УДК 004.81

КОГНИТИВНАЯ МОДЕЛЬ ОЦЕНКИ УРОВНЯ ДОСТОВЕРНОСТИ ИНФОРМАЦИИ В СИНТЕЗИРУЕМОЙ НАУЧНО-ПРОИЗВОДСТВЕННОЙДОКУМЕНТАЦИИ

[1]Монахов М.Ю., [1]Семенова И.И.

[1]ФГБОУ ВПО «Владимирский государственный университет имени А.Г. и Н.Г. Столетовых», Владимир, e-mail: osobaii@gmail.com

В работе поднимается проблема оценки достоверности информации, на основе которой в системе синтезируется новая научно-техническая документация. Показано, что перспективным является путь интеллектуализации процесса, но в условиях множества дестабилизирующих факторов информация, поступающая извне и циркулирующая в системе, подвержена различным искажениям. В результате предлагается построить когнитивную модель, используя которую в сочетании со сценарным подходом возможно выявить закономерности влияния различных комбинаций дестабилизирующих факторов на достоверность информации в системе, причем на микро- и макроуровнях, а также на всех этапах преобразования информации.

Ключевые слова: автоматизация синтеза документации, когнитивное моделирование, оценка достоверности информации, дестабилизирующие факторы.

COGNITIVE MODEL OF ASSESSMENT THE LEVEL INFORMATION RELIABILITY OF SYNTHESIZING SCIENTIFIC-PRODUCTION DOCUMENTATION

[1]Monakhov M. Yu., [1]Semenova I.I.

[1]Vladimir State University named after A.G. and N.G. Stoletovs, Vladimir, e-mail: osobaii@gmail.com

The article raises the problem of assessing the reliability of information on the basis of which new scientific and technical documentation in a system is synthesized. It is shown that a promising way is intellectualization process. It is should be noted that because of very many destabilizing factors the information coming from the outside and circulating in the system is subject to many distortions. As a result it is proposed to build a cognitive model using which within scenario approach make it possible to reveal regularity of influence of various combinations of destabilizing factors on the reliability of the information in the system both on micro- and macrolevels, and at all stages of the information transformation.

Keywords: automated synthesis of documentation, cognitive modeling, evaluation of the information reliability, destabilizing factors.

Развитие современных систем автоматизации синтеза научно-производственной документации идет по пути интеллектуализации (Бурдо Г.Б., Венцов Н.Н., Головицина М.В., Евгенев Г.Б., Кондаков А.И., Курейчик В.М., Лебедев Б.К., Палюх Б.В., Тарасов В.Б. и др.). Анализ работ показал, что вопросам интеллектуального анализа данных с целью извлечения знаний из накопленных баз данных разработанной документации в совокупности уделяется мало внимания. Одним из зарекомендовавших себя методов является поиск ассоциативных правил, нашедший свое применение в других областях (Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И., Чубукова И.А. и др.).

Для повышения эффективности процесса создания технологической как одной из видов научно-производственной документации по уже сформированному комплекту конструкторской документации, с учетом специфики и потребностей методов интеллектуального анализа данных был получен комплекс решений [4, 5 и др.] по этапам, представленным на рис. 1. В соответствии с выработанными методиками в системе создается

база инженерных знаний, которая хранит накопленные устойчивые закономерности (правила), связывающие элементы конструкторской и технологической документации, а также учитывающие последовательности операций в них. Для создания такой базы знаний используется база данных, содержащая структурированный набор взаимосвязанных данных конструкторской и технологической документации. А для создания и наполнения такой базы данных необходим инструментарий по обработке накопленных на предприятиях комплектов технической документации в виде файлового архива, которыми пользуются технологи при создании новых комплектов документации с учетом факта разработки документации в разных системах.

