них, а также последующей их обработки. Помимо этого необходимость анализа информационного состава документации обуславливается возможностью применения различных алгоритмов для решения проблемы хранения и подготовки. Информационный анализ позволил определить процентное соотношение постоянной и переменной информации для каждого документа в отдельности, а также сделать обобщение на все типы документов. Данный анализ позволяет сделать вывод, что многие документы относятся к слабоформализуемым. В связи с этим для решения вышеозначенной проблемы целесообразно использовать лексикологический синтез. Данный факт позволяет в дальнейшем при использовании лексикологического синтеза попробовать применить новые способы формирования индексной последовательности с целью еще большего уменьшения объема готового документа.

Ключевые слова: документ, постоянная информация, переменная информация, лексикологический синтез, слабоформализуемый документ, хранение

ANALYSIS OF TEXT DOCUMENTS INFORMATION AT ENTERPRISES OF MICROELECTRONIC INDUSTRY

^{1,2}Chernikov B.V., ³Kremer E.A.¹LLC «Gazprom VNIIGAZ», Moscow, e-mail: bor-cher@yandex.ru;²Plekhanov Russian University of Economics, Moscow;³Moscow Institute of Electronic Technology, Moscow, e-mail: kremerea@gmail.com

Management and design actions in the microelectronic industry are carried out with the help of documents. Document management is performed in paper form in most cases. However, due to the growth of the number and volume of accumulated documents, the problem of storing and preparation of documents arises. In this case, the automated procedures of documents' formation begin to be a subject of special interest. The most interesting aspects of this problem are procedures, which allow not only to reduce time for creation of the document, but also to considerably reduce its volume of storage in electronic form. Various documentation systems are used at the enterprises of microelectronic industry. Standardization is carried out through regulatory documents. In case of heterogeneity of the documents used at the enterprise their analysis is required in order to identify their information composition, determine the ratio of constant and variable information in them, as well as their subsequent processing. Besides importance of the information content of documentation analysis is conditioned by the possibility of applying various algorithms to solve the storage and preparation problem. Information analysis allowed to determine the percentage of constant and variable information for each document separately, and also to make a generalization to all types of documents. This analysis allows concluding that many documents are slightly formalized. The implementation of lexicological synthesis can provide a way to solve this problem. This fact allows further using lexicological synthesis to try to apply new ways of forming an index sequence with the aim of further reducing the volume of the finished document.

Keywords: document, permanent information, variable information, lexicological synthesis, slightly formalized document, storage

В настоящее время документооборот на предприятиях в большинстве случаев осуществляется в бумажном виде, несмотря на широкое распространение персональных компьютеров и применение на предприятиях систем электронного документооборота. Данный факт обуславливает необходимость наличия архива бумажных документов значительных объемов.

Преобладание бумажных документов над электронными копиями определяется рядом причин, к которым можно отнести следующие:

- требования документов, регламентирующих документооборот внутри предприятия (помимо внутренних стандартов предприятия сюда входят государственные и ведомственные стандарты);

- недостаточное распространение электронных средств защиты текстовых документов, таких как электронная подпись;

- требования предоставления документов регулирующим органам, вышестоящим организациям и партнерам в бумажном виде;

- высокая юридическая значимость.

Создание, редактирование и подготовка документов осуществляется при помощи различных программ, среди которых преобладают Microsoft Word, OpenOffice Writer и LibreOffice Writer. Данные программы при всех преимуществах имеют ряд недостатков:

- возможность появления ошибок вследствие человеческого фактора при наборе документа с клавиатуры;

- недостаточная полнота документов;

- большой объем работ, который необходимо выполнить при создании и подготовке документа.

Персональные компьютеры часто используются на предприятиях лишь в качестве печатных машинок, а локальная сеть предприятия – в роли файлообменника. Следствием является недостаточная эффективность использования персональных компьютеров внутри предприятий.

