



КЛАССИФИКАЦИОННАЯ МОДЕЛЬ ТЕКСТОВ НОРМАТИВНОЙ БАЗЫ ПРИ ЭКСПЕРТИРОВАНИИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Шостак И. В., Бутенко Ю. И.

*Национальный аэрокосмический университет им. Н.Е. Жуковского
«Харьковский авиационный институт»
г. Харьков, ул. Чкалова 17, тел 7884735
e-mail: iv_shostak@rambler.ru, iuliabutenko@yandex.ru*

Экспертиза программного обеспечения в значительной мере определяет реальные возможности в обеспечении необходимого уровня безопасности и качества в отраслях и социальных сферах, связанных с системами критического применения. В то же время экспертиза ПО является слабо формализованным и слабо структурированным видом профессиональной деятельности. Велик субъективизм и влияние человеческого фактора на результаты экспертных оценок [1].

Динамика развития международной и национальной нормативной базы программной инженерии обуславливает необходимость постоянного проведения работ по формированию гармонизированных нормативных профилей, отражающих реальные потребности повышения качества продукции в конкретных прикладных областях [2].

Государственные и отраслевые стандарты, нормативно-методические документы предприятий и общие требования спецификаций ПО относятся к документам, обладающим высокой степенью формализованности и, как следствие, внутренней иерархической структурой. При поиске информации в коллекции сложно структурированных текстов стандартов недостаточно получить лишь список релевантных документов в качестве поисковой выдачи по причине больших объемов и высокой сложности документов. Повышение эффективности поиска в таких документах может быть достигнуто, если в качестве поисковой выдачи будут получены не только документы, но и цитаты из них — точные дословные выдержки из текста, обладающие смысловой законченностью. Цитаты могут быть получены с помощью анализа иерархической структуры текстов стандартов и НД, и далее могут быть уточнены с применением синтаксического анализа. В результате может быть получена компактная поисковая выдача, в которой отсечен значительный объем информации, нерелевантный запросу.

Для автоматизации процесса формирования нормативного профиля необходимо установить структуру языка стандартов, как совокупности элементов. Одним из средств структурной формализации текста выступает рубрикация, являющаяся внешним выражением его композиционной структуры. Основную часть текста стандарта принято делить на разделы, подразделы, пункты, подпункты. Обязательным является наличие заголовков для разделов и подразделов. Заголовки (подзаголовки) являются важным средством рубрикации и, следовательно, элементом структурной



формализации. Они в предельно сжатой, краткой и лаконичной форме позволяют отразить тематику и основную идею выделенной части документа, выступая тем самым как важные единицы сообщения, передающие определенную информацию [3].

Структурные элементы стандарта принято делить на элементы передней части: титульный лист, предисловие, содержание, введение; и элементы основной части: название, область применения, нормативные ссылки, термины и определения понятий, обозначения и сокращения, требования к объекту стандартизации, приложения и библиографические данные [4]. Структурные элементы, за исключением элементов «Титульный лист», «Предисловие», «Название», «Требования к объекту стандартизации», приводят по необходимости, в зависимости от особенностей стандартизуемого объекта.

Анализ моделируемых текстов показывает, что они строятся по единой строго установленной схеме и содержат определенный набор элементов, прописанный в ДСТУ 1.5 -2003 [4]. Закрепление за структурными элементами постоянного места делает документ более удобным для зрительного восприятия, упрощает его обработку и обуславливает возможность использования технических средств для формирования нормативного профиля.

За основу возьмем построение стандартов, предложенное в [5]. Текст стандарта целесообразно разделить на три части: предварительная часть, рекомендации и требования и информативная часть. К предварительной части относится введение, назначение, нормативные ссылки, термины и определения, общие сведения. К рекомендациям и требованиям относят разделы стандарта, в которых непосредственно описаны требования. Информативная часть содержит дополнительную информацию в виде приложений, не вошедшую в стандарт.