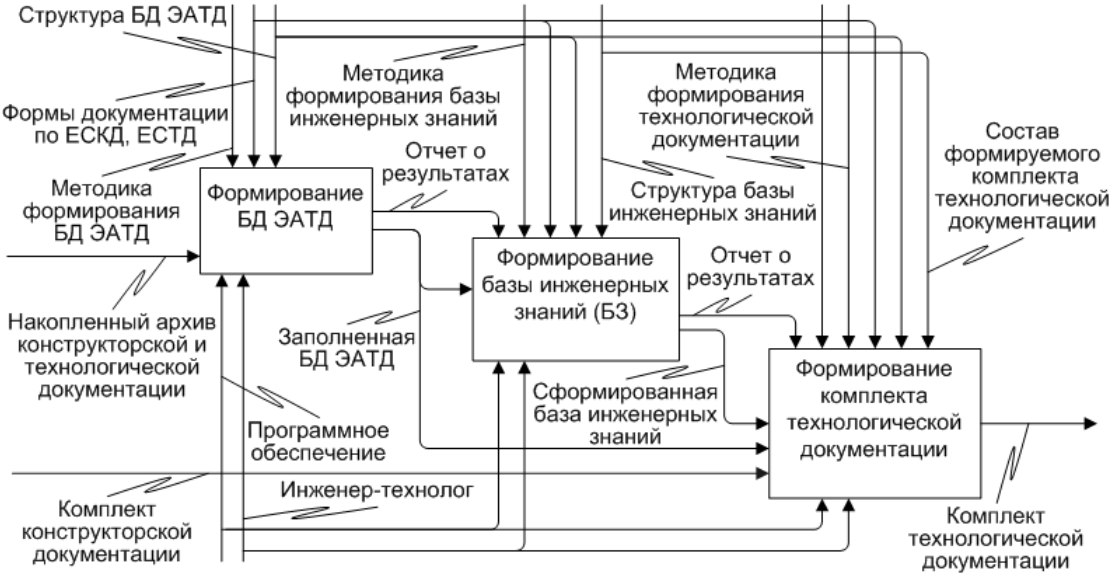

Рис. 1. Схема процесса автоматизации синтеза технологической документации (БД ЭАТД – база данных электронного архива технической документации)

При этом аспект достоверности циркулирующих в системе данных и выявленных знаний в работах не затрагивается, по умолчанию было принято решение, что формируемая база знаний достоверна и сформирована на достоверных источниках информации (ИстИ). Хотя практика показывает, что дестабилизирующие факторы, снижающие достоверность циркулирующей информации, имеют место и оказывают существенное влияние на процессы.

Таким образом, налицо потребность в комплексном подходе к исследованию и обеспечению достоверности данных в информационных системах (ИС) поддержки научно-производственных процессов при автоматизации синтеза научно-производственной документации.

В ИС циркулируют данные – обработанные сообщения, представленные в формализованном виде (в виде цифрового кода), пригодном для передачи и переработки в некотором информационном процессе для решения задач лицом, принимающим решения

(ЛПР). В них может присутствовать информация об объекте или явлении. Заметим, что ЛПР не является непосредственным наблюдателем события или явления (объекта), а должно использовать данные об объектах, которые получает от некоторого ИстИ, который является либо непосредственным «наблюдателем» явления, либо транслирует данные, получаемые от других источников, в лучшем случае, от первоисточников.

Информационные сообщения до времени «скрыты» в ИстИ, проявляются же в виде данных в момент инициирования задач ЛПР путем фиксации на физических носителях или при передаче по физическому каналу связи в ИС. Эти данные сохраняются, подвергаются обработке, представляются ЛПР, которое принимает информационное сообщение.

Данные, получаемые при кодировании сообщений, в силу различных обстоятельств могут оказаться правдивыми (либо правдоподобными, неправдоподобными), полными (либо недостаточно полными для задачи ЛПР), актуальными (либо устаревшими для решаемой задачи) и т.п. по многим качественным градациям. Степень доверия к таким данным определяется их «смысловой (семантической)» и временной искаженностью. Следовательно, необходимо говорить о достоверности ИстИ (точнее, о степени доверия потребителя к конкретному источнику), который искажает (неосознанно или с умыслом) формируемые им данные, делая информацию, содержащуюся в них, недостоверной.

