

РАЗРАБОТКА ОБОБЩЕННОЙ МОДЕЛИ ПРЕДСТАВЛЕНИЯ СТРУКТУР ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Рассматривается модифицированная технология формирования электронных документов на основании документов-шаблонов. Как развитие данного подхода предлагается модель перехода между различными моделями представления такой структуры, что позволяет формировать выходные документы в формате произвольных процедурных и описательных языков разметки на основании документов-шаблонов, представленных описательным языком разметки.

1. Актуальность исследования

Актуальность темы обусловлена в первую очередь стремительным развитием web-базируемых информационных систем, а также многообразием существующих форматов электронных документов (ЭД), которые используются при работе с такими системами. Быстрота разработки и постоянного развития систем такого класса требует постоянного вовлечения заказчика в процесс их развития в целом и формирования документов-шаблонов для их использования в рамках систем. Сложность процесса вовлечения заказчика обусловлена тем, что формат документов-шаблонов в существующих технологиях формирования ЭД является специфическим и требует специальных знаний для формирования их структуры, а также ограниченным набором используемых форматов. Применение произвольного языка разметки выходного документа позволяет на этапе проектирования максимально вовлечь заказчика, а также расширить спектр устройств, при помощи которых осуществляется доступ к web-базируемым информационным системам, и использовать для просмотра ЭД то программное обеспечение, которое установлено.

2. Постановка задачи и цель

Предложенная в [1] модифицированная технология формирования ЭД предполагает их создание на основании документов-шаблонов, форматом которых является произвольный описательный язык разметки (ОЯР). В связи с этим для представления структуры документа-шаблона в [2] разработана унифицированная модель, которая не зависит от ОЯР, используемого в качестве формата документа-шаблона. Такая модель позволяет формально представить структуру документа-шаблона и разработать методы ее преобразования. Форматом выходных документов могут быть произвольные языки разметки описательного и процедурного типов. При этом для формирования выходного документа в формате произвольного ОЯР используется измененная структура документа-шаблона, представленная унифицированной моделью. Однако такая модель не позволяет формировать выходные документы в формате произвольного процедурного языка разметки (ПЯР). Это связано с отличиями между процедурными и описательными языками разметки. Для реализации задачи формирования выходных документов в формате произвольного ПЯР разработана модель представления структур ЭД с абсолютным позиционированием элементов, рассмотренная в [3]. Измененная структура документа-шаблона, представленная такой моделью, используется для формирования выходных документов в формате произвольного ПЯР.

Для решения поставленной задачи процесс формирования выходного документа в модифицированной технологии представляется последовательностью преобразования моделей представления структуры документа-шаблона. Изначально документ-шаблон представлен моделью, которая соответствует ОЯР, являющегося его форматом, а затем структура документа-шаблона представляется унифицированной моделью. После преобразования такой структуры возможны два варианта. Первый вариант предполагает получение модели структуры модифицированного документа-шаблона, соответствующей ОЯР, в формате которого должен быть сформирован выходной документ. Во втором варианте осуществля-

ется переход к структуре модифицированного документа-шаблона, которая представлена моделью с абсолютным позиционированием документов, что позволяет получить модель представления структуры документа-шаблона, соответствующей ПЯР, в формате которого должен быть сформирован выходной документ. Целью данного исследования является разработка модели перехода между различными моделями представления структур ЭД, для достижения которой необходимо решить задачу построения такой модели, а также описать процесс перехода между моделями представления ЭД в модифицированной технологии.

3. Модель перехода между различными моделями представления структур ЭД

Представим обобщенную модель структуры электронного документа C_{DM} в следующем виде:

$$C_{DM} = (Ob_{DM}, \Phi_{DM}), \quad (1)$$

где Ob_{DM} – объекты, представленные моделями; Φ_{DM} – отображения между объектами.

