

УДК 681.3

А.В. ПЕТРИЧЕНКО¹, Е.А. МОСПАН²¹Харьковский национальный университет радиоэлектроники, Украина²Украинско-немецкая компания «ПрофИТсофт», Украина

РАЗРАБОТКА МОДЕЛИ ПРЕДСТАВЛЕНИЯ СТРУКТУР ЭЛЕКТРОННЫХ ДОКУМЕНТОВ С АБСОЛЮТНЫМ ПОЗИЦИОНИРОВАНИЕМ ЭЛЕМЕНТОВ

В статье описаны результаты исследования моделей представления структур электронных документов, форматом которых являются процедурные языки разметки. Такие модели обеспечивают возможность использования процедурных языков разметки в качестве форматов выходных документов в технологиях формирования электронных документов в информационных системах. На основании выявленных недостатков существующих моделей разработана модель представления структур электронных документов с абсолютным позиционированием элементов, которая не зависит от типа процедурного языка разметки. Такая модель обеспечивает возможность выдачи выходного документа в формате произвольного процедурного языка разметки на основании модифицированного документа-шаблона в технологиях формирования электронных документов.

Ключевые слова: электронный документ, документ-шаблон, технология, модель структуры электронного документа, процедурный язык разметки.

Введение

Реализация функциональных задач информационных систем непосредственно связана с разработкой пользовательских интерфейсов по взаимодействию с данными, которые к ним относятся. Одним из способов представления данных функциональных задач являются электронные документы (выходные документы). В WEB-ориентированных информационных системах в качестве форматов таких документов используются, как правило, процедурные языки разметки [1]. Это прежде всего связано с тем, что такие языки разметки гарантируют одинаковое позиционирование элементов как на экране монитора, так и при печати, обладают средствами по шифрованию и архивированию документов. Среди наиболее распространенных процедурных языков разметки следует выделить следующие: PDF, PostScript, LaTeX и другие [2 – 3]. В работе [4] рассмотрена технология формирования электронных документов, которая предполагает создание выходных документов на основании шаблонов, форматом которых являются описательные языки разметки. В качестве форматов формируемых документов в указанной технологии используются как описательные, так и процедурные языки разметки. Для формирования выходного документа в технологии структура документа-шаблона изменяется на основании данных, соответствующих актуальному состоянию функциональной задачи [4]. Затем осуществляется пре-

образование измененной структуры документа-шаблона к формату выходного документа.

В процедурных языках разметки структура электронного документа представляется двумя способами: формальным и в виде объектной иерархической модели. Формальный способ записи структуры электронного документа представляет собой файл, формат которого определяется спецификацией процедурного языка разметки. Такой файл может быть открыт в соответствующем текстовом процессоре, благодаря чему структура электронного документа будет визуализирована в понятном пользователю виде. Объектная иерархическая модель представляет собой последовательность специализированных объектов. Среди них следует выделить два основных класса: объекты, представляющие собой метаинформацию о документе, и команды. К первому классу объектов следует относить: описание состава и размера страниц; используемые графические элементы (рисунки) и шрифты; метаинформацию о методах шифрования и архивирования. Ко второму классу относят команды процедурного языка разметки, благодаря которым осуществляется вывод данных как на экран монитора в текстовом процессоре, так и на печать. Однако, структура и состав объектов иерархической модели электронного документа уникальна в рамках одного процедурного языка разметки. В связи с этим в существующих технологиях формирования электронных документов использование различных процедурных языков разметки затруднено [5].

В общем виде модель структуры электронного документа в формате некоторого процедурного языка разметки может быть представлена следующим образом:

$$M_{PML} = \langle \{OBJECT_i\}, \{COMMAND_j\} \rangle, \quad (1)$$

$$i = 1 \dots n; j = 1 \dots m,$$

где M_{PML} – модель структуры электронного документа в формате некоторого процедурного языка разметки;

$OBJECT_i$ – объект первого класса в контексте процедурных языков разметки;

$COMMAND_j$ – объект второго класса в контексте процедурных языков разметки.

