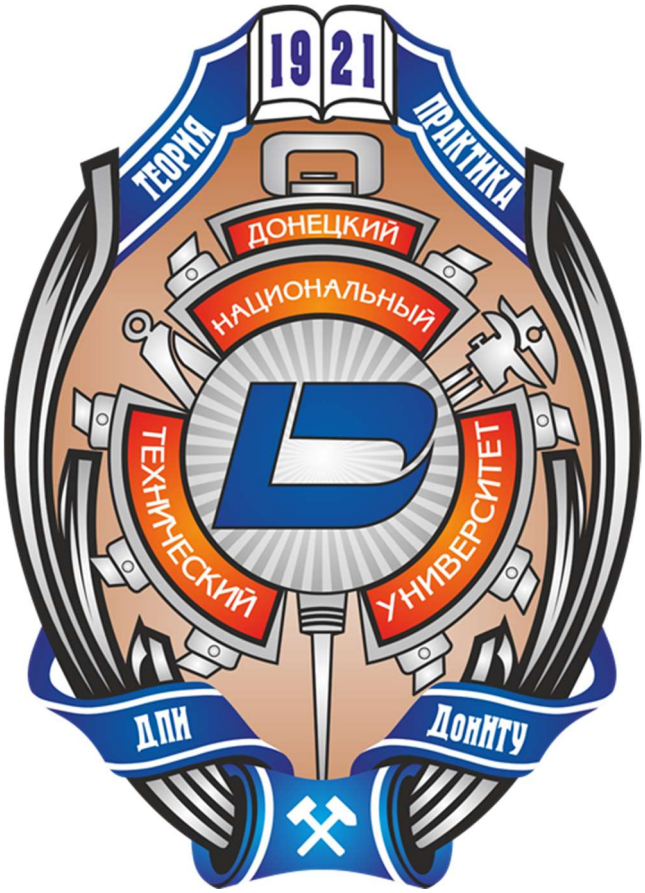
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
ДОНЕЦКОЙ НАРОДНОЙ РЕСПУБЛИКИ



ИНФОРМАТИКА И КИБЕРНЕТИКА

4 (22)

Донецк – 2020

УДК 004.3+004.9+004.2+51.7+519.6+519.7

ИНФОРМАТИКА И КИБЕРНЕТИКА, № 4 (22), 2020,
Донецк, ДонНТУ.

Представлены материалы по вопросам приоритетных направлений научно-технического обеспечения в области информатики, кибернетики и вычислительной техники.

Материалы предназначены для специалистов народного хозяйства, ученых, преподавателей, аспирантов и студентов высших учебных заведений.

Рекомендовано к печати ученым советом ГОУ ВПО «Донецкий национальный технический университет» Министерства образования и науки ДНР. Протокол № 6 от 25 декабря 2020 г.

Свидетельство о регистрации СМИ: серия ААА № 000145 от 20.06.2017.

Приказ МОН ДНР № 135 от 01.02.2019 о включении в Перечень рецензируемых научных изданий ВАК ДНР.

Контактный адрес редакции
ДНР, 83001, г. Донецк, ул. Артема, 58, ГОУ ВПО «ДонНТУ»,
4-й учебный корпус, к. 36., ул. Кобозева, 17.
Тел.: +38 (062) 301-07-35, +38 (071) 334-89-11
Эл. почта: infcyb.donntu@yandex.ru
Интернет: http://infcyb.donntu.org

СОДЕРЖАНИЕ

Информатика и вычислительная техника

Инженерное образование

Информатика и вычислительная техника

УДК 004.93:681.5.012:004.62(043.5)

Исследование алгоритма адаптивного нелинейного оптимального сглаживания многопараметрических данных измерений

И. Л. Щербов
Донецкий национальный технический университет, г. Донецк
scherbov@yandex.ru

Аннотация
Исследование алгоритмов нелинейного оптимального адаптивного сглаживания многопараметрических данных измерений позволяет сделать вывод, что алгоритм, оптимизирующий структуру сглаживающего полинома, более эффективен, чем алгоритм оптимизирующий степень сглаживающего полинома. Средние показатели качества и эффективности, при применении алгоритма оптимизирующего структуру сглаживающего полинома, практически сохраняют свои величины при уменьшении среднеквадратичных отклонений измерительных средств. С целью обеспечения высокой эффективности и качества сглаживания длительность интервала сглаживания необходимо выбирать в 1,5 и более раз превышающей время корреляции ошибок измерений.