В модифицированной технологии формирования ЭД используются следующие модели представления структур ЭД:

- модели представления структур ЭД, форматом которых является произвольный ОЯР;
- унифицированная модель M_1 представления структуры ЭД, не зависящая от ОЯР, который является его форматом;
- модель M_p представления структуры ЭД с абсолютным позиционированием элементов, не зависящая от ПЯР, который является его форматом;
- модели представления структур ЭД, форматом которых является произвольный ПЯР.

Модели M_1 и M_p подробно рассмотрены в [1] и [3] соответственно. Опишем модели представления структур ЭД, форматом которых являются произвольные ОЯР Descriptive Markup Language (DML) и ПЯР.

Тогда модель представления структуры ЭД в сигнатуре модели M_a , форматом которого является язык DML, опишем следующим образом:

$$M_{DML} = \langle \overline{O_{DML}}, \overline{\Phi_{DML}} \rangle, \quad (2)$$

где M_{DML} – модель представления структуры ЭД, форматом которого является язык DML; $\overline{O_{DML}}$ – множество объектов, составляющих структуру ЭД, представленного моделью M_{DML} ; $\overline{\Phi_{DML}}$ – множество отображений между объектами множества $\overline{O_{DML}}$.

Определим множества $\overline{O_{DML}}$ и $\overline{\Phi_{DML}}$ на основании спецификации языка разметки DML и представим следующим образом:

$$\begin{aligned} \overline{O_{DML}} &= \langle \overline{O_{DML_1}}, \overline{O_{DML_2}}, \dots, \overline{O_{DML_a}}, \dots, \overline{O_{DML_n}} \rangle, a = \overline{1, n}; \\ \overline{\Phi_{DML}} &= \langle \overline{\Phi_{DML_1}}, \overline{\Phi_{DML_2}}, \dots, \overline{\Phi_{DML_b}}, \dots, \overline{\Phi_{DML_m}} \rangle, b = \overline{1, m}; \end{aligned} \quad (3)$$

здесь $\overline{O_{DML_a}}$ – множество объектов типа «а», каждый из элементов множества $\overline{O_{DML_a}}$ представляет собой некоторый объект структуры ЭД, форматом которого является язык DML; n – количество типов элементов в спецификации языка DML; $\overline{\Phi_{DML_b}}$ – множество отображений типа "b", относящихся к паре объектов множества $\overline{O_{DML}}$; m – количество типов отображений между парами объектов множества $\overline{O_{DML}}$ в спецификации языка DML.

Соответственно обозначим некоторый ПЯР Procedural Markup Language (PML). Тогда модель представления структуры ЭД в сигнатуре модели M_a , форматом которого является язык PML, опишем следующим образом:

$$M_{PML} = \langle \overline{O_{PML}}, \overline{\Phi_{PML}} \rangle, \quad (4)$$

где M_{PML} – модель представления структуры ЭД, форматом которого является язык PML; $\overline{O_{PML}}$ – множество объектов, составляющих структуру ЭД, представленного моделью M_{PML} ; $\overline{\Phi_{PML}}$ – множество отображений между объектами множества $\overline{O_{PML}}$.

Определим множества $\overline{O_{PML}}$ и $\overline{\Phi_{PML}}$ на основании спецификации языка разметки DML и представим следующим образом:

$$\begin{aligned} \overline{O_{PML}} &= \langle \overline{O_{PML_1}}, \overline{O_{PML_2}}, \dots, \overline{O_{PML_k}}, \dots, \overline{O_{PML_q}} \rangle, k = \overline{1, q}; \\ \overline{\Phi_{PML}} &= \langle \overline{\Phi_{PML_1}}, \overline{\Phi_{PML_2}}, \dots, \overline{\Phi_{PML_l}}, \dots, \overline{\Phi_{PML_r}} \rangle, l = \overline{1, r}; \end{aligned} \quad (5)$$

здесь $\overline{O_{PML_k}}$ – множество объектов типа "k", определяющих модель представления структуры M_{PML} ; q – количество типов элементов в спецификации языка PML; $\overline{\Phi_{PML_l}}$ – множество отображений типа "l", относящиеся к паре объектов множества $\overline{O_{PML}}$; r – количество типов отображений между парами типами объектов множества $\overline{O_{PML}}$ в спецификации языка PML.

