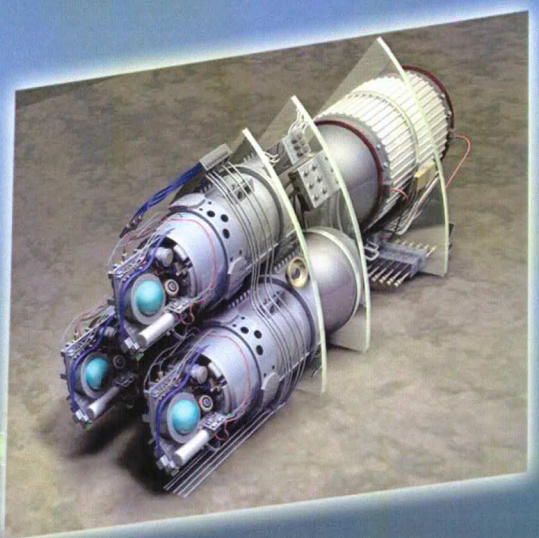
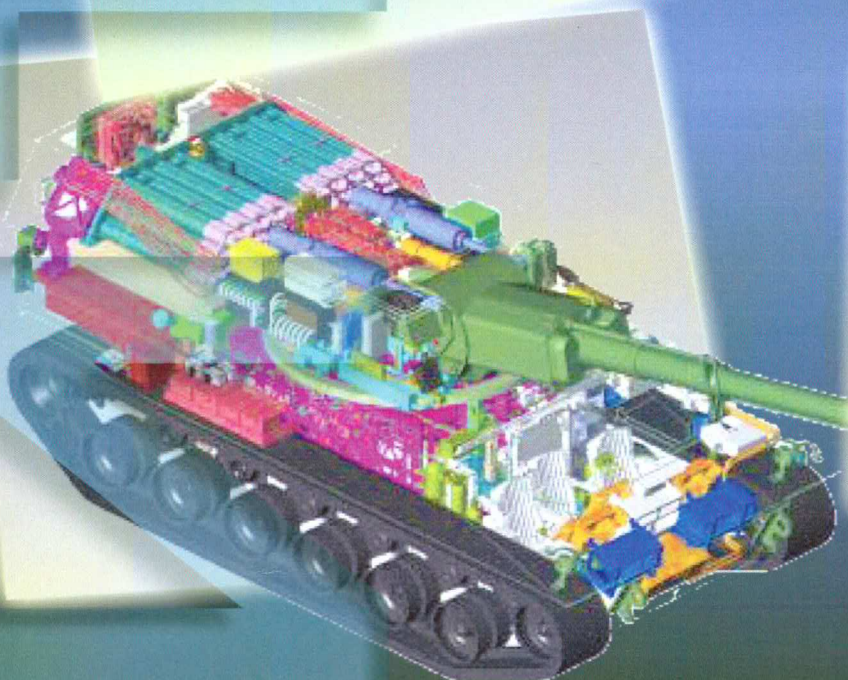
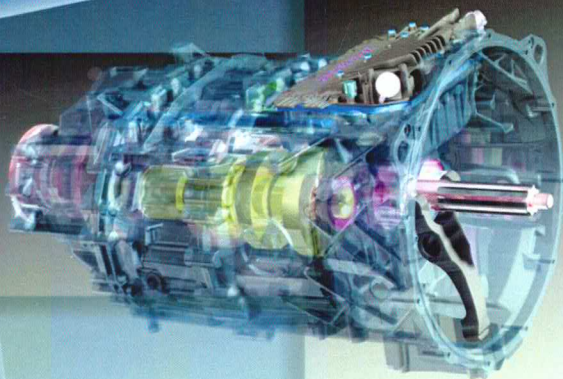
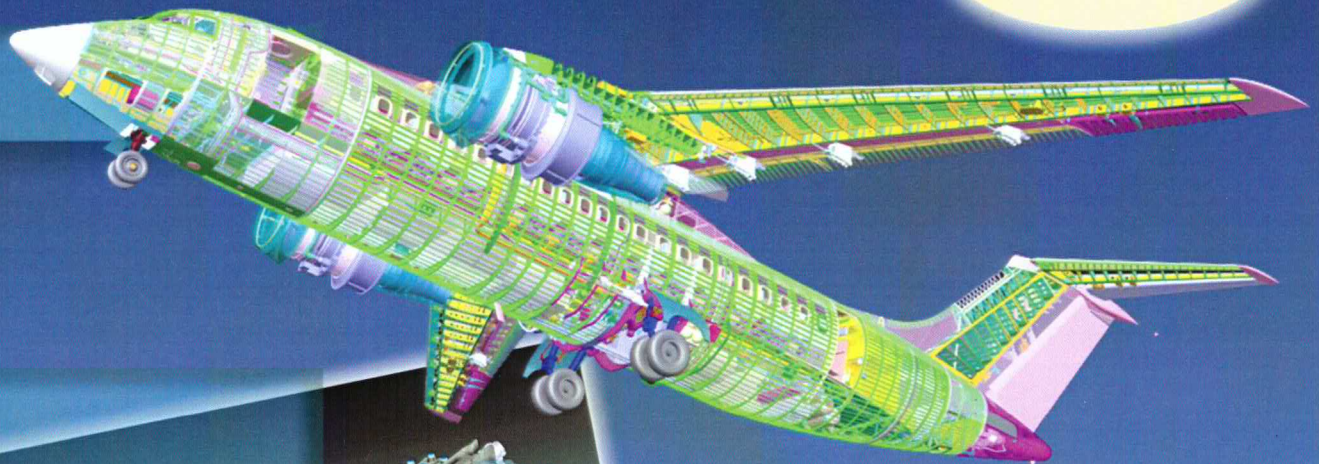


ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

в проектировании и производстве

№ 2' 2014



ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ПРОЕКТИРОВАНИИ И ПРОИЗВОДСТВЕ

Федеральное государственное унитарное предприятие "Всероссийский научно-исследовательский институт межотраслевой информации — федеральный информационно-аналитический центр оборонной промышленности"

№ 2 (154) 2014

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Основан в 1976 г.

Главный редактор

Д. А. Леманский, канд. техн. наук, начальник научно-образовательного центра ОАО «ГСКБ "Алмаз-Антей"»

Заместители главного редактора:

В. Н. Гридин, д-р техн. наук, проф., директор ЦИТИ РАН; А. И. Островерх, д-р техн. наук, проф., зам. генерального директора ГКНПЦ им. М. В. Хруничева

Ответственный секретарь

Б. С. Глотов

Редакционный совет:

И. В. Безденежных, канд. техн. наук, ФГУП "ВИМИ"; Е. И. Бронин, канд. техн. наук, ОАО «ГСКБ "Алмаз-Антей"»; А. И. Громов, канд. хим. наук, НИУ "Высшая школа экономики"; Ю. В. Давыдов, канд. техн. наук, ОАО "Туполев"; В. Ф. Евстафьев, д-р техн. наук, проф., ФГУП "ВИМИ"; В. Д. Костюков, канд. техн. наук, ГКНПЦ им. М. В. Хруничева; В. Н. Кофанов, д-р техн. наук, проф., ИИЭМ; Е. И. Митрушкин, д-р техн. наук, ОАО «НИИАА»; М. А. Пирогова, канд. техн. наук, НИУ МЭИ; Е. В. Судов, д-р техн. наук, НИЦ SALS-технологий "Прикладная логистика"; А. В. Цыркков, д-р техн. наук, проф., ГКНПЦ им. М. В. Хруничева; Е. И. Шульгин, д-р техн. наук, НИЦ при президиуме АИИ

☎ редакции (495) 491-77-67
E-mail: izdanie@vimi.ru, ivleva@vimi.ru
http://www.vimi.ru

Информационные технологии
в проектировании и производстве:
Науч.-техн. журн./ФГУП "ВИМИ", 2014.
№ 2. С. 1–76.

Редактор Г. А. Никитин
Заместитель редактора М. А. Николенико
Компьютерная верстка: Н. В. Ильина,
Т. А. Жамальдинова

Печатано в печать 14.05.2014.
Формат 60×84 1/8. Бумага офсетная.
Тираж печ. л. 8,8. Уч.-изд. л. 9,5.
Заказ 1815. Тираж 450 экз.
Адрес редакции: 125993, Москва,
Вильямовское ш., 77, ФГУП "ВИМИ".
Печатано в ООО «САМ ПОЛИГРАФИСТ».
125090, Москва, Протопоповский пер., д. 6.
Тел.: 79378. 11 статей.
Журнал зарегистрирован
в Роскомпечати 14.08.1998 г. Рег. № 018050.