Достоверность ИстИ – апостериорная оценка, полученная в результате наблюдения за его «информационной активностью». Источнику можно верить или не верить – он субъект, а значит, может лгать или говорить правду, может быть достоин веры (достоверен) или недоверия. Рассчитать достоверность информации (степень доверия) данного конкретного источника можно статистически как, например, процент совпадающих по смыслу сообщений от этого источника с сообщениями других (проверенных) источников в общем количестве его сообщений за некоторый промежуток времени (например, год).

Далее, в процессе передачи, хранения, обработки и представления данных могут также возникнуть их искажения (синтаксические и семантические), вносимые внешней средой (включая реализуемые угрозы целостности и доступности), ненадежными техническими и программными средствами, человеческими факторами и т.д. Каждое звено прохождения информации накладывает на нее свой (информационный) фильтр и вносит свои искажения. Природа таких искажений случайна. Отсюда мы можем говорить о соответствии данных объекту или явлению только в вероятностном смысле.

В итоге «достоверность информации», которую мы оцениваем, – это априорная оценка вероятности того, что сообщение для ЛПР при решении определенной задачи будет содержать неискаженные данные. Важным моментом в данном определении является «нужность» (важность или значимость) информации для ее потребителя (для решения задачи

ЛПР). Достоверность информации – это её свойство адекватно отражать действительность, которое проявляется в: аутентичности D_A – соответствии информации об объекте его действительному состоянию; полноте D_C – способности отражать все существенные в рамках поставленной задачи свойства объекта; актуальности D_T – способности отражать состояние объекта с задержкой, допустимой в поставленных условиях; целостности D_I – неизменности в процессах хранения, передачи и представления.

В связи с тем, что все эти показатели достоверности (ПД) информации характеризуют достоверность с разных сторон, её необходимо определять как множество оценок $\{D_A, D_C, D_T, D_I\}$ соответственно. На выделенные показатели влияет в свою очередь множество дестабилизирующих факторов (ДФ), под которыми будем понимать согласно [3] внутренние или внешние события в ИС, следствием которых является либо изменение, либо уничтожение информационных ресурсов (ИР), или процессов (ИП). Внутренние свойства ИС, определяющие возможность возникновения ДФ, будем называть структурно-функциональными недостатками (СФН) ИС. Анализ типовых ИС промышленных предприятий выявил наиболее значимые и распространенные классы ДФ и СФН ИС.

Обратимся к процессу автоматизации синтеза технологической документации и выделим модели отображения, которые, по сути, представляют основные этапы преобразования информационных ресурсов в ИС (рис. 2).

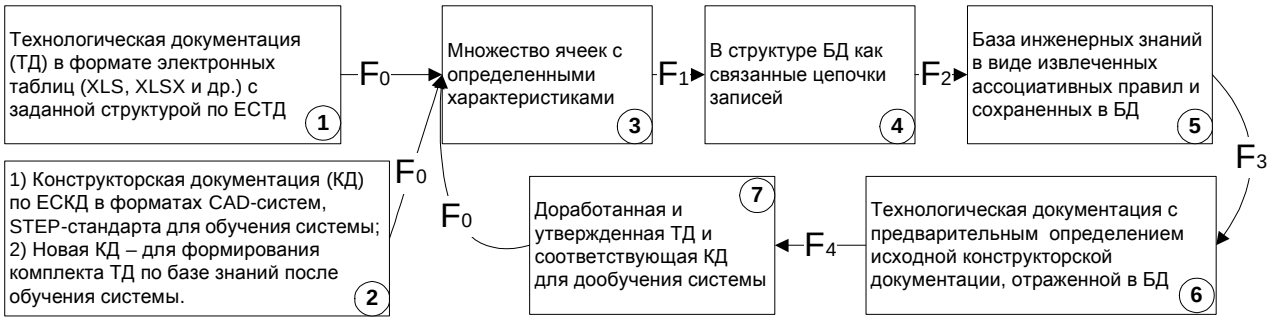

Рис. 2. Этапы преобразования информационных ресурсов на примере процесса автоматизации синтеза технологической документации