Определим множество объектов Ob_{DM} модели C_{DM} (1) более подробно следующим образом, учитывая, что документ-шаблон может быть представлен произвольным ОЯР, а выходной документ – произвольным ПЯР или ОЯР в модифицированной технологии формирования ЭД:

$$Ob_{DM} = \langle \{M_{DML_i}\}, M_1, M_p, \{M_{PML_j}\} \rangle, i = \overline{1, n}, j = \overline{1, m}, \quad (6)$$

где $\{M_{DML_i}\}$ – модели представления структур ЭД, форматом которого является произвольный ОЯР; M_1 – унифицированная модель представления ЭД; M_p – модель представления структуры ЭД с абсолютным позиционированием элементов; $\{M_{PML_j}\}$ – модели представления структур ЭД, форматом которого является произвольный ПЯР.

Пусть ОЯР DML' – формат документа-шаблона, структура которого представлена моделью $M_{DML'}$. Предположим, что выходной документ необходимо сформировать в двух форматах ПЯР PML' и ОЯР DML'' , структуры которых представлены моделями $M_{PML'}$ и $M_{DML''}$ соответственно. Тогда в рамках модифицированной технологии формирования ЭД для создания выходного документа в формате DML'' на основании документа-шаблона, форматом которого является DML' , необходимо последовательное преобразование модели $M_{DML'}$ в унифицированную модель M_1 , а затем M_1 – в модель $M_{DML''}$. В случае формирования выходного документа в формате PML' , помимо преобразования $M_{DML'}$ в M_1 , необходимо преобразование M_1 в модель M_p представления структуры ЭД с абсолютным позиционированием элементов, а затем M_p в модель $M_{PML'}$. Исходя из сказанного выше, необходимо сформировать функциональную взаимосвязь между описанными моделями. Определим следующие функциональные отображения:

- множество отображений $\{f_{DP_i}\}$ между множеством моделей $\{M_{DML_i}\}$ и моделью M_1 – (DP – Document Pattern);
- отображения f_{LM} между моделями M_1 и M_p – (LM – Layout Model);
- множество отображений $\{f_{ODD_i}\}$ между моделью M_1 и множеством моделей $\{M_{DML_i}\}$ – (ODD – Output Document Descriptive);
- множество отображений $\{f_{ODP_i}\}$ между моделью M_p и множеством моделей $\{M_{PML_i}\}$ – (ODP – Output Document Procedural).

Рассмотрим указанные функциональные отображения более подробно. Благодаря множеству отображений $\{f_{DP_i}\}$ в рамках модифицированной технологии формирования ЭД существует возможность перейти от некоторой модели $M_{DML'}$ представления структуры документа-шаблона, которая зависит от языка разметки, являющегося его форматом, к унифицированной модели M_1 , не зависящей от ОЯР исходного документа-шаблона. Покажем функциональную связь между моделями $M_{DML'}$ и M_1 . В [1] подробно рассмотрены объекты, которые составляют иерархическую структуру ЭД в рамках модели M_1 . Среди таких объектов отметим следующие: E_d – объект, представляющий документ в целом; E_s – объект, представляющий секции документа; E_l – логический объект; E_{pc} – контейнер физических объектов; E_p – физический объект. Основным структурным элементом модели M_1 является логический объект E_l , который может соответствовать любому структурному элементу, образующему иерархическую структуру в рамках модели $M_{DML'}$. Стиль (контейнер свойств) S объекта E_l позволяет конкретизировать тип объекта в рамках модели $M_{DML'}$, например, параграф, таблица, ячейка таблицы и т.д. В свою очередь физический объект E_p также может иметь несколько прототипов в рамках модели $M_{DML'}$, например, текст, рисунок и т.д. Все необходимые свойства объекта E_p содержатся в стиле S контейнера физических объектов E_{pc} . В [1] определены отображения между описанными объектами. Для описания отображений между моделями $M_{DML'}$ и M_1 важны следующие отображения: отображение $\Phi_{E_s}^{E_d}$, которое определяет, что секция E_s является частью документа E_d ; отображение $\Phi_{E_l}^{E_s}$, которое показывает, что объект E_l является дочерним элементом относительно объекта E_s ; отображение $\Phi_{E_{pc}}^{E_l}$, которое показывает, что объект E_{pc} является дочерним элементом относительно объекта E_l ; отображение $\Phi_{E_p}^{E_{pc}}$ определяет, что физический объект E_p является дочерним элементом относительно объекта E_{pc} ; отображения $\Phi_S^{E_l}$ и $\Phi_S^{E_{pc}}$, которые показывают, что стиль S принадлежит логическому объекту E_l или контейнеру физических объектов E_{pc} соответственно.