При подготовке документов высокая скорость набора символов не обуславливает быструю готовность документа. Значительная часть времени уходит на поиск нужной информации, компоновку, многократную пересылку документов между исполнителями, согласование различных аспектов, а также утверждение различными отделами, которые имеют отношение к документу. Исправить создавшееся положение можно, пересмотрев традиционные процедуры подготовки документов. Процесс документирования информации должен отвечать следующим требованиям:

- максимальная формализация;

- минимальное время создания документов;

- автоматизация формирования конкретного документа при слабой его формализации.

Данные требования соответствуют российским и европейским стандартам [1, 2]. Большое значение имеет простота использования и производительность, что отвечает требованиям открытых стандартов.

Актуальность и цель исследования

Во множестве публикаций рассматривались вопросы информационного и документационного обеспечения на предприятиях. Зачастую рассматриваются вопросы, связанные с электронным документооборотом. Целью данных публикаций являлось повы-

шение эффективности обработки, циркулирования и обмена электронными документами внутри предприятия, а также создание эффективных систем поиска по текстовым документам. Также рассматриваются проблемы поиска новых, более эффективных способов сжатия объема электронных документов при хранении. В работе [3] проводится поиск оптимальных методов и алгоритмов сжатия для различных типов текстовых файлов. В работах [4, 5] затрагивается вопрос повышения безопасности информационных систем.

В связи с текущим состоянием документационного обеспечения на предприятиях микроэлектронной промышленности необходимы исследования в области технологий автоматизированного формирования документов. Актуальность исследований в этом направлении обусловлена необходимостью уменьшения объемов электронных документов при их хранении, а также реализацией государственной программы «Информационное общество» [6], которая направлена на ускорение внедрения информационных технологий во все сферы деятельности и решения следующих задач:

- повышение эффективности государственного управления;

- повышение информационной открытости власти;

- развитие информационных технологий;

- развитие информационных систем (порталов);

- увеличение числа квалифицированных специалистов;

- максимальное использование интеллектуального потенциала.

Важной проблемой на современных предприятиях микроэлектронной промышленности является необходимость увеличения объемов систем хранения данных (СХД), поскольку количество создаваемых документов постоянно растет, а потребность в хранении более старых документов никуда не пропадает. Бесконечно увеличивать объемы СХД невыгодно как с экономической, так и с технологической точки зрения, поскольку это влечет за собой траты на техническое обеспечение, на увеличение площадей, занимаемых СХД, а также на обслуживание данного парка техники. Поэтому целесообразно уменьшать объем самих документов. Для уменьшения объема документов применяют сжатие данных. Сжатие данных – это преобразование, выполняемое в пределах одного файла по определенному алгоритму. В основе сжатия данных лежит принцип устранения избыточности, которое выполняется посредством замены длинных последовательностей более мелкими.

Недостатком сжатия файлов является необходимость наличия уже готового файла к началу выполнения процедуры сжатия. При использовании такого метода нельзя сжимать файл в процессе его создания, а значит, есть возможность перехвата текстового документа злоумышленником, когда файл еще не защищен.

Для сокращения объема документов, которые создаются в различных предприятиях, можно применять лексикологический синтез. Особенно эффективен он при создании слабоформализуемых документов. Слабоформализуемые документы – полнотекстовые, табличные или смешанные документы, содержание которых существенным образом связано с произвольной, меняющейся в каждой конкретной ситуации структурой [7].

Так как внедрение лексикологического синтеза связано с использованием информации, содержащейся в документах микроэлектронной промышленности, то необходим информационный анализ данных документов.

Целью данной статьи является анализ информации, содержащейся в документах предприятия микроэлектронной промышленности.

Состав документации предприятия микроэлектронной промышленности

Автоматизация документооборота предприятия заключается в комплексном подходе к решению задач разработки, согласования, распространения, поиска и архивации документов. Основным объектом в этом процессе является документ [8].

Процесс документооборота рассматривается как движение документов с момента их создания или получения до завершения исполнения, отправки адресату или передачи в архив [9]. Структура документационного обеспечения управления (ДОУ) отражена на рис. 1.

