

## МЕТОДИКА КОЛИЧЕСТВЕННОЙ ОЦЕНКИ ИНФОРМАЦИОННОГО НАПОЛНЕНИЯ ДОКУМЕНТОВ КОМПЛЕКТА ПРОЕКТНОЙ ДОКУМЕНТАЦИИ

**В работе рассмотрена методика синтеза проектных документов, обеспечивающая повышение их качества за счет использования квазиструктурированных моделей их информационного наполнения. Рассмотрена математическая интерпретация модели квазиструктурированного информационного наполнения электронных документов, предложены параметры количественной оценки информационного наполнения документа по его модели.**

**Ключевые слова:** моделирование документов; квазиструктурированная информация; проектная документация.

### Введение

Этап проектирования сложных технических объектов является очень важным, так как именно на этом этапе закладывается «фундамент» для разрабатываемого объекта. Использование документации ГОСТ при проектировании способствует повышению качества подготавливаемой документации.

Как правило, для создания нового документа в качестве основы используют аналогичный документ, подготовленный для другого проекта. В этом случае высока вероятность ошибки, обусловленная неточностями, связанными с возможной нестыковкой фрагментов старого документа и вновь вносимых в документ фрагментов. Старый документ используется проектировщиками в качестве шаблона, который представлен в виде модели документа, подготовленной в соответствии с требованиями ГОСТ, с занесенным в нее информационным наполнением.

Перспективной является разработка методики подготовки проектной документации, обеспечивающая повышение качества документации, достигаемое за счет использования квазиструктурированных моделей информационного наполнения проектных документов.

Таким образом, проектировщику для создания нового документа необходима модель проектного документа.

### Постановка задачи

Информационное наполнение проектных документов может быть описано квазиструктурированной моделью, в которой можно выделить некую структуру, которая заранее целиком или частично неизвестна, либо может меняться с течением времени [1]. В проектных документах присутствуют как фрагменты, обяза-

тельные к использованию по требованиям ГОСТ, так и необязательные фрагменты, а также те, которые могут повторяться в документе некоторое число раз и т. д.

Квазиструктурированная модель информационного наполнения  $S$  представляет собой графовую модель вида:

$$S = \langle root, sObj, LObj, minOccurs, maxOccurs, sMet, Obj\_smet \rangle, \quad (1)$$

где  $root$  – корневой объект,  $root \in sObj$ ;

$sObj$  – конечное множество объектов, каждый из которых содержит фрагмент информационного наполнения документа (текст, рисунок и т. д.) или выполняет роль контейнера для одного или нескольких объектов.

Для объектов-контейнеров доступны следующие метасвойства:  $smet_c$  – определяет объект в качестве контейнера;  $mixed$  – разрешает использование объектов-потомков в произвольном порядке;

$LObj$  – отображение, определенное на множестве  $sObj$ , такое, что  $sObj \xrightarrow{LObj} \{obj_1, \dots, obj_n\}$ , где  $obj_i \in sObj$  – дочерний объект;  $n$  – количество дочерних объектов;

$Obj\_met$  – отображение, определенное на множестве  $sObj$ , такое, что

$sObj \xrightarrow{Obj\_smet} \{smet_c \mid smet_c, mixed \mid smet_1, \dots, smet_k\}$ , где  $smet_i \in sMet$  – метасвойство ограничения на содержимое объекта;

$minOccurs$  – функция, определяющая минимально возможное количество раз использования объекта в модели;

$maxOccurs$  – функция, определяющая максимально возможное количество раз использования объекта в модели.

Рассмотрим пример графического представления документа (рисунок 1).

Документ, представленный на рисунке 1, состоит из пяти объектов. Объект *A* – выполняет роль контейнера для объектов *B* и *C*, объект *B* – выполняет роль контейнера для объектов *D* и *E*. Объекты *A*, *B*, *D* – обязательно должны быть использованы при разработке документа, объект *C* – является необязательным к использованию, объект *E* в рассматриваемом примере должен быть использован от трех до пяти раз.