Введение

Развитие средств получения и обработки информации сложных технических систем перемещающихся в пространстве по сложным траекториям, придает особую актуальность вопросам повышения оперативности метрологического контроля обработки траекторной информации, достоверности оценки ее точности, сокращению сроков обработки. Наиболее перспективным подходом к решению задачи является автоматизация и управление таким технологическим процессом на основе созданного математического, информационного и алгоритмического обеспечения.

По полученным данным с помощью аппаратуры информационно-измерительных комплексов проводится анализ траектории движения летательных аппаратов при проведении послеполётной обработке поступившей информации.

Данный анализ необходим для точной и достоверной оценке информации о траектории летательного аппарата с целью предупреждения принятия неправильных решений о характеристиках испытуемого объекта или качестве его бортовых навигационных систем.

Таким образом, повышение точности и достоверности определения параметров пространственного положения летательных аппаратов за счет создания информационной технологии процесса контроля совместной обработки избыточных данных траекторных измерений является актуальной научно-технической задачей и имеет практическое значение.

С цель решения вышеуказанной задачи в работах [1, 2, 3, 4] проанализирована работа алгоритма нелинейного оптимального адаптивного сглаживания многопараметрических данных измерений. Рассмотрен метод последовательных приближений для решения задачи по определению максимально правдоподобной оценки вектора измерений. Также рассмотрена система базисных функций и вектор коэффициентов сглаживающего полинома и произведен сравнительный анализ двух клеточно-матричных структур базисных функций для осуществления сглаживания путем совместной обработки данных внешнетраекторных измерений. Предложено решение задачи о нахождении начального приближения вектора коэффициентов сглаживающего полинома.

Цель и задачи исследования

Разработка метода, составляющего математическое описание технологии совместной обработки избыточной траекторной информации и проведение анализа устойчивости аппаратно-программного комплекса послеполетной обработки траекторной информации к аномальным ошибкам данных.

Решение данных вопросов позволит получить независимые оценки коэффициентов сглаживающего полинома, обосновать выбор критерия проверки статистических гипотез и осуществить оценку устойчивости аппаратно-программного комплекса к аномальным ошибкам обрабатываемых данных.

Основной материал исследования

Для объективного исследования алгоритма нелинейного оптимального адаптивного сглаживания многопараметрических данных измерений необходимо знать результаты сглаживания при точно известной степени (структуре) сглаживающего полинома,

осуществляемое методом наименьших квадратов (МНК) или методом максимального правдоподобия (ММП), которое приводит к наилучшим (по точности) результатам сглаживания [1, 5]. Назовем такое сглаживание идеальным сглаживанием. Оно возможно только при моделировании, так как стохастический характер реальных траекторий трудно совместим с их высокой априорной определённостью.

При выборе показателей качества и эффективности необходимо исходить из задач и целей, преследуемых при сглаживании [2, 1].

Рассмотрим эти показатели:

Выигрыш в точности в i-ой точке j-траектории рассчитывается по формуле:

$$W_{ij} = \frac{\hat{\sigma}_{\xi_{ij}}}{\hat{\sigma}_{\tilde{\xi}_{ij}}}, \quad (1)$$

которая представляет собой отношение среднеквадратической ошибки в *i*-ой точке *j*-траектории до и после сглаживания.

Данное отношение показывает, во сколько точность после сглаживания превышает в *i*-ой точке *j*-траектории точность до сглаживания. Средний выигрыш в точности по *j*-траектории определяется по формуле:

$$W_{j_{\text{ср}}} = \frac{1}{n}\sum_{i=1}^{n} W_{ij}, \quad (2)$$

где *n* – число точек на *j*-траектории (интервале измерения).

Средний выигрыш в точности в i-ой точке по N траекториям определяем:

$$W_{i_{\text{ср}}} = \frac{1}{N}\sum_{i=1}^{N} W_{ij}, \quad (3)$$

где *N* – число обрабатываемых траекторий.