На предприятиях микроэлектронной промышленности, как правило, используют различные системы документации (таблица). Для стандартизации применяют систему нормативных документов, к которым относятся государственные стандарты, отраслевые стандарты, стандарты предприятий.

Состав унифицированной системы документации

Вид документации	Документы
Организационно-распорядительная	Кадровая документация Приказы Штатное расписание
Организационно-правовая	Устав организации Должностные инструкции Регламенты
Информационно-справочная	Служебные записки Справки Акты
Плановая	Финансовые планы Бизнес-планы Производственные планы Планы развития
Учетная	Накладные Табели Книги учета
Отчетная	Отчеты подразделений Сводные отчеты

Рис. 1. Структура документационного обеспечения управления предприятием

Ряд документов можно создавать с помощью программных продуктов, сопровождающих создание конструкторских образцов. К такой документации относятся документы, связанные со строго унифицированными компонентами. Значительная часть документов относится к технологической документации (Единая система технологической документации и ГОСТы серии 3). Такие документы формируются при организации и поддержке производственных процессов, причем программные продукты не предусматривают возможности создания эксплуатационной, ремонтной, а часто и непосредственно технологической документации. Поэтому, исследуя возможности уменьшения размеров электронных документов во время хранения, особое внимание следует уделить технологическим, эксплуатационным и ремонтным документам.

*Информационный состав документации.
Категорирование типов информации*

Объем и содержание информации в документе могут изменяться. Наличие или отсутствие информации в документе зависит от различных факторов. Стандартных совпадающих ситуаций практически не существует.

Для исследования возможности применения тех или иных алгоритмов уменьшения размера документов необходимо проанализировать структуру содержащейся в них

информации. Вся информация традиционно делится на постоянную и переменную. Постоянная информация – неизменная информация, которая используется в течение длительного периода времени без каких-либо изменений. Данный тип информации может содержаться, например, в справочных или нормативных данных, а также во всех шаблонных заготовках документов. Переменная информация отражает фактические количественные и качественные характеристики деятельности предприятия, которые необходимо закрепить в документе. К переменной информации можно отнести всю информацию, которую необходимо вводить в документ при каждом его заполнении. Комплект технической документации на предприятии включает в себя следующие группы [10]:

- конструкторские документы (Единая система конструкторской документации (ЕСКД) и нормативная документация (ГОСТы серии 2);
- технологические документы (Единая система технологической документации (ЕСТД), ГОСТы серии 3);
- программные документы (на автоматизированные системы, регламентируются ГОСТами серии 34; на программные продукты – ГОСТами серии 19);
- эксплуатационные документы;
- ремонтные.

Рис. 2. Состав информации по документам: а) акт о готовности оборудования; б) акт о браке продукции; в) акт о передаче дел увольняемого руководителя; г) акт передачи оборудования в ремонт; д) акт предписаний; е) усредненное распределение информации по документам

Помимо технической документации на предприятии существуют также организационно-распорядительные и информационно-справочные документы. Первые можно отнести к наиболее стандартизированным, поскольку они значительно унифицированы по форме и содержанию. Информационно-справочные документы включают в себя различные справки, сводки, отчеты и протоколы.

Содержание постоянной и переменной информации различных документов предприятия показано на рис. 2.

При уменьшении объема документов существенный выигрыш может дать наличие большого объема постоянной информации, так как это позволит использовать возможность замены достаточно большой последовательности маленькой последовательностью при хранении однотипных документов. Однако, исходя из проведенного анализа видно, что объем постоянной информации по сравнению с переменной невелик, что не позволяет использовать преимущество постоянной информации. Очевидно, что все рассмотренные документы можно отнести к слабоформализуемым, поскольку доля постоянной информации в них мала.