Модель отображения текстовой конструкторской и технологической документации в БД может быть представлена в виде суперпозиции отображений: $F_0 \circ F_1 : R_1 \to R_2$, $F_0 : \{(i_n, d, r, c)\} \to \{(k, j)\}$, $F_1 : \{(k, j)\} \to \{\tau, \alpha)\}$, где R_1 – неструктурированное множество данных, содержащихся в конструкторской и технологической документации; R_2 – структурированное множество атрибутов таблиц БД; *n* – порядковый номер документа в архиве; i_n – тип *n*-го документа; *d* – данные в *r*-й строке, *c*-м столбце таблицы *n*-го документа; $n = \overline{1, n_{R_1}}$; n_{R_1} – количество документов в архиве; $1 \le i_n \le 10$; $r = \overline{1, r_n}$, $c = \overline{1, c_n}$, r_n и c_n –

количество строк и столбцов в таблице *n*-го документа; $j=\overline{1,L_{ik}}$, $k=\overline{1,K_i}$; L_{ik} – количество атрибутов *k*-й группы записей в документе *i*-го типа; K_i – количество групп записей в документе *i*-го типа; τ – порядковый номер таблицы БД, α – порядковый номер атрибута τ-й таблицы; $\tau=\overline{1,\tau_{R_2}}$, τ_{R_2} – количество таблиц БД; $\alpha=\overline{1,a_\tau}$, α_τ – количество атрибутов в τ-й таблице.

Модель отображения структурированных данных из БД в базу инженерных знаний с ассоциативными правилами, отражающими устойчивые закономерности между фрагментами технологической и конструкторской документации, может быть представлена в виде отображения: $F_2 : R_2 \to R_3$, $F_2 : \{(\tau,\alpha)\} \to \{(x,y)\}$, где R_3 – структурированное множество правил в БЗ; τ – порядковый номер таблицы БД; α – порядковый номер атрибута τ-й таблицы; $\tau=\overline{1,\tau_{R_2}}$; τ_{R_2} – количество таблиц БД; $\alpha=\overline{1,a_\tau}$; α_τ – количество атрибутов в τ-й таблице, $\{(x,y)\}$ – множество извлеченных правил (инженерных знаний).

Модель отображения БД ЭАТД и базы инженерных знаний в комплект формируемой технологической документации может быть представлена в виде суперпозиции: $F_3 = F_3^{'} \circ F_3^{''}; F_3 : C \to D$, $F_3^{'} : C \to S_1 \cup S_2$, $F_3^{''} : S_1 \cup S_2 \to D$, где C – исходный комплект конструкторской документации; D – формируемый комплект технологической документации; S_1– структура БД ЭАТД для R_2; S_2 – структура базы инженерных знаний для хранения R_3.

Отображение F_4, по сути, представляет собой совокупность действий ЛПР, которое проводит проверку синтезированной документации и доработку в случае необходимости.

Было выделено и систематизировано множество ДФ, оказывающих влияние на уровень достоверности информации в разрезе D_A, D_C, D_T, D_I при реализации представленных на рис. 2 моделей отображения. Например, к ДФ, нарушающим аутентичность информации, отнесены: ошибки идентификации объекта (информация о другом объекте), ошибки идентификации свойства объекта (информация об ином свойстве объекта), ошибки измерения / оценки свойств объекта (измерительный / оценочный механизм не обладает требуемыми точностными характеристиками), ошибки преобразования (преобразование разнородных данных сопряжено с искажениями и нарушением эквивалентности), ошибки ЛПР и т.п.