Статьи рецензируются.

© Федеральное государственное унитарное предприятие "Всероссийский научно-исследовательский институт межотраслевой информации — федеральный информационно-аналитический центр оборонной промышленности" (ФГУП "ВИМИ"), 2014

СОДЕРЖАНИЕ

Инновационные технологии решения задач проектирования, моделирования и производства

- Бахвалов Ю. О., Костюков В. Д., Пугаченко С. Е., Шканов Д. А., Глотов Б. С.* Проблемы наукоемких технологий в отечественной ракетно-космической технике 3
- Гридин В. Н., Мазера Р. Б., Михайлов В. Ю.* Концепция, понятия и методы безопасного информационного взаимодействия 14
- Балабанов А. В.* Кодирование и оценка структур данных интерактивных систем объемного геометрического моделирования 22
- Дебелов В. А., Козлов Д. С.* Алгоритм фотореалистического рендеринга окрашенных полупрозрачных кристаллов 25
- Огородникова О. М.* О проблемах интеграции вычислительного материаловедения в цифровое машиностроение 30

Прикладные задачи применения информационных технологий

- Огнев И. В., Огнев А. И., Парамонов П. А.* Метод выделения речи на основе анализа распределения локальных экстремумов сигнала в системах автоматического распознавания 35
- Афанасьев В. В., Игнатенко А. В., Тисевич И. О.* Калибровка вторичного источника освещения по фотографиям 41
- Адамова А. А., Власов А. И.* Визуальное моделирование адаптации подготовки производства к выпуску новой продукции 46
- Митрушкин Е. И.* Инженерный анализ системы массового обслуживания 56
- Гусейнов А. Г.* Разработка методологической основы структурного моделирования ГПС с использованием базы знаний 63
- Клышинский Э. С., Калачёв Я. Б., Жаднов В. В.* К вопросу об автоматизации проверки полноты отчетной документации 68

следований структурной модели синтезированной ГПС обеспечивают требования заданных критериев ТЗ, то приступают к этапу рабочего проектирования, иначе корректируют необходимые параметры и исходные данные на предыдущих этапах.

Литература

1. Солнцев Р. И. и др. Автоматизация проектирования гибких производственных систем. — Л.: Машиностроение, 1990 г.

2. Норенков И. П., Арутюнян Н. М. Эволюционные методы в задачах выбора проектных решений // Электронный журнал "Наука и образование". 2007. № 9.

3. Симанженков К. А., Тясто С. А. Формализация структурного синтеза автоматизированных производственных систем в машиностроении // Вестник МГТУ «Станкин». 2011. № 4 (16).

4. Гусейнов А. Г. Компьютерная технология проектирования нестандартных механических модулей с применением базы знаний // Вестник ВГТУ. 2012. Т. 8. № 7-1. С. 38—42.

5. Алиев Р. А., Ахмедов М. А., Мамедов Дж. Ф., Гусейнов А. Г. Создание инструмента автоматизированного проектирования нестандартных элементов гибкой производственной системы. "Автоматизация и современная технология". — М.: Машиностроение, 2010. № 1.

DEVELOPMENT OF METHODOLOGICAL BASIS OF STRUCTURAL SIMULATION OF FMS WITH USING KNOWLEDGE BASE

A. G. Guseynov

Sumgait State University, Sumgait, Azerbaijan republic

In the paper development of methodical basis of structural simulation of flexible manufacture system (FMS) with application of intelligence units on the base of knowledge base is considered. It was done analyses of traditional methods and defined stages of works of executive.

The new architecture of the system of structural simulation of FMS is proffered. The developed architecture of the system supports creation of knowledge base on there bases levels.

Keywords: structural simulation, flexible manufactures system, knowledge base, computer experiments, structure synthesis.

Гусейнов Агил Гамид оглы, ученый секретарь СГУ.
E-mail: AQIL.55@mail.ru

Статья поступила в редакцию 25 февраля 2013 г.