Основным недостатком такой модели является то, что множества объектов $OBJECT_i$ и $COMMAND_j$ уникальны для конкретного процедурного языка разметки. Следовательно, эта модель не может применяться для представления структуры электронного документа в формате произвольного процедурного языка разметки.

В связи с тем, что в рамках технологии формирования электронных документов в качестве документов-шаблонов используются описательные языки разметки, модель представления структуры выходного документа должна учитывать особенности, соответствующие этим языкам разметки [4]. В общем виде структуру электронного документа в формате описательного языка разметки можно представить ориентированным графом вида:

$$E = (\{V_i\}, \{A_j\}), i = 1 \dots n; j = 1 \dots m, \quad (2)$$

где E – ориентированный граф, отражающий структуру электронного документа в формате описательного языка разметки;

V_i – вершина ориентированного графа E ;

A_j – ребро ориентированного графа E .

Следует отметить, что вершинам V_i графа E соответствует элементы структуры электронного документа в контексте некоторого описательного языка разметки, а ребро A_j определяет отношения между ними. Граф E представим следующим образом:

Ориентированный граф E имеет только один корень, а уровень его вершин определяется кратчайшим путем до корня. Вершине нулевого уровня (корню), как правило, соответствует объект структуры электронного документа, который представляет его в целом. Вершины первого уровня соответствуют объектам, представляющим собой метаинфор-

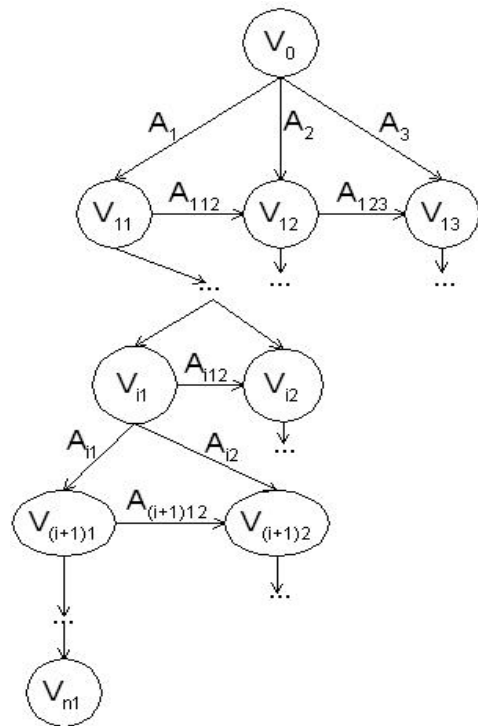


Рис. 1. Ориентированный граф E , который представляет структуру электронного документа

мацию о страницах, составляющих электронный документ. Вершины, располагающиеся в графе E со второго по последний (n) уровень, соответствуют различным объектам, которые входят в состав структуры электронного документа. Их тип, последовательность, а также состав их свойств определяются спецификацией того описательного языка разметки, в формате которого представлен рассматриваемый электронный документ.

Постановка задачи

На основании проведенного анализа можно сделать вывод о том, что на сегодняшний день не существует такой модели представления структуры электронных документов, форматом которых являются процедурные языки разметки, которая бы обеспечивала возможность формирования документов в рамках технологии, рассмотренной в работе [4]. В связи с этим возникает задача разработки модели представления структур электронных документов (форматом которых является процедурный язык разметки), которая может быть получена на основании измененной структуры документа-шаблона (форматом которого является описательный язык разметки). Далее такую модель будем называть моделью представления структур электронных документов с абсолютным позиционированием элемен-

тов. Выделим основные требования к разработке подобной модели:

- структура электронного документа, представляемая моделью с абсолютным позиционированием, должна быть получена на основании документа-шаблона, форматом которого является описательный язык разметки;

- модель должна позволять сформировать выходной документ, форматом которого является произвольный процедурный язык разметки, что предполагает наличие информации о физических размерах элементов структуры электронного документа и их координат относительно листа определенного размера.