На языке теории категорий задача классификации объектов языка стандартов сводится к разработке и исследованию классификационной модели объектов языка стандартов, которая удовлетворяет следующим принципам:

– единства универсума, т.е. существования в некоторой категории \mathfrak{R} , которая описывает классификацию объектов языка стандартов, единого инициального объекта, который отвечает понятию «надсистема-класс»:

$$\exists! a_{\perp} \in \text{Ob}\mathfrak{R} \forall a \in \text{Ob}\mathfrak{R} \exists a_a \in \text{Mor}\mathfrak{R} a_a : a_{\perp} \rightarrow a \text{Mor}\mathfrak{R}(a_{\perp}, a) = \{a_a\} \quad (1)$$

– иерархичности, т.е. каждый объект категории \mathfrak{R} , которая описывает классификацию объектов языка стандартов, является вершиной конуса и, кроме того, для каждого объекта категории \mathfrak{R} существует некоторая подкатегория, для которой он является инициальным объектом:

$$\forall a \in \text{Ob}\mathfrak{R} \exists \text{Mor}\mathfrak{R}(a) \neq \emptyset; \forall a \in \text{Ob}\mathfrak{R} \exists \mathfrak{R}_a \subseteq \mathfrak{R} : a \in \text{Ob}_i \mathfrak{R}_a, \quad (2)$$

– параметричности, т.е. в некоторой категории \mathfrak{R} , которая описывает классификацию объектов языка стандартов, существует подкатегория \mathfrak{R}_{ce} , описывающая свойства любых элементов классифицируемой предметной области:

$$\exists \mathfrak{R}_{ce} \subset \mathfrak{R}. \quad (3)$$

Выше обоснована возможность представления языка стандартов в виде иерархической трехуровневой структуры, где при реализации



информационной технологии формирования нормативного профиля низший уровень – информативная часть, средний – рекомендации и требования, а верхний уровень – предварительная часть. Категорная модель языка стандартов при этом имеет следующий вид:

Рассмотрим категорию A . Класс ObA является конечным множеством с разбивкой:

$ObA = \bigcup_{i=0}^n A_i$, где $A_i = \{a_m^i\}_{m=1}^{k_i}$ – множество объектов, которые отвечают системам-классам i -го уровня иерархии ($k_0 = 1$, т.е. на нулевом (верхнем) уровне иерархии находится только один объект a_1^0 (система-класс предварительной части), $k_1 = 2, 3$, т.е. на первом и втором уровнях иерархии находятся только два объекта a_1^1 и a_2^1 (два системы-классы – рекомендации и требования и информативная часть соответственно), $k_2 \geq 2i$ и $k_{i+1} \geq 2ki$ ($i = \overline{2, n-1}$)).

Класс $MorA$ может быть описан таким образом. Для двух элементов

$$a_{m_1}^i a_2^1, a_{m_2}^j a_2^1 \quad (i = \overline{1, n}): \quad Mor(a_{m_1}^i, a_{m_2}^j) = \begin{cases} \emptyset, \text{ при } m_1 \neq m_2, \\ \{1_{a_{m_1}^i}\} \text{ при } m_1 = m_2 \end{cases} \quad (4)$$

Для произвольной пары (a_s^{i+1}, a_r^i) : $Mor(a_s^{i+1}, a_r^i) = \emptyset$, где $i = \overline{0, n-1}$, $r = \overline{1, k_i}$, $s = \overline{1, k_{i+1}}$. Для любого элемента $a_r^i \in A_i$ ($i = \overline{1, n}$):

$$Mor(a_s^{i-1}, a_r^i) = \begin{cases} \{\alpha_{s,r}^{i-1,i}\} \text{ при } s = s_r, \\ \emptyset \text{ при } s \neq s_r \end{cases} \quad (5)$$

$$\begin{cases} \{\alpha_{s,r}^{i-1,i}\} \text{ при } s = s_r, \\ \otimes \text{ при } s \neq s_r. \end{cases}$$

Для элемента a_1^0 подмножества A_0 : $Mor(a_1^0, a_1^0) = \{1_{a_1^0}\}$; для произвольной пары (a_2^1, a_r^i) : $Mor(a_2^1, a_r^i) = \emptyset$, где $i = \overline{2, n}$, $r = \overline{1, k_i}$.

Морфизмы категории A как связи между ее объектами описывают родовидовые связи между соответствующими системами-классами. После построения категории A , любому объекту любого уровня отвечает единый морфизм и единый объект верхнего уровня, связанный с этим объектом с помощью данного морфизма:

$$A_i \ni a_r^i \leftrightarrow a_{s_r}^{i-1} \in A_{i-1}; a_r^i \leftrightarrow a_{s_r,r}^{i-1,i} \in Mor(a_{s_r}^{i-1}, a_r^i), \quad (6)$$

где $i = \overline{1, n}$.

Для любого объекта $a_1^0 \in A_0$ морфизмом является $1_{a_1^0}$, который обозначим $a_{1,1}^{0,0} = 1_{a_1^0}$.