К ВОПРОСУ ОБ АВТОМАТИЗАЦИИ ПРОВЕРКИ ПОЛНОТЫ ОТЧЕТНОЙ ДОКУМЕНТАЦИИ

Э. С. Клышинский, канд. техн. наук; Я. Б. Калачёв; В. В. Жаднов, канд. техн. наук
Московский институт электроники и математики национального исследовательского университета «Высшая школа экономики», Москва, Россия

Рассмотрен новый метод автоматизации определения соответствия технического задания и итогового отчета в ходе его приемки. Предложен метод предварительной оценки степени соответствия отчета техническому заданию, где использовано выделение значимых фрагментов технического задания, поиск соответствующих им элементов отчета и проверка степени покрытия последнего. Разработанный метод, например в отличие от косинусной меры сходства, дает лучшее разделение отчетов по критерию хорошего и плохого изложения материала.

Ключевые слова: информационные технологии, электронный документооборот, проверка документации, автоматическая обработка текстов.

Оформление корректной документации при проектировании изделий является залогом успешного выполнения проекта. Качественно написанное техническое задание (ТЗ) на проект и

проектная документация снижает шансы на срыв поставки изделия. Важную роль играет и отчет о выполненных работах, так как он позволяет повторить проделанные работы и (или)

разобраться в них. Особую роль документация играет при работе в соответствии с принципами ИПИ (CALS)-технологии [1].

Существует несколько стандартов на оформление технической документации. В первую очередь это ГОСТы, по которым оформляется документация в государственных учреждениях, на промышленных предприятиях и т. д. В качестве альтернативы им можно привести стандарты International Standard Organization, которые дают рекомендации по составу документации. В области электротехники и телекоммуникаций — серия рекомендаций European Telecommunications Standards Institute, в области разработки программного обеспечения можно использовать рекомендации Rational Unified Process от IBM [2] и Microsoft Solutions Framework [3]. Хотя они не утверждены в качестве государственных стандартов, договор может определить работу в соответствии с этими рекомендациями или по стандартам предприятий, описывающих состав и содержание документации.

Проверка текста отчета обычно осуществляется экспертами и занимает много времени. Недобросовестный исполнитель может пытаться спрятать низкое качество отчета за его объемом, цитатами, уходом в смежную область, применением бюрократического стиля изложения текста и др. Для создания "первого эшелона обороны" в помощь экспертам необходимо создать автоматизированную систему, проверяющую степень соответствия отчетов или технической документации тексту ТЗ и помогающую принимать решения о проведении детальной экспертизы документации. Подобная система выявит наиболее очевидные несоответствия, которые должны быть проверены специалистом. На вход система должна получать тексты документов, содержащих постановку задачи и требования к проекту, а также отчетные документы. Выходом системы является оценка степени сходства и связности фрагментов текста, их смысловой нагруженности.

Для определения близости документов или их фрагментов используется, например, метод шинглирования, который основывается на выделении последовательностей слов длины k с последующей оценкой вероятности совпадения документов [4–6]. Данный метод обладает высокой скоростью, но в данном случае может быть применен лишь для выявления заимствованных фрагментов отчетов. Для сравнения текстов ТЗ и отчетов больше подходят методы определения тематической близости документов. Пусть для двух документов вычислены вектора частот встречаемости слов \mathbf{a} и \mathbf{b} . Вектора определены на множестве всех слов, встречающихся в обоих документах. В этом случае косинусная мера сходства двух документов определяется следующим образом [7]

$$\cos(\mathbf{a}, \mathbf{b}) = \frac{\sum a_i \cdot b_i}{\sqrt{\sum a_i \cdot b_i} \cdot \sqrt{\sum b_i \cdot b_i}}.$$

Развитием этого метода является использование векторов частот не отдельных слов, а их сочетаний различной длины. В этом случае точность определения степени сходства документов возрастает [8]. Косинусная мера сходства документов хорошо зарекомендовала себя при решении различных задач, но определяет лишь сходство тематики документов.

К настоящему времени хорошо проработаны методы хранения технической документации с использованием PDM/PLM-систем [9]. Также существуют методики, учитывающие специфику обработки технических документов [10]. Современные достижения в области ИПИ-технологий дают основу для создания системы контроля качества документации, но не позволяют решить задачу определения полноты документации.