Разработка модели представления структуры электронных документов с абсолютным позиционированием элементов

Опишем модель представления структур электронных документов с абсолютным позиционированием элементов (M) следующим образом:

$$M = \langle \bar{O}, \bar{\Phi} \rangle, \quad (3)$$

где \bar{O} – множество объектов, составляющих структуру электронного документа, представленных моделью M ;

$\bar{\Phi}$ – множество отношений между элементами множества \bar{O} .

Исходя из первого требования к разработке модели представления структур электронных документов с абсолютным позиционированием, можно сделать вывод о том, что множества объектов \bar{O} и множество отношений между ними $\bar{\Phi}$ определяются на основании элементов структуры электронного документа, форматом которого является описательный язык разметки. Как правило, одному элементу структуры электронного документа, форматом которого является описательный язык разметки, соответствует более чем один объект в модели M . Рассмотрим такую особенность элементов, которые соответствуют страницам электронного документа. В описательных языках разметки декларируется лишь метаинформация о типах страниц, которые входят в состав документа. Их конечное число определяется в момент визуализации документа текстовым процессором. В процедурных языках разметки страница электронного документа соответствует физической странице, которую видит пользователь в текстовом процессоре или в отпечатанном варианте. Такая особенность характерна и для других элементов структуры электронного документа в

контексте описательного языка разметки по отношению к объектам модели M .

Для перехода от модели представления структуры электронного документа с абсолютным позиционированием элементов к модели вида наличие сведений только лишь о физических размерах элементов и координатах элементов структуры электронного документа недостаточно. Это связано с тем, что команды процедурных языков разметки оперируют дополнительными параметрами, такими как цвет, тип, размер шрифта и прочее. Таким набором обладает каждый элемент структуры электронного документа в контексте описательного языка разметки, который далее будем называть контейнером свойств или стилем. В общем виде стиль объекта структуры электронного документа, представленный описательным языком разметки, можно представить следующим образом:

$$S = \langle p_1, p_2, \dots, p_i, \dots, p_{n-1}, p_n \rangle, i = \overline{1, n}, \quad (4)$$

где p_i – свойство стиля S ;

n – количество свойств в стиле S .

Свойство p_i опишем кортежем вида:

$$p_i = \langle K, V \rangle, \quad (5)$$

где K – ключ (название) свойства p_i ;

V – значение свойства p_i .

Опишем объекты модели M , которые соответствуют элементам структуры электронного документа, структура которого представлена ориентированным графом E . Объект O_d описывает элемент структуры электронного документа, который соответствует корню графа E и представляет документ в целом. Объект O_d имеет следующий вид:

$$O_d = \langle \{O_{s_1}, \dots, O_{s_i}, \dots, O_{s_{n-1}}, O_{s_n}\}, S_d \rangle, i = \overline{1, n}, \quad (6)$$

где O_{s_i} – объект O_s , который принадлежит объекту O_d ;

S_d – контейнер свойств (стиль), относящийся к документу в целом, такие как сведения о типе документа, авторе, алгоритме шифрования и прочее;

n – количество объектов O_{s_i} , которые принадлежат объекту O_d .

Объект O_s представляет собой страницу электронного документа и описывает элементы структуры электронного документа, которые соответствуют вершинам первого уровня графа E . Представим объект O_s следующим образом:

$$O_s = \langle \{O_{pa_1}, \dots, O_{pa_j}, \dots, O_{s_m}\}, S_s \rangle, j = \overline{1, m}, \quad (7)$$

где O_{raj} – объект printable area (область печати) O_{pa} , который принадлежит объекту O_s ;

S_s – контейнер свойств (стиль) страницы электронного документа, который содержит информацию о размерах страницы, величине отступов и т.д.; m – количество объектов O_{raj} , принадлежащих объекту O_s .