Объектами модели $M_{DML'}$ являются: DO'_{DML} – объект, который соответствует документу в целом; SO'_{DML} – объект, который соответствует секции документа; LO'_{DML} , LO''_{DML} – структурные объекты в рамках модели $M_{DML'}$; PO'_{DML} , PO''_{DML} – физические объекты в рамках модели $M_{DML'}$. Между указанными выше объектами модели $M_{DML'}$

выделим следующие отображения: $\Phi_{SO'_{DML}}^{DO'_{DML}}$ – отображение между объектами DO'_{DML} и

SO'_{DML} , которое определяет принадлежность секций к документу; $\Phi_{LO'_{DML}}^{SO'_{DML}}$, $\Phi_{LO''_{DML}}^{SO'_{DML}}$ –

отображения между объектами SO'_{DML} и LO'_{DML} , LO''_{DML} , которые определяют принадлежность структурных объектов к секции, определяя тем самым иерархическую структуру

секции ЭД; $\Phi_{PO'_{DML}}^{LO'_{DML}}$, $\Phi_{PO''_{DML}}^{LO''_{DML}}$ – отображения между структурными объектами LO'_{DML} , LO''_{DML} и физическими PO'_{DML} , PO''_{DML} , которые определяют порядок следования физических объектов в иерархической структуре ЭД. Отразим функциональную связь между объектами и отображениями моделей $M_{DML'}$ и M_1 посредством отображения f'_{DP} , принадлежащего множеству $\{f_{DP_i}\}: M_{DML'}$ (рис.1).

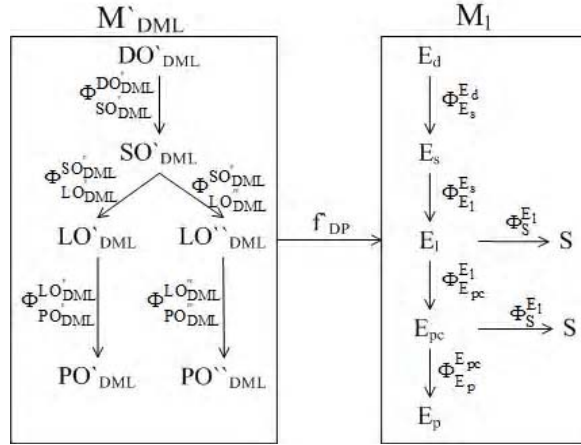


Рис. 1 Функциональная связь между моделями M_1 и $M_{DML'}$

Определим отображение f'_{DP} следующим образом:

$$f'_{DP} : M_{DML'} \rightarrow M_1, \quad (7)$$

Отображение f'_{DP} устанавливает следующие функциональные связи между объектами и отображениями моделей M_1 и $M_{DML'}$: объекту E_d соответствует DO'_{DML} ; объекту E_s соответствует SO'_{DML} ; отображению $\Phi^{E_d}_{E_s}$ соответствует $\Phi^{DO'_{DML}}_{SO'_{DML}}$; объекту E_l в совокупности со своим стилем S и отображением $\Phi^{E_l}_S$ соответствуют объекты LO'_{DML} и LO''_{DML} ; объекту E_{pc} в совокупности со своим стилем S и отображением $\Phi^{E_{pc}}_S$, а также дочерним физическим объектом E_p , определяемому на основании отображения $\Phi^{E_{pc}}_{E_p}$, соответствуют объекты PO'_{DML} и PO''_{DML} ; отображению $\Phi^{E_l}_{E_{pc}}$ соответствуют отображения $\Phi^{LO'_{DML}}_{PO'_{DML}}$, $\Phi^{LO''_{DML}}_{PO''_{DML}}$.