Еще одним направлением, в котором проводятся исследования, является разработка формальных моделей текста документов [11], но для создания подобных систем необходимы большие онтологии предметных областей, не всегда доступные разработчикам. В настоящее время проводятся работы по автоматизированному составлению онтологий [12], но и эти работы еще не доведены до программного обеспечения, удобного для применения. Наиболее разработанными в теоретическом и практическом плане являются работы в области формализации выделения спецификаций систем [13], однако и они далеки от широкого распространения.

Исходя из вышеизложенного можно сделать вывод о том, что существует потребность в создании теоретического метода автоматизации определения соответствия технического задания и итогового отчета и его практической реализации. Кратко изложим основные этапы такого метода.

Из текста ТЗ выделяются предложения, содержащие характеристики разрабатываемого изделия. Из выделенных фрагментов извлекаются все пары стоящих рядом слов (коллокации), незначимые коллокации удаляются.

Текст отчета разбивается на фрагменты, для которых также выделяется список коллокаций, после чего находится максимум меры сходства абзацев со значимыми фрагментами ТЗ. Это значение и будет мерой соответствия абзаца тексту технического задания. Полученный результат выдается лицу, принимающему решение (ЛПР), в визуальной форме. С его помощью ЛПР определяет меру соответствия отчета ТЗ, а также определяет необходимость дальнейшего анализа текста отчета или его фрагментов.

Рассмотрим каждый из этапов более подробно.

Представим текст как упорядоченное множество предложений: $\mathbf{t} = \langle s_i \rangle$. Представим предложение как упорядоченное множество слов: $s_i = \langle w_{ij} \rangle$. Под словосочетанием будем понимать упорядоченное множество слов: $\mathbf{c} = \langle w \rangle$. Будем говорить, что словосочетание входит в i -е пред-

ложение ($c \in s_i$), если в s_i существует контактное подмножество, эквивалентное c .

Пусть $K = \{c\}$ — список ключевых коллокаций, вводящих требования к изделию, поставленные заказчиком (например, "должно обладать/состоять/ ...", "обеспечивает", "служит"). Тогда предложение s , входящее в текст ТЗ, называется значимым, если $\exists c \in K: c \subset s$. Значимое предложение входит в значимый фрагмент, содержащий в себе одно или несколько предложений

$$f = \langle t, s, e \rangle,$$

где t — текст, в который входит фрагмент, s — номер начального предложения фрагмента, e — номер последнего предложения фрагмента.

По результатам анализа текстов ТЗ были разработаны следующие правила, определяющие границы значимых фрагментов ТЗ:

если ключевая фраза встречается в предложении, после которого идет перечисление, то выделяется и весь текст до конца перечисления (например, "система должна состоять из следующих подсистем: ...");

если ключевая фраза встречается в предложении, находящемся в связанном тексте, то выделяется предложение целиком;

если фраза встречается отдельно (например, заголовок "необходимо"), то выделяется весь следующий абзац.

Эксперименты показали, что качество работы метода возрастает, если помимо значимого предложения в фрагмент включается одно предложение до и два предложения после значимого, так как они чаще всего связаны по смыслу. Предыдущее предложение часто вводит некоторое определение или определяет общее направление, последующие расшифровывают требования.

На первом шаге метода по тексту ТЗ ищутся ключевые фразы, к которым применяются приведенные выше правила. Если условие выполняется, выделяется очередной значимый фрагмент, который заносится в список $F = \{f\}$. Два значимых фрагмента могут быть объединены вместе, если их границы пересекаются или между ними нет значимого текста: если $f_m = \langle t, s_1, e_1 \rangle$ и $f_{m+1} = \langle t, s_2, e_2 \rangle$: $e_1 \geq s_2$, то $f_m = \langle t, s_1, e_2 \rangle$, а f_{m+1} удаляется.

На втором шаге проводится выделение коллокаций из значимых фрагментов. Для значимых фрагментов рассчитывается вектор признаков

$$a = \langle \{w, f\} \rangle,$$

где $w \in f$ — коллокация, а f — ее частота встречаемости.

Из вектора признаков отсеиваются коллокации с частотами выше 0,75 и ниже 0,25 от максимальной. Это позволяет избавиться от служебных слов и авторских особенностей текста, отсеять редко встречающиеся сочетания. В итоге будет сформировано множество векторов признаков ТЗ $S_1 = \{a\}$.