Основными элементами, которые формируют структуру электронного документа в контексте модели с абсолютным позиционированием элементов, являются объекты printable area (область печати) O_{pa} и физические элементы (O_p). Объект O_{pa} определяет некоторую область на странице документа, относительно которой позиционируются находящиеся на ней элементы. В структуре электронного документа, форматом которого является описательный язык разметки, объекту O_{pa} соответствуют такие элементы, как абзац (параграф), ячейка таблицы и другие. В ориентированном графе E , такие элементы располагаются, начиная со второго уровня. Объекту O_p соответствуют такие элементы, как текст и рисунок, то есть те, которые несут основную смысловую нагрузку в электронном документе. Объекты O_{pa} и O_p имеет следующий вид:

$$O_{pa} = \langle x, y, w, h, S_{pa}, \{O_{pa_1}, \dots, O_{pa_k}, \dots, O_{pa_o}\}, \{O_{p_1}, \dots, O_{p_q}, \dots, O_{p_r}\} \rangle, \quad k = \overline{1, o}, q = \overline{1, r}, \quad (8)$$

$$O_p = \langle x, y, w, h, S_p, VALUE \rangle,$$

где x, y – координаты объекта O_{pa} или O_p ;

w, h – ширина и высота объекта O_{pa} или O_p ;

S_{pa}, S_p – контейнер свойств (стиль) объекта O_{pa} или O_p ;

O_{pa_k} – объект, представляющий printable area (область печати) и принадлежащий объекту O_{pa} ;

O_{p_q} – объект, представляющий физический элемент и принадлежащий объекту O_{pa} ;

o – количество объектов O_{pa_k} , принадлежащих объекту O_{pa} ;

r – количество объектов O_{p_q} , принадлежащих объекту O_{pa} ;

VALUE – содержимое физического элемента O_p .

На рис. 2 показана структура электронного документа, представленная моделью M .