На основании изложенного выше можно утверждать, что любому объекту или отображению модели $M_{DML'}$ соответствует объект или отображение в модели M_1 , что подтверждает функциональную связь между указанными моделями. В общем виде отображение между моделями из множества $\{M_{DML_i}\}$ в модель M_1 может быть представлено как:

$$f_{DP_i} : M_{DML_i} \rightarrow M_1, i = \overline{1, n}. \quad (8)$$

Отображения из множеств $\{f_{ODD_i}\}$ (между моделью M_1 и моделями из множества $\{M_{DML_i}\}$) и $\{f_{ODP_i}\}$ (между моделью M_p и множеством моделей $\{M_{PML_i}\}$) раскрываются по аналогии с отображениями из множества $\{f_{DP_i}\}$ и в общем виде могут быть представлены следующим образом:

$$f_{ODD_i} : M_1 \rightarrow M_{DML_i}, i = \overline{1, n}, \quad (9)$$

$$f_{ODP_j} : M_p \rightarrow M_{PML_j}, j = \overline{1, m}. \quad (10)$$

Рассмотрим более подробно отображение f_{LM} , которое определяет функциональную связь между моделями M_1 и M_p . Объекты и отображения модели M_p описаны в [3], из которых необходимо выделить следующие: O_d представляет собой объект в целом в рамках модели M_p ; объект O_s представляет собой страницу ЭД; объект область печати O_{pa} является основным логическим элементом структуры ЭД в рамках модели M_p ; физический объект O_p ; отображение $\Phi^{O_d}_{O_s}$ определяет, из каких элементов O_s состоит ЭД документ, представленный O_d ; отображение $\Phi^{O_{pa}}_{O_s}$ определяет, из каких дочерних элементов O_{pa} состоит O_s ; отображение $\Phi^{O_{pa}}_{O_p}$ определяет, из каких дочерних элементов O_p состоит O_{pa} ; отображения $\Phi^{O_{pa}}_S$ и $\Phi^{O_p}_S$ определяют связь между объектами O_{pa} и O_p и их стилями S соответственно.

Отразим функциональную связь между объектами и отображениями моделей M_1 и M_p посредством отображения f_{LM} следующим образом (рис.2).

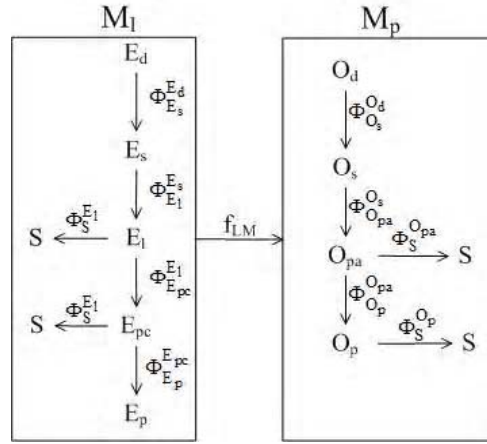


Рис. 2. Функциональная связь между моделями M_1 и M_p

Определим отображение f_{LM} следующим образом:

$$f_{LM} : M_1 \rightarrow M_p. \quad (11)$$

Отображение f_{LM} устанавливает следующие функциональные связи между объектами и отображениями между моделями M_1 и M_p : объекту E_d соответствует O_d ; объекту E_s соответствует O_s ; отображению $\Phi_{E_s}^{E_d}$ соответствует $\Phi_{O_s}^{O_d}$; объекту E_1 в совокупности со своим стилем S и отображением $\Phi_S^{E_1}$ соответствует объект O_{pa} в совокупности со своим стилем S и отображением $\Phi_S^{O_{pa}}$; объекту E_{pc} в совокупности со своим стилем S и отображением $\Phi_S^{E_{pc}}$, а также дочерним физическим объектом E_p , определяемому на основании отображения $\Phi_{E_p}^{E_{pc}}$, соответствует объект O_p в совокупности со своим стилем S и отображением $\Phi_S^{O_p}$.