На третьем шаге текст отчета разбивается на абзацы, для которых формируется список коллокаций с частотами их встречаемости (вектор признаков b : $S_2 = \{b\}$). Значимость абзаца с номером j вычисляется как максимум косинусной меры сходства вектора b с векторами a ТЗ или равна нулю, если найденная значимость ниже заданного порога

$$v_j = \max_i \cos(a_i, b_j),$$

где: $a_i \in S_1$, а $b_j \in S_2$.

На четвертом шаге ЛПП получает информацию о покрытии отчета фрагментами ТЗ в виде точечной диаграммы. Так как в работе метода возможны ошибки при выделении свойства или значимого фрагмента, эксперт может получить более подробную информацию о фрагментах отчета и ТЗ: соответствие значимых фрагментов ТЗ, список коллокаций и т. д.

По текстам ТЗ и отчетов формировались точечные диаграммы. На них точка соответствует 100, а строка — 10 000 символам текста. Черные точки показывают части текста, содержащие ключевые слова.

Проверка метода проводилась в два этапа. На первом этапе экспертам давали ознакомиться с содержанием ТЗ и отчета и высказать свое мнение относительно их содержания. После этого документы проверялись при помощи разработанного метода. На втором этапе проводилась кросс-проверка документации. Все ТЗ проверялись со всеми отчетами, чтобы проверить гипотезу о том, что максимум совпадения для качественно написанных отчетов должен находиться на соответствующих им ТЗ.

На рис. 1, а представлена диаграмма разбора ТЗ, содержащего ненужную информацию.

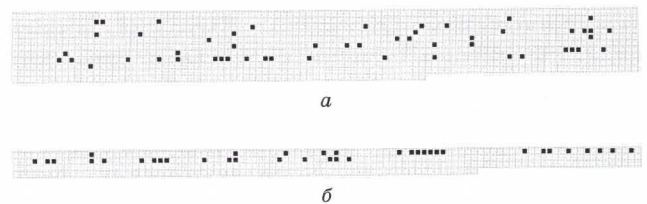


Рис. 1. Визуализация результатов анализа неудачного (а) и удачного (б) ТЗ

В нем большая часть текста говорит о составе и планах организации, проводимой ею научной работе. ТЗ, соответствующее рис. 1, б, написано в строгом стиле и по требованиям ГОСТа. Ключевые предложения найдены в середине текста в разделе, описывающем требования к изделию. Заключительная часть ТЗ относится к срокам разработки, требованиям к рабочим местам, требованиям к интерфейсу. Хотя число ключевых фрагментов в первом и втором случае почти одинаково, второй текст выигрывает из-за сжатости текста и точно поставленных требований.

На рис. 2, а показана диаграмма для отчета, полного "воды".

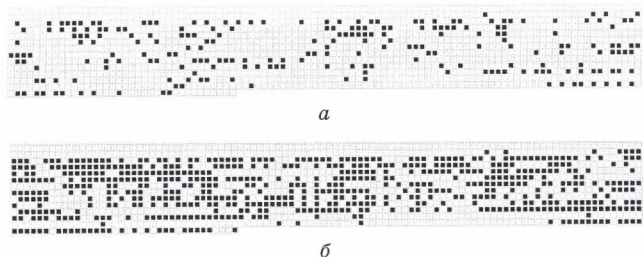


Рис. 2. Точечная диаграмма для неудачного (а) и качественного (б) отчета

Блоки из компактно расположенных 5—10 черных точек описывают заявленные в ТЗ требования. Отдельно стоящие черные квадраты соответствуют единичным коллокациям (например, в заголовке). Здесь на более чем 130 000 знаков отчета было найдено лишь 470 коллокаций, относящихся к ТЗ (считая единичные вхождения в заголовках). Максимальная длина связного текста, имеющего отношение к одному из значимых фрагментов ТЗ, — 700 символов.

На рис. 2, б представлен качественно написанный отчет, в котором ключевые коллокации встречаются везде, за исключением начала (содержание, авторы, введение) и конца отчета (юридический и экономический разделы). При длине отчета свыше 130 000 знаков найдено более 3500 коллокаций. Максимальная длина текста, имеющего значимые фрагменты ТЗ, — 1500 знаков.