Средний выигрыш в точности по *N* траекториям, каждая из которых содержит *n* точек:

$$W_{\text{ср}} = \frac{1}{n}\frac{1}{N}\sum_{i=1}^{n}\sum_{i=1}^{N} W_{ij} = \frac{1}{N}\sum_{i=1}^{N} W_{j_{\text{ср}}}. \quad (4)$$

Показатель эффективности в i-ой точке j-траектории μ_i – отношение приращения точности в i-ой точке j-траектории при сглаживании исследуемым методом к приращению точности при идеальном сглаживании:

$$\mu_{ij} = \frac{\hat{\sigma}_{\xi_{ij}} - \hat{\sigma}_{\tilde{\xi}_{ij}}}{\hat{\sigma}_{\xi_{ij}} - \hat{\sigma}_{\tilde{\xi}_{ij\text{ид}}}}, \quad (5)$$

где $\hat{\sigma}_{\xi_{ij}}$ – оценка среднеквадратичного отклонения (СКО) в *i*-ой точке *j*-траектории до сглаживания;

$\hat{\sigma}_{\tilde{\xi}_{ij}}$ – оценка СКО в *i*-ой точке *j*-траектории после сглаживания;

$\hat{\sigma}_{\tilde{\xi}_{ij\text{ид}}}$ – оценка СКО в *i*-ой точке *j*-траектории после сглаживания МНК при априорно известной степени (структуре) сглаживающего полинома.

Средний показатель эффективности метода по *j*-траектории:

$$\mu_{j_{\text{ср}}} = \frac{1}{n}\sum_{i=1}^{n} \mu_{ij}, \quad (6)$$

где μ_{ij} – показатель эффективности в *i*-ой точке *j*-траектории.

Средний показатель эффективности метода по *N* траекториям:

$$\mu_{i_{\text{ср}}} = \frac{1}{N}\sum_{i=1}^{N} \mu_{ij}. \quad (7)$$

Средний показатель эффективности метода по *N* траекториям, каждая из которых содержит *n* точек:

$$\mu_{\text{ср}} = \frac{1}{n}\frac{1}{N}\sum_{i=1}^{n}\sum_{i=1}^{N} \mu_{ij} = \frac{1}{N}\sum_{i=1}^{N} \mu_{j_{\text{ср}}}. \quad (8)$$

Так как выигрыш в точности по траектории может оказаться по величине меньше единицы, то вводится показатель проигрыша, как величина обратная выигрышу, и определяется оценка вероятности проигрышей в точности в результате сглаживания.

Получение показателей качества и эффективности метода сглаживания осуществлялось путем имитационного моделирования на ЭВМ [6, 7]. В процессе моделирования истинные реализации первичных координат суммировались с реализациями ошибок измерений первичных координат, распределенных по нормальному закону.

Так как наивысшее качество сглаживания обеспечивается в середине интервала сглаживания и ухудшается на его концах, то предлагаются для использования и такие показатели, как средний выигрыш и средняя эффективность на концах и в середине интервала сглаживания.

Исследованиям подвергались следующие методы сглаживания [1, 2].

Первый метод - нелинейное оптимальное адаптивное сглаживание с оптимизацией степени сглаживавшего полинома. Сущность этого метода стоит в том, что проверка значимости коэффициентов сглаживающего полинома начинается с последнего компонента вектора оценок коэффициентов сглаживающего полинома и заканчивается, когда очередной проверяемый компонент оказывается выше порогового уровня. Он и все предшествующие коэффициенты считаются значимыми и принимаются равными оценкам соответствующих компонент. Все последующие компоненты считаются незначимыми и принимаются равными нулю.

Второй метод - нелинейное оптимальное адаптивное сглаживание с оптимизацией структуры сглаживающего полинома. Сущность этого метода состоит в том, что проверка значимости коэффициентов сглаживающего полинома начинается с последнего компонента вектора оценок коэффициентов сглаживающего полинома и осуществляется тройками компонент. Если внутри тройки проверяемый компонент выше порогового уровня, то он остаётся в составе вектора без изменений, а те компоненты, которые меньше порогового уровня, приравниваются к нулю. Процесс проверки значимости компонент продолжается до тех пор, пока в составе очередной тройки компонент два компонента не окажутся выше порога (структура I) или все три компонента не окажутся выше порога (структура II). Все предшествующие компоненты вектора считаются значимыми и принимаются равными оценкам соответствующих коэффициентов сглаживающего полинома.

Значения пороговых уровней выбираются из распределения Фишера в зависимости от уровня значимости и числа степеней свободы и записываются в виде таблиц [8 - 10].