Объекту-контейнеру *A* соответствует метасвойство ограничения  $smet_c$ , а для объекта-контейнера *B* определено дополнительно метасвойство  $mixed$ . Объект *C* представлен числовым наполнением, т. е. ему соответствует метасвойство ограничения  $smet_2$ . Объекты *D*, *E* имеют символическое информационное наполнение, которому соответствует метасвойство ограничения  $smet_1$ .

Таким образом, документ может быть представлен с помощью модели следующим образом:

$$\begin{aligned}
 &root = \{A\}; sObj = \{A, B, C, D, E\}; \\
 &LObj(A) = \{B, C\}, LObj(B) = \{D, E \mid E, D\}, \\
 &LObj(C) = \{ \}, LObj(D) = \{ \}, LObj(E) = \{ \}; \\
 &Obj\_smet(A) = \{smet_c\}, Obj\_smet(B) = \{smet_c, mixed\}, \\
 &Obj\_smet(C) = \{smet_2\}, Obj\_smet(D) = \{smet_1\}, \\
 &Obj\_smet(E) = \{smet_1\}; \\
 &minOccurs(A) = 1, \maxOccurs(A) = 1; \\
 &minOccurs(B) = 1, \maxOccurs(B) = 1; \\
 &minOccurs(C) = 0, \maxOccurs(C) = 1; \\
 &minOccurs(D) = 1, \maxOccurs(D) = 1; \\
 &minOccurs(E) = 3, \maxOccurs(E) = 5.
 \end{aligned}$$

### Практическая реализация

Модель *S* может быть реализована с помощью XML-технологии. Проектный документ будет представлен в виде XML-документа, а модель его квазиструктурированного информационного наполнения – в виде XSD-модели.

Стоит отметить, что Microsoft Word начиная с версии 2003 Professional поддерживает технологию XML, а проектные документы, как правило, формируются именно этим текстовым процессором.

При работе с XML-документами стандартом определены два уровня «правильности» [2]: правильно построенный (Well-formed), действительный (Valid).

Правильно построенный документ соответствует всем правилам синтаксиса XML. Действительный документ дополнительно соответствует некоторым семантическим правилам.

Это более строгая дополнительная проверка корректности документа на соответствие заранее определенным, но уже внешним правилам в целях минимизации количества ошибок, например структуры и состава данного, конкретного документа или семейства документов.

Рассмотрим подробнее процесс валидации XML-документов, для этого рассмотрим схему, изображенную на рисунке 2.

Из анализа схемы валидации можно сделать вывод, что использование схемы (модели документа) обеспечивает контроль корректности вводимых данных в XML-документ. Документ сохраняется без выдачи сообщения об ошибке только при полном соответствии модели документа.

### Пример использования

Для примера рассмотрим модель документа «Руководство пользователей АРМ» (АРМ – автоматизированное рабочее место) из пакета документов рабочей документации, подготовленного в соответствии с требованиями ГОСТ 19.505-79 «Руководство оператора. Требования к содержанию и оформлению».

Модель документа «Руководство пользователей АРМ» состоит из сегментов (рисунок 3). Каждый сегмент отвечает за описание раздела документа, а он в свою очередь подразделяется на более мелкие сегменты и объекты.