Для кросс-проверки были использованы тексты 6 ТЗ и 9 отчетов. Отчетам присвоен номер соответствующих ТЗ. ТЗ с номерами 1—3 написаны по одной тематике, отчеты 5 и 6 написаны по близким тематикам. Отчет с номером 0 не связан ни с одним из ТЗ. Отчеты 3 и 6 представлены в двух версиях. Вторая версия отчетов, отмеченная знаком "+", содержит исправления найденных заказчиком недостатков.

Результаты проверки показаны в табл. 1. Соответствия ТЗ и отчетов выделены рамкой. Результаты удачных проверок выделены темным фоном. Успешные проверки с другими отчетами показаны светлым фоном.

Таблица 1

Результаты кросс-проверки для предложенного метода

| Отчеты | Технические задания | | | | | |
|--------|---------------------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0,521 | 0,157 | 0,192 | 0,032 | 0,025 | 0,072 |
| 2 | 0,394 | 0,592 | 0,543 | 0,056 | 0,054 | 0,062 |
| 3 | 0,37 | 0,39 | 0,158 | 0,05 | 0,049 | 0,05 |
| 3+ | 0,494 | 0,45 | 0,535 | 0,045 | 0,051 | 0,054 |
| 4 | 0,032 | 0,032 | 0,066 | 0,032 | 0,002 | 0,031 |
| 5 | 0,032 | 0,009 | 0,02 | 0,307 | 0,057 | 0,095 |
| 6 | 0,006 | 0,011 | 0,007 | 0,002 | 0,031 | 0,638 |
| 6+ | 0,006 | 0,009 | 0,006 | 0,002 | 0,016 | 0,725 |
| 0 | 0,011 | 0,043 | 0,035 | 0,006 | 0,006 | 0,017 |

Как видно из табл. 1, разработанный метод и программное обеспечение определило высокое

качество отчетов, написанных для технических заданий 1—3 и 6. При этом результат работы системы для отчета 3 и 6 совпал с мнением заказчика. Отчет 0 не показал совпадений ни для одного из ТЗ.

ТЗ 4 и 5 не предполагали подробного описания результатов работы и требований к ним. Также в ТЗ 5 требовалось дать рекомендации по улучшению изделия, что усложнило поиск соответствия. Отчет 4 содержал информацию по предметной области ТЗ 5, в связи с чем их сходство выше.

Для проверки метода были вычислены значения косинусной меры сходства между ТЗ и отчетами. Для этого использовались частоты встречаемости отдельных слов, отсеивание слов не проводилось. Результаты кросс-проверки для косинусной меры сходства сведены в табл. 2.

Таблица 2

Результаты кросс-проверки для косинусной меры сходства

| Отчеты | Технические задания | | | | | |
|--------|---------------------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0,533 | 0,546 | 0,546 | 0,198 | 0,263 | 0,292 |
| 2 | 0,532 | 0,554 | 0,880 | 0,151 | 0,169 | 0,162 |
| 3 | 0,554 | 0,779 | 0,579 | 0,183 | 0,205 | 0,191 |
| 3+ | 0,534 | 0,763 | 0,612 | 0,191 | 0,204 | 0,192 |
| 4 | 0,331 | 0,238 | 0,326 | 0,189 | 0,212 | 0,167 |
| 5 | 0,418 | 0,302 | 0,331 | 0,182 | 0,317 | 0,243 |
| 6 | 0,161 | 0,091 | 0,116 | 0,089 | 0,091 | 0,443 |
| 6+ | 0,163 | 0,091 | 0,117 | 0,089 | 0,091 | 0,443 |
| 0 | 0,257 | 0,174 | 0,207 | 0,139 | 0,175 | 0,185 |

Как видно из табл. 2, косинусная мера успешно определяет тексты с общей тематикой, но не показывает качество отчетной документации: максимальные значения достигаются на ТЗ, не соответствующих данному отчету; разделимость хороших и плохих отчетов отсутствует.

Проверка результатов показала повышение точности работы метода по сравнению с разработанными ранее. Проблему представляют ТЗ, описывающие лишь основные цели работы. Кроме того, даже при соответствии 70 % и выше, детальная экспертиза необходима, так как метод не гарантирует полностью достоверных результатов. Для решения этой задачи необходимо применять специализированные методы (например, для экспертизы отчета по надежности электронных средств — методику [14]).