Для построения когнитивной модели оценки уровня достоверности информации следует рассматривать семейство моделей по уровню обобщения информации в системе: на макроуровне в хранилище данных предприятия – это множество научно-производственных отчетов документированных и недокументированных на основе запросов и правил; далее по

мере детализации – это множество правил, которые можно получить путем применения методов извлечения знаний; множество запросов, которые можно получить при комбинировании единиц информации путем применения языков запросов; на микроуровне – это множество единиц информации, которые хранятся в одной или нескольких БД, хранилищах как неделимые единицы, извлеченные из первичных документов и/или введенные оператором, и/или импортированные средствами ETL-процессов. Данное четырехуровневое представление согласуется с этапами преобразования информационных ресурсов на рис. 2. Для оценки достоверности информации на более высоком уровне необходимо получение оценок достоверности с низшего уровня. Следовательно, имеем семейство взаимосвязанных разноуровневых моделей. В данной работе приведем пример модели микроуровня для отражения влияния аутентичности информации на уровень достоверности (рис. 3). Она может быть выражена в таких измеряемых показателях, как: вероятность безошибочного распознавания единиц информации P_{A1} из ИстИ в процессе реализации отображения F_0, вероятность неискажения информации во входных (в ИС) данных от конкретного источника информации P_{A2} или уверенность в том, что при вводе и передаче данных от ИстИ не допущено искажения информации в процессе отображения.

Используем подход к формированию четкой когнитивной карты [1] для получения представлений о характере зависимостей при внесении возмущений в систему взаимосвязанных параметров.

Представим когнитивную модель согласно [1] как $\Phi = < G, V, F >$, где $G = < X, E >$ – ориентированный граф; $X = \{x_i\}$ – множество вершин (концептов), $i = \overline{1, N}$; $E = \{e_{ij}\}$ – множество дуг, $j = \overline{1, N}$; $V = \{v^{x_i}\}$ – множество параметров вершин, $i = \overline{1, N}$; $F = F(V, E) = f(v_i, v_j, e_{ij})$ – функционал преобразования дуг, ставящий в соответствие каждой дуге весовой коэффициент w_{ij}. Учет возмущений (импульсов), изменяющих значения параметров в вершинах когнитивной модели, выполняется по правилу изменения значений вершин согласно подходу импульсного моделирования: $v_i(t+1) = v_i(t) + \sum_{j=1}^{N} w_{ji} \cdot p_j(t)$, где w_{ij} – вес связи между вершинами x_j и x_i (влияние x_j на x_i); $p_j(t) = v_j(t) - v_j(t-1)$ – величина изменения j-й вершины на шаге моделирования t.

Одним из ключевых вопросов является определение значений весов связей. Согласно [2 и др.], возможно использовать экспертные оценки, мягкие вычисления, теорию нечетких множеств, нейронные сети. Определим экспертным путем силы влияния связей (веса). Пусть w_{21}=0,75; w_{32}=0,5; w_{34}=0,5; w_{42}=0,5; $w_{41'}$ =0,1; $w_{42'}$ = –0,5; $w_{43'}$ = –0,5; $w_{44'}$ =0,25; w_{51}=0,25;

w_{61}=0,25; $w_{71'}$ =0,1; $w_{72'}$ = –0,25; $w_{73'}$ = –0,25; $w_{74'}$ =0,25; w_{17}= –0,25.