Сегменты «abbr\_list» и «user\_guide» показаны на схеме пунктиром. При описании доку-

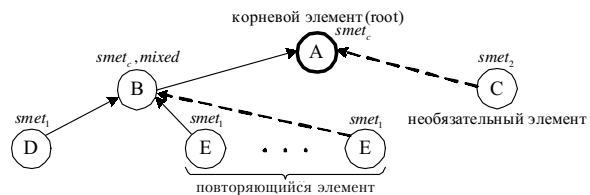


Рисунок 1. Граф информационного наполнения документа

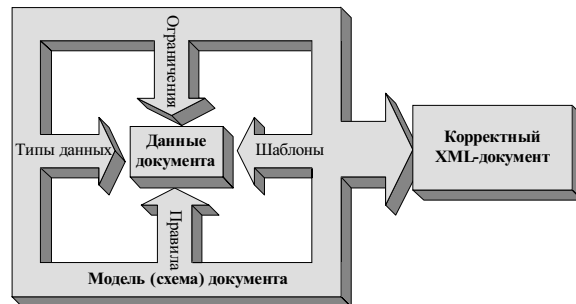


Рисунок 2. Схема валидации XML-документа

мента эти элементы являются необязательными и могут не присутствовать в документе.

Сегмент «title» описывает информационное наполнение титульного листа документа (рисунок 4), которое состоит из объектов «organization» – организация-разработчик, «project\_name» – название проекта, «doc\_name» – название документа, «doc\_type» – тип документа и двух сегментов: сегмента «cipher» – полный шифр документа, состоящего из двух атрибутов «cipher\_project» – шифра проекта и «cipher\_doc» – шифра документа, и сегмента «dev\_dt\_plc» – место и год создания документа, в кото-

ром определены два атрибута: «dev\_place» и «dev\_date» – место создания документа и дата создания документа соответственно.

На втором листе документа размещена информация о руководителях проектной организации, которая описана с помощью сегмента «chiefs» (рисунок 5). Этот сегмент включает в себя следующие объекты: «engineer\_institute» – главный инженер института, «engineer\_project» – главный инженер проекта, «chief\_department» – начальник отдела. Каждый из перечисленных объектов будет содержать Ф. И. О. соответствующего руководителя.

Далее использован сегмент «list\_of\_executors» – список исполнителей документа (рисунок 6). Этот сегмент состоит из списка авторов документа, для каждого из которых определены его Ф. И. О. и занимаемая им должность. Количество исполнителей документа неограниченно.

Содержание документа описано с помощью объекта «contents», который представляет собой блок неограниченного размера.

Сегмент «abbr\_list» позволяет описать сокращения, использованные в документе. Структура этого сегмента представлена на рисунке 7. Количество использованных аббревиатур в документе неограниченно. Каждое сокращение определено с помощью двух объектов: «abbr\_name» – аббревиатура и «abbr\_decryption» – расшифровка аббревиатуры.

Состав комплекта руководств пользователей описан с помощью сегмента «user\_guide» (рисунок 8). Количество руководств комплекта неограниченно. Для каждого руководства определены следующие объекты: «cipher\_guide» – шифр руководства и «name\_guide» – название руководства.

Для описания функционального и эксплуатационного назначения АРМ используется сегмент «appointment» (рисунок 9). Для него определены следующие объекты: «functional\_appointment» – функциональное назначение, «maintenance\_appointment» – эксплуатационное назначение, а также с помощью объекта «function\_name» – функции АРМ может быть описано неограниченное количество функций, реализуемых АРМом.