На основании сказанного выше отображения Φ_{DM} модели C_{DM} можно представить следующим образом:

$$\Phi_{DM} = \langle \{f_{Dpi}\}, f_{LM}, \{f_{ODDi}\}, \{f_{ODPi}\} \rangle, i = \overline{1, n}, j = \overline{1, m}. \quad (12)$$

Используя ранее введенные обозначения, отразим переход между моделями представления структур ЭД и модифицированной технологией формирования ЭД на следующей диаграмме (рис.3):

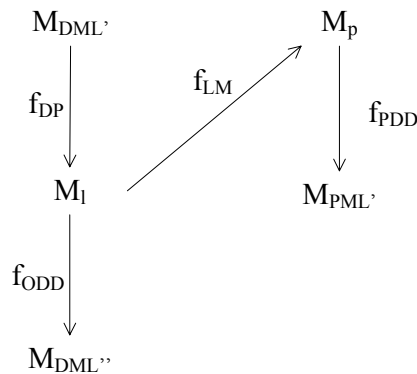


Рис. 3. Диаграмма, отражающая процесс перехода между моделями представления ЭД в модифицированной технологии формирования ЭД

Для перехода от структуры документа-шаблона, форматом которого является ОЯР DML', к выходному документу, форматом которого является ОЯР DML'', необходимо применить композицию отображений:

$$MT_{DML''}^{DML'} = f_{DP} \times f_{ODD} \cdot \quad (13)$$

А для формирования выходного документа в формате ПЯР PML':

$$MT_{PML'}^{DML'} = f_{DP} \times f_{LM} \times f_{PDD} \cdot \quad (14)$$

Таким образом, на основании полученных отображений $MT_{DML''}^{DML'}$ и $MT_{PML'}^{DML'}$ можно представить переход между моделями представления ЭД в модифицированной ИТ формирования ЭД. Такие отображения представляют частный случай формирования ЭД на основании конкретного формата документа-шаблона и заданного формата выходного документа. Однако определенные в модели S_{DM} объекты и морфизмы позволяют сформировать выходной документ в произвольном ОЯР или ПЯР на основании документа-шаблона, форматом которого является произвольный ОЯР.

Научные и практические результаты и выводы

На основании проведенного анализа и исследований разработана модель перехода между различными представлениями структуры электронного документа, которая составляет основу модифицированной технологии и позволяет формировать выходные документы в формате произвольных процедурных и описательных языков разметки на основании документов-шаблонов, представленных ОЯР. Полученная модель перехода позволила реализовать программные модули, соответствующие разработанным моделям представления структур ЭД и методам их преобразования, что в свою очередь дало возможность использовать данные разработки при проектировании web-базируемых информационных систем различного класса.

Список литературы: 1. Чалый С.Ф., Кравченко Д.Л., Моспан Е.А. Разработка модели модифицированной технологии формирования электронных документов на основании шаблонов в WEB-ориентированных информационных системах // Сборник научных трудов Харьковского университета воздушных сил. 2008. Вип. 3 (18). С. 135-138. 2. Левыкин В. М., Моспан Е. А. Разработка унифицированной модели представления структур электронных документов // Комп'ютерні науки та інформаційні технології. 2010. № 663. С. 39 - 45. 3. Петриченко О. В., Моспан Є. О. Разработка модели представления структур электронных документов с абсолютным позиционированием элементов. Радіоелектронні і комп'ютерні системи. 2009. № 3(37). С. 92-97.

Поступила в редколлегию 02.09.2012

Петриченко Александр Вячеславович, канд. техн. наук, ведущий научный сотрудник кафедры ИУС ХНУРЭ. Научные интересы: web-базируемые информационные системы, управление проектами. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 702-14-51.