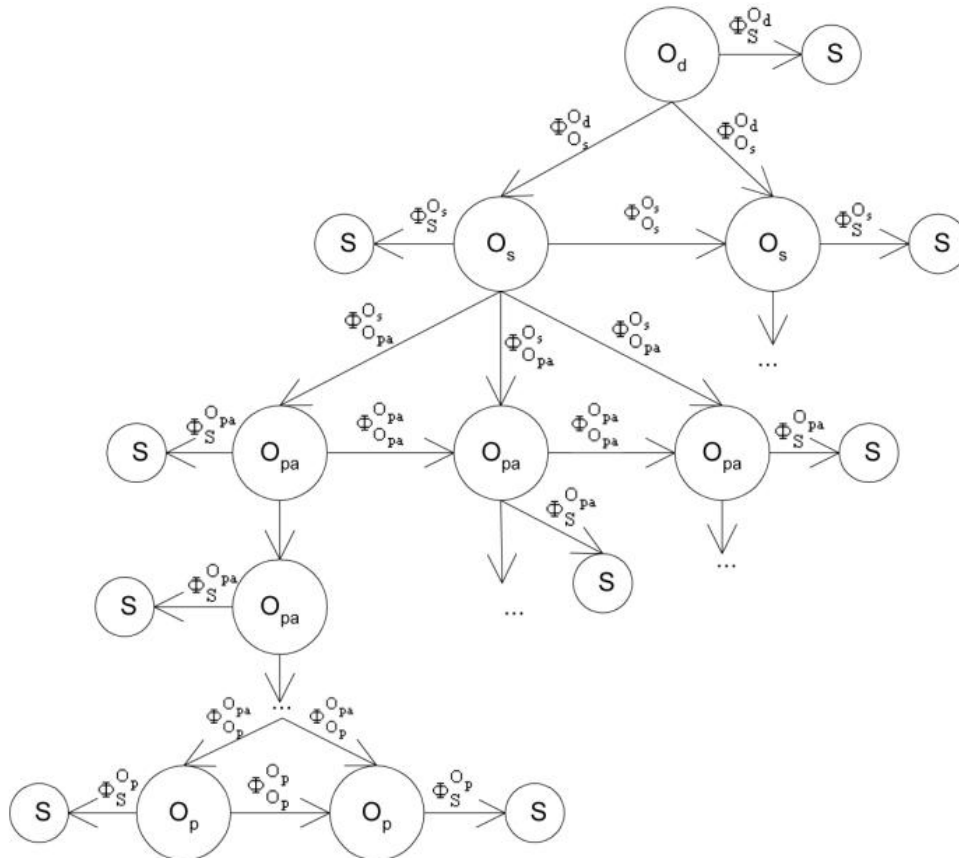


Рис. 2. Структура электронного документа в контексте модели M

Запишем модель M в общем виде:

$$M = \langle O_d, \overline{O_s}, \overline{O_{pa}}, \overline{O_p}, \overline{S}, \Phi_S^{O_d}, \overline{\Phi_{O_s}^{O_d}}, \overline{\Phi_{O_s}^{O_s}}, \overline{\Phi_S^{O_s}}, \overline{\Phi_{O_{pa}}^{O_s}}, \overline{\Phi_{O_{pa}}^{O_{pa}}}, \overline{\Phi_{O_{pa}}^{O_{pa}}}, \overline{\Phi_S^{O_{pa}}}, \overline{\Phi_{O_p}^{O_{pa}}}, \overline{\Phi_{O_p}^{O_p}}, \overline{\Phi_S^{O_p}} \rangle, \quad (9)$$

где O_d - объект, представляющий документ в целом;

$\overline{O_s}$ - множество объектов O_s , представляющих страницы документа;

$\overline{O_{pa}}$ - множество объектов printable area (область печати) O_{pa} ;

$\overline{O_p}$ - множество физических объектов O_p ;

\overline{S} - множество стилей S ;

$\Phi_S^{O_d}$ - отношение между объектом O_d и стилем S , определяющее связь между ними;

$\overline{\Phi_{O_s}^{O_d}}$ - множество отношений между объектами O_d и O_s , которое показывает, из каких страниц состоит электронный документ;

$\overline{\Phi_{O_s}^{O_s}}$ - множество отношений между двумя объектами O_s , определяющее порядок следования страниц в документе;

$\overline{\Phi_S^{O_s}}$ - множество отношений между объектами O_s и стилями S , определяющее связь между ними;

$\overline{\Phi_{O_{pa}}^{O_s}}$ - множество отношений между объектами O_s и O_{pa} , которое показывает, из каких объектов printable area (область печати) O_{pa} состоит страница документа O_s ;

$\overline{\Phi_{O_{pa}}^{O_{pa}}}$ - множество отношений между двумя объектами O_{pa} , определяющее порядок их следования в рамках общего контейнера;

$\overline{\Phi_{O_p}^{O_{pa}}}$ - множество отношений между двумя объектами O_{pa} , которое определяет, какие объекты printable area (область печати) O_{pa} являются дочерними элементами другого объекта O_{pa} ;

$\overline{\Phi_S^{O_{pa}}}$ - множество отношений между объектами O_{pa} и стилями S , определяющее связь между ними;

$\overline{\Phi_{O_p}^{O_{pa}}}$ - множество отношений между

объектами O_{pa} и O_p , которое показывает, из каких физических объектов O_p O_{pa} состоит объект printable area (область печати);

$\overline{\Phi_{O_p}^{O_p}}$ - множество отношений между двумя объектами O_p , определяющее порядок их следования в рамках общего контейнера

$\overline{\Phi_S^{O_p}}$ - множество отношений между объектами O_p и стилями S , определяющее связь между ними.