Далее в документе опреде-

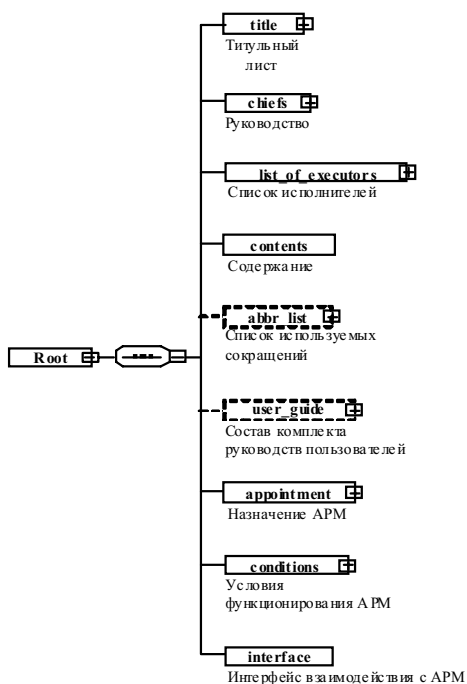


Рисунок 3. Базовый сегмент схемы

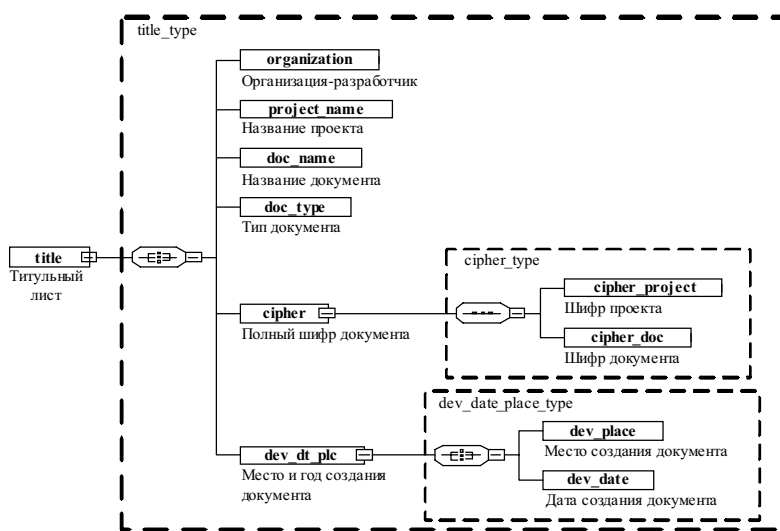


Рисунок 4. Сегмент титульного листа

лены условия функционирования АРМ. Их описание реализует сегмент «conditions» (рисунок 10), который в свою очередь состоит из сегментов «min\_system\_requirements» – минимальные системные требования, «adv\_system\_requirements» – рекомендуемые системные требования и объектов «conditions\_operator» – условия эксплуатации АРМ, «software» – список программного обеспечения (ПО).

Объект «conditions\_operator» определяет требования к рабочему месту оператора АРМ. Сегменты «min\_system\_requirements» и «adv\_system\_requirements» состоят из неограниченного количества объектов «hardware\_name» – наименование и «hardware\_metrics» – характеристика. Сегмент, формирующий рекомендуемые системные требования, является необязательным и может отсутствовать в документе. Эти сегменты позволяют описать минимальные и рекомендованные требования к компьютеру АРМ. Неограниченная последовательность объектов «software» позволяет задать перечень программных средств, требуемых для функционирования АРМ на компьютере.

Объект «interface» использован в модели для маркирования разделов, содержащих описание интерфейса взаимодействия с АРМом.

Предложенная модель документов «Руководство пользователей АРМ» позволяет описать информационное наполнение документов этого вида. Использование модели позволяет упростить процесс создания документа, так как в этом случае документ разрабатывается в соответствии со структурой и информационным наполнением, определенным моделью. На рисунке 11 изображен фрагмент документа, разработанного с использованием предложенной модели.

Использование моделей проектных документов обеспечивает автоматический

контроль над структурой и информационным наполнением, что позволяет снизить количество ошибок в документации и повысить качество документации.

Для документов, синтезированных по единой модели, может быть выполнена количественная оценка информационного наполнения относительно модели.