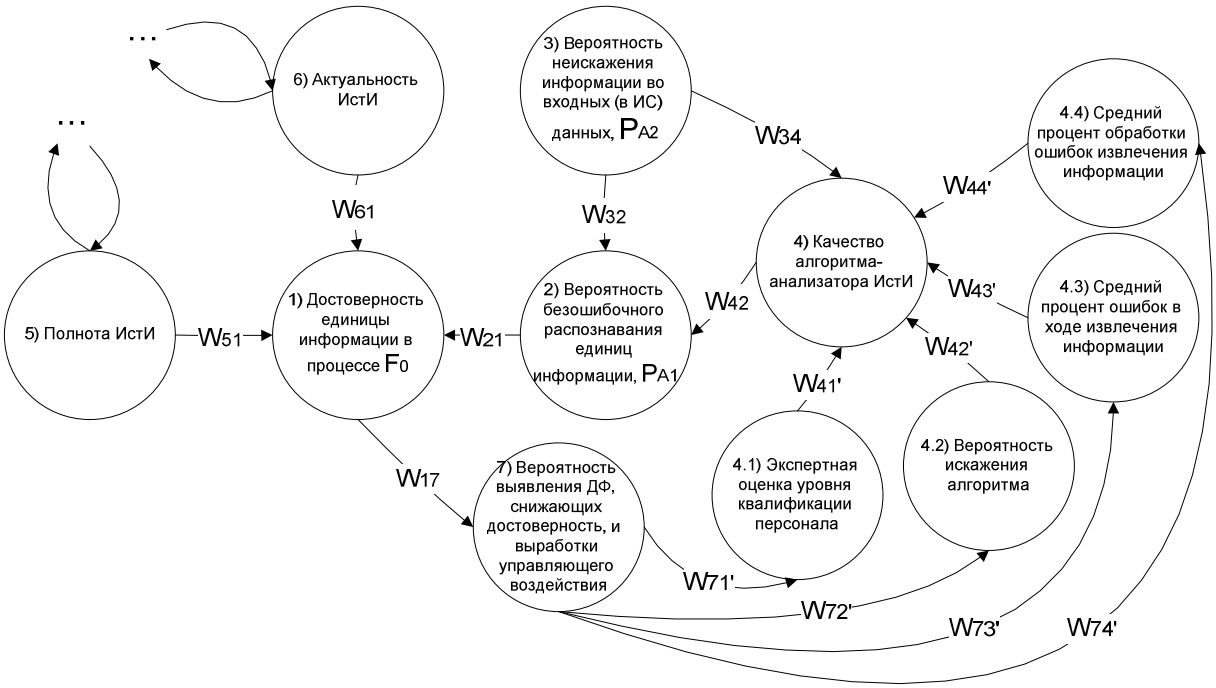

Рис. 3. Граф, отражающий связи между показателями, влияющими на достоверность единицы информации на примере процессов работы с научно-производственной документацией с обратным управляющим воздействием, где узлы 4.1–4.4 отражают влияние множества ДФ, а узел 7 – управляющее воздействие по нейтрализации ДФ

Анализ примера сценария (подача возмущающего воздействия «Повысился в результате некоторых мероприятий показатель среднего процента обработки ошибок извлечения информации на 5%»), приведенного на рис. 4 согласно модели на рис. 3, говорит о том, что возмущающее воздействие нейтрализуется на 9-й итерации, переводя систему в новое состояние, при этом показатель достоверности, таким образом, улучшается только на 0,5%. Реализация серии экспериментов с различными возмущающими воздействиями от узлов 4.1–4.4 позволит определить закономерности взаимовлияния параметров, отражающих уровень достоверности информации и ДФ, выявить приоритетные направления действий по устранению групп ДФ.

Развитие работы видится: в создании семейства когнитивных моделей всех уровней, описанных ранее, для выявления скрытых механизмов и закономерностей взаимовлияния параметров, отражающих уровень достоверности информации и дестабилизирующих факторов; в проработке методики проведения экспериментов по сбору статистики для уточнения значений сил влияния связей (весов). Кроме того, интересным направлением может стать сопоставление данных, полученных на когнитивных моделях, и известных эмпирических и аналитических оценках достоверности информации.