*Лексикологический синтез
как способ уменьшения объема
слабоформализуемых документов*

Многие документы, которые используются на предприятиях электронной промышленности, относятся к слабоформализуемым. Их содержание меняется в зависимости от конкретной ситуации. Наполнением данных документов могут быть текст, таблицы, а также смешанные варианты. Данные документы отличаются высокой вариативностью. Поэтому их структуризация требует разбиения этих документов на фрагменты вплоть до отдельных слов, чтобы иметь возможность рассматривать все варианты отдельных частей документа, и, следовательно, иметь представление о вариативности той или иной части документа и адекватно формировать соответствующие фрагменты.

Для сокращения объема слабоформализуемых документов для повышения эффективности сжатия данных целесообразно использовать возможности лексикологического синтеза, что, однако, требует пересмотра не только способов создания документов, но и изменения классификации типов используемой информации [11]. Лексикологический синтез документов позволяет существенно сократить объем слабоформализуемых документов, что достигается благодаря формированию индексной последовательности на основе лексикологического дерева с последующим ее сохра-

нением с целью восстановления документа при необходимости. Итоговый документ в сохраняемом варианте представляет собой индексную последовательность, что значительно сокращает его объем по сравнению с традиционными представлениями электронных текстовых документов. Измененный алгоритм формирования индексной последовательности, используемый во время синтеза, позволит еще больше сократить объем сохраняемых и передаваемых по сетям связи текстовых документов. Для этого необходимо создать эффективный алгоритм формирования индексной последовательности, разработать методологию автоматизации процесса ее формирования и единую технологию обработки документов на основе нового алгоритма формирования индексной последовательности.

Заключение

1. Исследован информационный состав документов микроэлектронной промышленности. Данный анализ позволил отнести многие документы, применяемые на предприятиях микроэлектронной промышленности, к разряду слабоформализуемых.

2. Отнесение документов к разряду слабоформализуемых позволяет применять лексикологический синтез с целью ускорения процесса формирования документов, а также уменьшения объема их хранения.

Список литературы

1. Model Requirements for the Management of Electronic Records (MoReq 2). – France: European Communities. – 2008. – P. 212.
2. Modular Requirements for Records Systems (MoReq 2010). – CEEA-CEE-CEEA, Bruxelles-Luxembourg. – 2011. – P. 520.
3. Петрянин Д.Л., Юрков Н.А. Сжатие текстовых данных // Труды Международного симпозиума «Надежность и качество». – Пенза: Пензенский государственный университет. – 2014. – Т. 1. – С. 328–330.
4. Шаньгин В. Защита информации в компьютерных системах и сетях. – М.: Литрес, 2017. – 593 с.
5. Андрианов В. и др. Обеспечение информационной безопасности бизнеса. – М.: Литрес, 2017. – 450 с.
6. Государственная программа Российской Федерации «Информационное общество (2011–2020 годы)» (утверждена постановлением правительства Российской Федерации от 15.04.2014 № 313) // Собрание законодательства РФ, 05.05.2014, № 18 (2 ч.), ст. 2159.
7. Черников Б.В., Кремер Е.А. Проблемы хранения и защиты слабоформализуемых документов на предприятиях микроэлектронной промышленности // Информатизация и связь. – 2017. – № 4. – С. 165–171.
8. Федеральный закон от 27 июля 2006 года № 149-ФЗ «Об информации, информационных технологиях и о защите информации» // Собрание законодательства Российской Федерации, 31.07.2006, № 31 (ч. 1), ст. 3448.
9. Черников Б.В. Информационный анализ документации промышленных предприятий // Вестник машиностроения. – 2013. – № 3. – С. 74–78.
10. Глаголев В.А. Проблемы стандартизации и повышения качества разработки технической документации на современную промышленную продукцию приборостроения // Менеджмент качества. – 2008. – № 4. – Ч. 1. – С. 68–78. – 2009. – № 1. – Ч. 2. – С. 57–65.
11. Черников Б.В. Способ автоматизированного лексикологического синтеза документов. – Патент РФ № 2253893, 2005.