По аналогии с многокритериальной оптимизацией технических систем [3] будем ис-

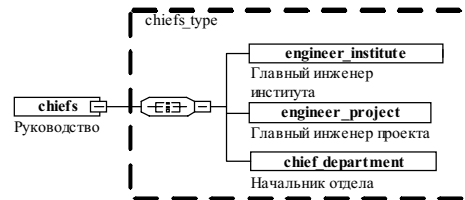


Рисунок 5. Сегмент второго титульного листа

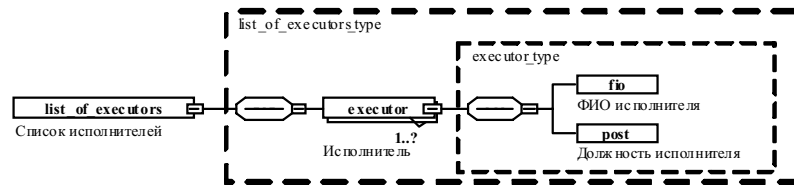


Рисунок 6. Сегмент «Список исполнителей документа»

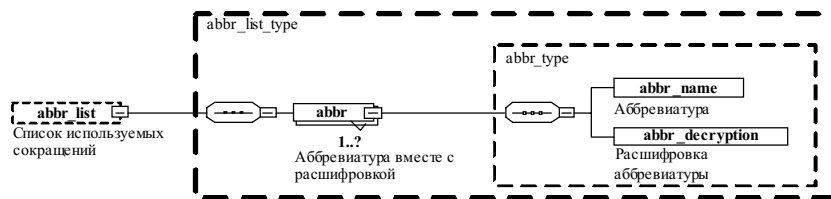


Рисунок 7. Сегмент «Список использованных сокращений»

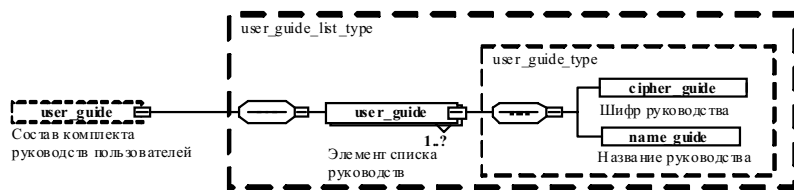


Рисунок 8. Сегмент «Состав комплекта руководств пользователей»

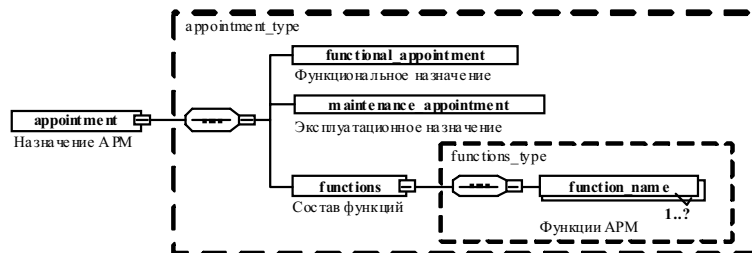


Рисунок 9. Сегмент «Назначение АРМ»

пользовать для методики оценки пять параметров.

В качестве первого параметра примем отношение количества использованных обязательных объектов к количеству обязательных к использованию объектов, определенных моделью документа. Данный параметр должен быть равен единице, так как использование обязательных параметров является условием применения модели документа:

$$\frac{obj\_count}{Obj(sObj)} = 1, \quad (2)$$

где *obj\_count* – функция, возвращающая количество обязательных объектов, использованных в документе;

*Obj* – отображение, ставящее в соответствие множеству *sObj* мощность множества, состоящего из элементов, удовлетворяющих условиям:  $Obj\_smet(sObj_i) \cap \{smet_c\} = \emptyset$ ;  $minOccurs(sObj_i) = 1$ ;  $maxOccurs(sObj_i) = 1$ ;  $LObj(sObj_i) = \emptyset$ , где  $sObj_i \in sObj$ .

Вторым параметром является количество использованных в документе необязательных объектов. Этот параметр должен быть максимизирован:

$$Uobj(sObj) \rightarrow max, \quad (3)$$

где *Uobj* – отображение, ставящее в соответствие множеству *sObj* мощность множества, состоящего из элементов, удовлетворяющих условиям:  $minOccurs(Obj_i) = 0$ ;  $maxOccurs(Obj_i) = 1$ ;  $LObj(sObj_i) = \emptyset$ , где  $sObj_i \in sObj$ .