Тем не менее, метод может применяться как часть автоматизированной системы ведения и хранения документации по проекту и помогать в принятии решений о доработке отчета или о его детальной экспертизе.

Работа выполнена при поддержке РГНФ
(грант № 12-04-00060) и РФФИ
(грант № 11-01-00793).

Литература

1. Яблочников Е. И., Молочник В. И., Миронов А. А. ИПИ-технологии в приборостроении: Учебное пособие. — СПб: СПбГУ ИТМО, 2008. — 128 с.
2. Кролл П., Крачтен Ф. Rational Unified Process — это легко. Руководство по RUP для практиков. Пер. с англ. — М.: КУДИЦ-ОБРАЗ, 2004. — 432 с.
3. Тернер М. Основы Microsoft Solutions Framework. — СПб.: Питер, 2008.
4. Клышинский Э. С., Антонова А. Ю. Об использовании мер сходства при анализе документов / Сб. тр. 13 Всерос. науч. конф. RCDL'2011. С. 246—250.
5. Broder S. Glassman, Manasse M. and Zweig G. Syntactic clustering of the Web. / Proc. of the 6th International World Wide Web Conference, April 1997.
6. Зеленков Ю. Г., Сегалович И. В. Сравнительный анализ методов определения нечетких дубликатов для WEB-документов / Сб. трудов 9 Всерос. науч. конф. RCDL'2007.
7. Маннинг К. Д., Рагхаван П., Шютце Х. Введение в информационный поиск. — М.: Вильямс, 2011. — 528 с.
8. Клышинский Э. С. Анализ комплексных мер тематического сходства документов / Научно-техническая информация. Сер. 2: Информационные системы и процессы. 2011. № 9. С. 6—11.
9. Колчин А. Что такое PDM? / PC Week. 2001. № 38.
10. Черников Б. В. Технологии подготовки документов на основе кибернетических методов. — М.: Финансы и статистика, 2009. — 206 с.
11. Тарасенко А. В. Разработка и исследование методов и моделей автоматической проверки текстов на соответствие требованиям технической документации: Автореферат диссертации на соискание ученой степени доктора технических наук по специальности 05.13.17. — г. Таганрог, 2009.
12. Волкова Г. А. Создание "онтологии всего". Проблемы классификации и решения. / Сб. трудов науч.-практич. семинара "Новые информационные технологии в автоматизированных системах". — М.: МИЭМ НИУ ВШЭ, 2013. С. 293—300.
13. Заболеева-Зотова А. В., Орлова Ю. А. Автоматизация процедур семантического анализа текста технического задания / Известия Волгоградского гос. технического университета. 2007. Т. 9. № 3. С. 52—55.
14. Жаднов В. В. Методические указания по проведению экспертизы конструкторского документа РР01 "Расчет надежности" для электронных модулей первого уровня с использованием технологии прогнозирования надежности АСОНИКА® — М.: МИЭМ НИУ ВШЭ, 2012. — 16 с.

ON AUTOMATIZATION OF REPORT DOCUMENTATION COMPLETENESS ASSESSMENT

E. S. Klyshinsky, Ya. B. Kalachyov, V. V. Zhadnov

Moscow Institute of Electronics and Mathematics of National Research University
"Higher School of Economics", Moscow, Russia

The paper introduces a new method of technical reports verification by finding the correspondence between an agreed technical statement and a submitted report during its acceptance process. The method makes a preliminary evaluation of such correspondence. For these purposes we select significant fragments of technical statement, mark the coinciding parts of report and evaluate the percentage of the marked text. Instead of the cosine similarity the proposed method allows to separate reports with correct and wrong description of material.

Keywords: information technology, electronic document flow, documentation completeness assessment, natural language processing.

Клышинский Эдуард Станиславович, доцент кафедры информационных технологий и автоматизированных систем.

E-mail: eklyshinsky@hse.ru

Калачёв Ярослав Борисович, аспирант кафедры информационных технологий и автоматизированных систем.

E-mail: Kalachyov-YB@sac.minenergo.gov.ru

Жаднов Валерий Владимирович, доцент, профессор кафедры радиоэлектроники и телекоммуникаций.

Тел. 8 (495) 916-88-80. E-mail: vzhadnov@hse.ru

Статья поступила в редакцию 11 января 2014 г.