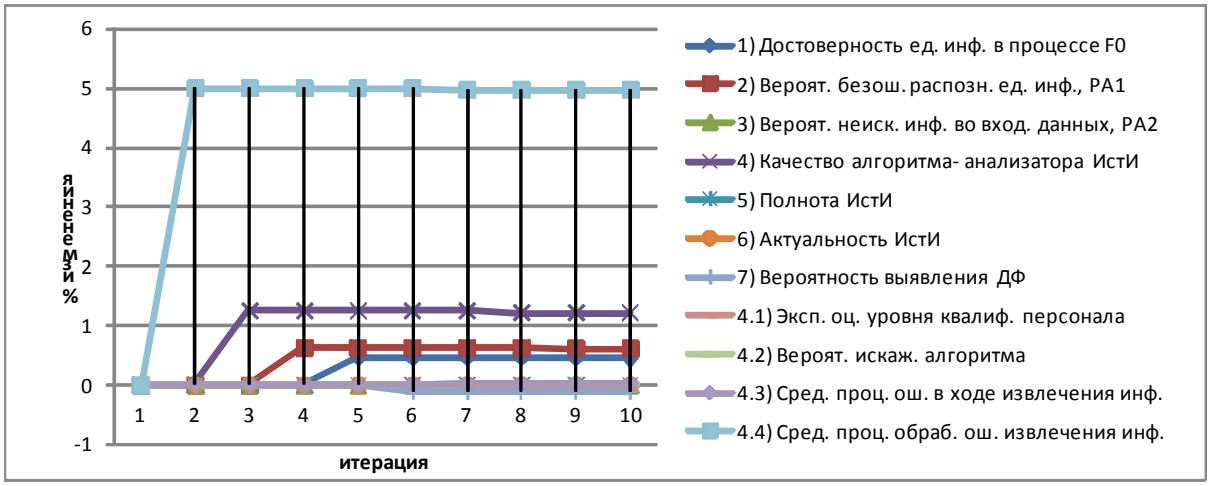

Рис. 4. Пример моделирования при подаче возмущающих воздействий

Качественные прогностические свойства проверенных когнитивных моделей позволят упростить сложный процесс анализа и выявления закономерностей во влиянии отдельных ДФ внутри системы на уровень достоверности информации в выходном продукте – синтезируемой научно-технической документации.

Работа выполнена при поддержке РФФИ № 13-07-97536 «Теоретические основы построения информационных систем обеспечения научно-производственных процессов с гарантированным уровнем достоверности информации синтезируемой технологической документации».

Список литературы

1. Асанов А.З., Мышкина И.Ю. Когнитивное моделирование при оценке сотрудников предприятий // Проблемы управления и моделирования в сложных системах : XIV Междунар. конф. – Самара: ФГБУН Самарский научный центр РАН, 2012. – С. 256–261. – Режим доступа: http://www.ssc.smr.ru/media/ipuss_conf/14/3_22.pdf.
2. Васильев В.И., Ильясов Б.Г. Интеллектуальные системы управления. Теория и практика: Учеб. пособие. – М.: Радиотехника, 2009. – 392 с.
3. Полянский Д.А. Методы контроля и обеспечения достоверности информации в АСУП: автореф. дис. ... канд. техн. наук: 05.13.06. – Владимир, 2010. – 15 с.
4. Семенова И.И., Толкачева Е.В. Автоматизация синтеза технологических решений и их документирования на основе извлечения инженерных знаний // Вестник Воронежского государственного технического университета. – 2011. – Т. 7. – № 4. – С. 76–80.
5. Семенова И.И., Толкачева Е.В. Автоматизация формирования электронного архива технической документации на базе файлового архива разработанных проектов// Вестник Воронежского государственного технического университета. – 2011. – Т. 7. – № 3.– С. 114–116.

Рецензенты:

Житников Б.Ю., д.т.н., профессор, профессор кафедры СТиИТ ВЮИ ФСИН России, г. Владимир.

Задорожный В.Н., д.т.н., доцент, профессор кафедры АСОИУ ОмГТУ, г. Омск.