Третий параметр характеризует количество повторяющихся объектов модели. Использование повторяющихся объектов модели должно быть максимизировано:

$$Robj(sObj) \rightarrow max, \quad (4)$$

где *Robj* – отображение, ставящее в соответствие множеству *sObj* мощность множества, состоящего из элементов, удовлетворяющих условиям:  $minOccurs(Obj_i) = 1$ ;  $maxOccurs(Obj_i) > 1$ ;  $LObj(sObj_i) = \emptyset$ , где  $sObj_i \in sObj$ .

Четвертым параметром является плотность использованных объектов в тексте документа. При формировании документа по его модели часть информационного наполнения не может быть формализована моделью. Данный параметр должен быть максимизирован, так как в процессе синтеза документа необходимо минимизировать неформализуемое информационное наполнение:

$$\frac{Cobj(sObj)}{char\_count} \rightarrow max, \quad (5)$$

где *Cobj* – отображение, ставящее в соответствие множеству *sObj* мощность множества, состоящего из элементов, удовлетворяющих условию:  $LObj(sObj_i) = \emptyset$ , где  $sObj_i \in sObj$ ; *char\_count* – функция, возвращающая количество символов в документе.

В качестве пятого параметра примем среднее значение количества символов внутри объекта. Данный параметр должен быть минимизирован, так как большой объем информационного наполнения внутри объекта повышает процент неформализованного информационного наполнения документа:

$$\frac{\sum_{i=1}^n Len(sObj_i)}{Cobj(sObj)} \rightarrow min, \quad (6)$$

где *len* – функция, возвращающая количество символов из объекта, а в рассмотрение прини-

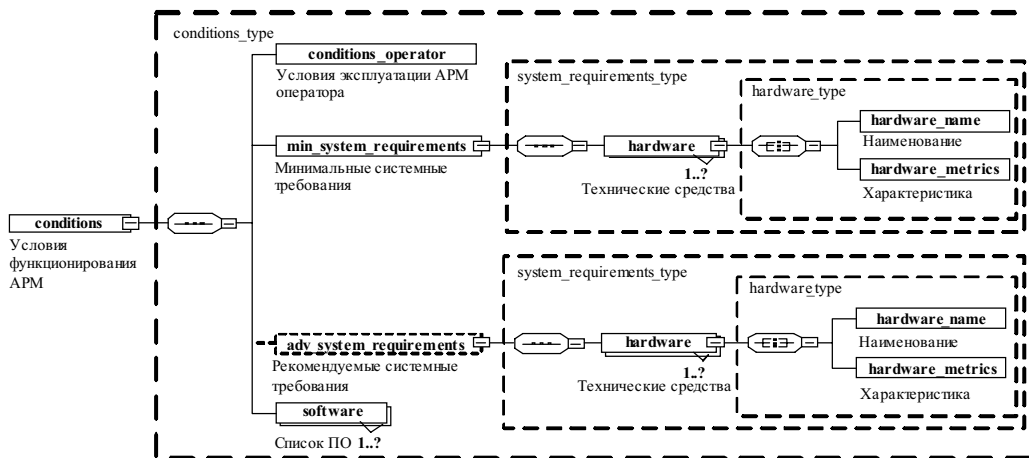


Рисунок 10. Сегмент «Условия функционирования АРМ»

маются только объекты  $sObj_i \in dObj$ , удовлетворяющие условию  $LObj(sObj_i) = \emptyset$ .

**Заклучение**

Рассмотренный набор (2–6) параметров реализует возможность сравнения документов, синтезированных по единой модели, в автоматизированном режиме без ознакомления с их информационным наполнением и позволяет решать задачи оптимизационного и имитационного моделирования для комплекта документов проектной документации.