Выводы

Разработанная модель представления структуры электронного документа с абсолютным позиционированием элементов позволяет описывать структуры электронных документов, форматом которых является произвольные описательные языки разметки. Эта модель обеспечивает возможность перехода к структурам документов, форматом которых являются процедурные языки разметки. Это позволяет в технологиях формирования электронных документов использовать шаблоны-документы в формате описательных языков разметки, и генерировать выходные документы в формате произвольного процедурного языка разметки.

Литература

1. *Lowagi Bruno IText in action: creating and manipulating PDF / Bruno Lowagi – Manning, 2007. – 657 p.*
2. *Goossens Michel The Latex Web Companion: Integrating TeX, HTML, and XML / Michel Goossens, Eitan M Gurari – Addison-Wesley, 1999. – 513 p.*
3. *Reid Glenn C. Thinking in PostScript / Glenn C. Reid – Addison-Wesley, 1999. – 221 p.*
4. *Левыкин В.М. Разработка модели формирования электронных документов в WEB-ориентированных информационных системах / В.М. Левыкин, Е.А. Моспан // АСУ и приборы автоматки. - 2008. - Вып. 144. - С. 54-58.*
5. *Чальый С.Ф. Разработка модели модифицированной технологии формирования электронных документов на основании шаблонов в WEB-ориентированных информационных системах / С.Ф. Чальый, Д.Л. Кравченко, Е.А. Моспан // Сборник научных трудов Харьковского университета воздушных сил. - 2008. - Вып. 3 (18). - С. 135-138.*

Поступила в редакцію 12.09.2009

Рецензент: д-р техн. наук, проф. А. Я. Кузємин, Харківський національний університет радіоелектроніки, Харків, Україна.

РОЗРОБКА МОДЕЛІ ПРЕДСТАВЛЕННЯ СТРУКТУР ЕЛЕКТРОННИХ ДОКУМЕНТІВ З АБСОЛЮТНИМ ПОЗИЦІОНУВАННЯМ ЕЛЕМЕНТІВ

О.В. Петриченко, Є.О. Моспан

У статті наведені результати дослідження моделей представлення структур електронних документів, форматом яких є процедурні мови розмічування. Такі моделі забезпечують можливість використання процедурних мов розмічування як форматів вихідних документів у технологіях формування електронних документів в інформаційних системах. На підставі виявлених недоліків існуючих моделей розроблена модель представлення структур електронних документів з абсолютним позиціонуванням елементів, яка не залежить від типу процедурної мови розмічування. Така модель забезпечує можливість видачі вихідного документа у форматі довільної процедурної мови розмічування на підставі модифікованого документа-шаблону в технологіях формування електронних документів.

Ключові слова: електронний документ, документ-шаблон, технологія, модель структури електронного документа, процедурна мова розмічування.

DEVELOPMENT OF MODEL FOR REPRESENTING ELECTRONIC DOCUMENT STRUCTURES WITH ABSOLUTE POSITIONED ELEMENTS

A.V. Petrichenko, E. A. Mospan

In the article are described results of investigating of models for representing electronic document structures, which have format of procedural markup languages. Such models allow to use procedural markup languages as formats for output documents in technologies of electronic documents generation in informational systems. On the basis of founded disadvantages in existing models new model for representing electronic document structures with absolute positioned elements was developed, which is independent from a type of procedural markup language. Such model allow to generate output documents in a format of optional procedural markup language on the basis of modified document pattern in technologies of electronic documents generation.

Key words: electronic document, document pattern, technology, model for representing document structure, procedural markup language.

Петриченко Александр Вячеславович – канд. техн. наук, доцент кафедри Информационных управляющих систем Харьковского национального университета радиоэлектроники, Харьков, Украина, e-mail: alexander@profitsoft.com.ua.

Моспан Евгений Александрович – аспирант кафедры Информационных управляющих систем Харьковского национального университета радиоэлектроники, начальник отдела Java ООО «ПрофИТсофт», Харьков, Украина, e-mail: eugene.mospan@mail.ru.