Конкретизацией приведенной выше модели языка стандартов является представление с помощью конструктора онтологий Protégé 4.2 фрагментов текстов стандартов, относящихся к конкретным объектам сертификации или предметным областям. Так на Рис.1 представлен фрагмент системы стандартов относящихся к объекту типа «атомная станция» из предметной области «ядерная энергетика».

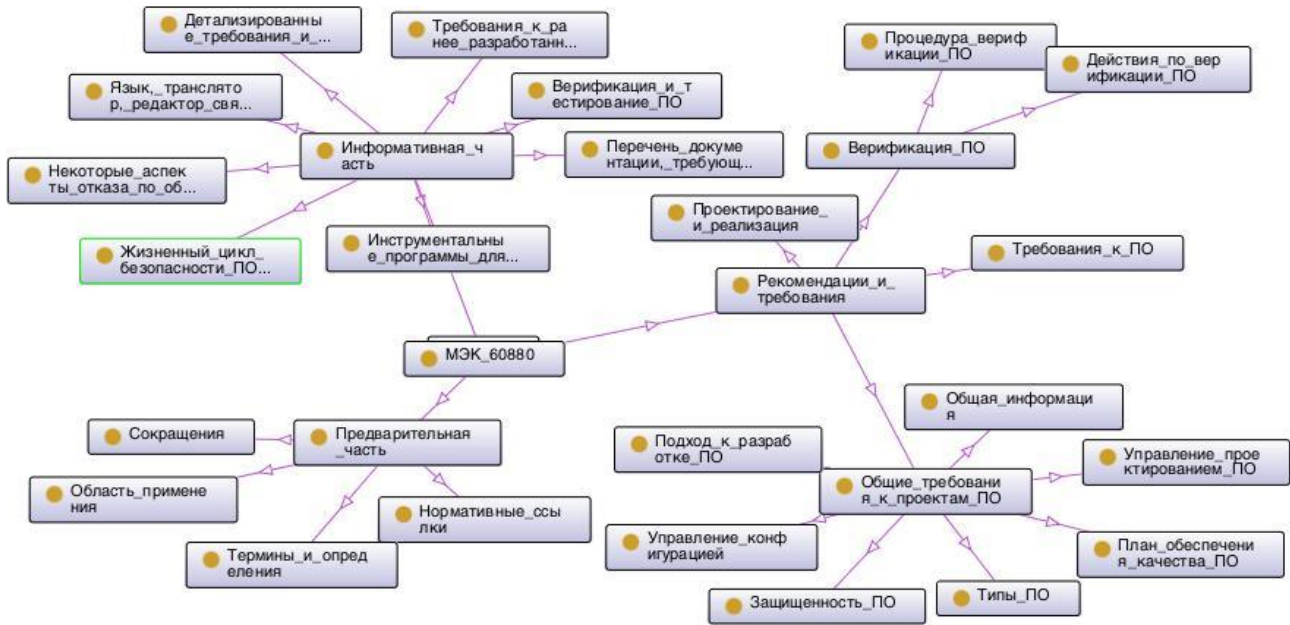


Рис. 1. Фрагмент стандарта МЭК 60880, построенный в Protégé 4.2

Следующим этапом реализации модели является синтез на ее основе системы онтологий: онтологии верхнего уровня, онтологии предметной области, онтологии источника знаний, онтологии задач и методов, онтологии запроса и онтологии-приложения. Онтологическая система станет ядром диалоговой системы поддержки принятия решений аудитором сертификационного центра.

Список литературы

1. Шостак И. В., Бутенко Ю. И., Шостак Е. И. Знаниеориентированные методы формирования нормативных профилей к системам критического применения на основе онтологий // Радиоэлектронные и компьютерные системы, 2010, №5, с. 104-108.
2. Нормативная база программной инженерии в разработке систем с интенсивным использованием программного обеспечения / Б.М. Конорев, Л.Ф. Пудовкина, И.Б. Сироджа, О.Е. Федорович. уч. пос. – Харьков: Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», 2001. – 162с.
3. Нелюбин Л. Л. Перевод и прикладная лингвистика. – М.; Высш. школа, 1983 – 207с.
4. ДСТУ 1.5:2003 Національна стандартизація. Правила побудови, викладення, оформлення та вимоги до змісту нормативних документів
5. Ястребенецкий М.А. Управление старением критических систем // Радиоэлектронные и компьютерные системы, 2008, №6, С. 114-121.