12.05.2010

Функциональное назначение

«functional\_appointment» Функциональным назначением АРМ оператора БПТО и К является предоставление оператору возможности автоматизированного управления системой учета и транспортировки метанола. functional\_appointment\*

Эксплуатационное назначение

«maintenance\_appointment» АРМ оператора БПТО и К предназначен для эксплуатации на технологическом объекте БПТО и К. Конечными пользователями АРМ являются операторы БПТО и К. maintenance\_appointment\*

Состав функций

АРМ оператора обеспечивает возможность выполнения следующих функций:

«functions» - function\_name контроль давления метанола на входе в БСМ D;  
 - function\_name контроль расхода метанола на выходе из БПТО и К D;  
 - function\_name контроль давления метанола на выходе из БПТО и К D;  
 - function\_name контроль температуры метанола на выходе из БПТО и К D;  
 - function\_name контроль плотности метанола на выходе в БПТО и К D;  
 - function\_name контроль массового расхода метанола на выходе в БПТО и К D;  
 - function\_name сигнализация превышения плотности метанола на БПТО и К D;  
 - function\_name включение/отключение режима опрессовки метанолопровода D;  
 - function\_name сигнализация понижения давления в метанолопроводе D. functions\* appointment\*

									Лист
									9
Изм.	Кол.	Лист	№ док.	Подпись	Дата	07.082.088.3876-АСУ ИЗ.1			

Рисунок 11. Фрагмент документа, синтезированного по модели

**Список использованной литературы:**

1. Полищук, Ю.В. Моделирование системы обработки квазиструктурированных электронных документов / Ю.В. Полищук, Т.А. Черных // Вестник компьютерных и информационных технологий, 2009. №5. – С. 41-44.
2. Полищук, Ю.В. Автоматизация разработки технической документации с применением XML технологии / Ю.В. Полищук, Т.А. Черных // Нефтепромысловое дело. – 2008. – №11, – С. 112-115.
3. Гмошинский, В.Г. Теоретические основы инженерного прогнозирования / В.Г. Гмошинский, Г.И. Флиорент. – М.: Наука, 1973. – 304 с.

**Работа выполнена при поддержке ФЦП НК-136/П1 П1004 «Разработка технологии хранения и обработки квазиструктурированных данных»**

Сведения об авторах:

Полищук Юрий Владимирович, доцент кафедры математического обеспечения информационных систем Оренбургского государственного университета, кандидат технических наук  
 460018, г. Оренбург, пр-т Победы, 13, ауд. 2132, тел. (3532)372534, e-mail: youra\_polishuk@bk.ru

Черных Татьяна Александровна, аспирант кафедры математического обеспечения информационных систем Оренбургского государственного университета  
 460018, г. Оренбург, пр-т Победы, 13, ауд. 2132, тел. (3532)372534, e-mail: chatty84@mail.ru

Polishchuk Y.V., Chernykh T.A.

The procedure of quantitative assessment of the information value of the documents of the complete set of project documentation

The work examines the procedure of the synthesis of project documents, which ensures increase in their quality due to the use of the quasi-structure models of their information filling. The mathematical interpretation of the model of the quasi-structure information filling of electronic documents is examined and the parameters of the quantitative assessment of the information value of a document according to its model are proposed.

The key words: the simulation of documents; the quasi-structure information; project documentation.

**Bibliography**

1. Polishchuk Ju.V. Modeling System for kvazistrukturiro-bathroom electronic documents / Ju.V. Polishchuk, T.A. Chernykh / Vestnik komp'yuternykh i informacionnykh tekhnologijj, 2009. №5. – S. 41-44.
2. Polishchuk Ju.V. Automating the development of technical documentation using XML technologies / Ju.V. Polishchuk, T.A. Chernykh / Neftepromyslovoe delo. – 2008. – №11 – S. 112-115.
3. Gmoshinsky V.G. Theoretical foundation engineering forecasting / V.G. Gmoshinsky, G.I. Florent. – M.: Nauka, 1973. – 304 pp.