Алгоритм эксперимента. Оценка показателей качества и эффективности методов осуществлялась путём математического моделирования на ЭВМ. Моделировались значения вторичных координат, которые по формулам простых методов [3] пересчитывались в первичные координаты с учётом местоположения и типа измерительных средств. Затем к полученным значениям первичных координат прибавлялись значения ошибок измерений, распределённых по нормальному закону. Полученные таким образом первичные данные измерений подвергались идеальному и нелинейному оптимальному адаптивному сглаживанию. По результатам обработки оценивались средние показатели качества и эффективности методов. В качестве идеального (по точности) сглаживания применялось нелинейное оптимальное сглаживание МНК при точно известной до обработки структуре полинома.

Условия эксперимента. Моделирование данных измерений осуществлялось следующим образом [1]:

- фиксировались значения вторичных координат X(t) = 10000 м и Y(t) = 10000 м во времени, а координата *Z* изменялась по закону:

$$Z(t) = \sum_{k=0}^{4} b_k (t - t_0)^k, \qquad (9)$$

где *Z(t)* – моделируемые истинные значения Z во времени;

t – текущий момент времени;

t0 – момент времени, соответствующий середине интервала сглаживания;

b_k – коэффициенты полинома ($b_0 = 10^4$; $b_1 = 50, 100, 150\ \text{с}^{-1}$; $b_2 = 0.5, 10,\ \text{с}^{-2}$; $b_3 = 0, 0.5, 1\ \text{с}^{-3}$; $b_4 = 0, 0.5, 0.1\ \text{с}^{-4}$);

- число точек на интервале сглаживания n = 5, 7, 9, 11, 13, 15, 17, 19, 21, 25, 27 при частоте дискретизации 1 Гц;

- число траекторий, по которым определялись средние показатели качества и эффективности метода, для каждого варианта размещения измерительных средств, равнялось 81 и определялось числом всевозможных сочетаний коэффициентов моделирующего полинома;

- моделировались данные измерений от двух трехкоординатных радиолокационных станций (РЛС) с координатами местоположения относительно старта (0, 0, 0); (0, 0, 0) и (0, 0, 0); (0, 0, 7000);

- среднеквадратическая ошибка измерений по дальности $\sigma_R = 40; 15; 1; 0.1$ м, по угловым координатам $\sigma_{\alpha,\beta} = 7'; 1'; 0.5'; 0.1'; 0.03'; 0.01'$;

- автокорреляция ошибок измерений моделировалась экспоненциальной зависимостью $exp(\varepsilon_\xi|\tau_k|)$ с временем корреляции $\tau_k = \varepsilon_\xi^{-1} = 0, 1, 3\ 5$ с;

- поиск оптимальной степени (структуры) сглаживающего полинома осуществлялся по критерию Фишера при доверительной вероятности 0,95;

- максимально возможная степень сглаживающего полинома принята равной 6.

Наиболее характерные реализации моделируемых зависимостей по координате Z приведены на рис. 1 [1].

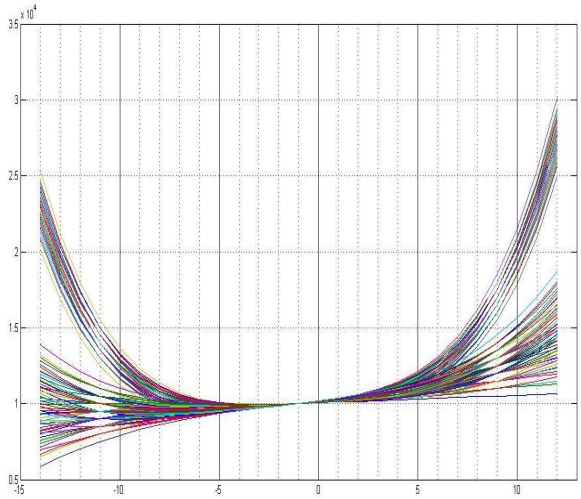


Рисунок 1 - Характерные реализации моделируемых зависимостей по координате Z

Начало отсчёта совмещено с серединой интервала сглаживания. Слева, по оси ординат, указаны значения истинного параметра для

нижнего пучка траекторий, справа – для верхнего. Для сравнения методов сглаживания обработке подвергались многопараметрические данные измерений с некоррелированными ошибками измерений. СКО по дальности принималась равной 40 м, по угловым координатам 7′.

Результаты исследований оценивались по средним показателям качества и эффективности в точках интервала сглаживания по 81 траектории для каждого варианта размещения измерительных средств и различных интервалов сглаживания. На рис.2-4 показаны результаты при n= 13, 17 и 27.

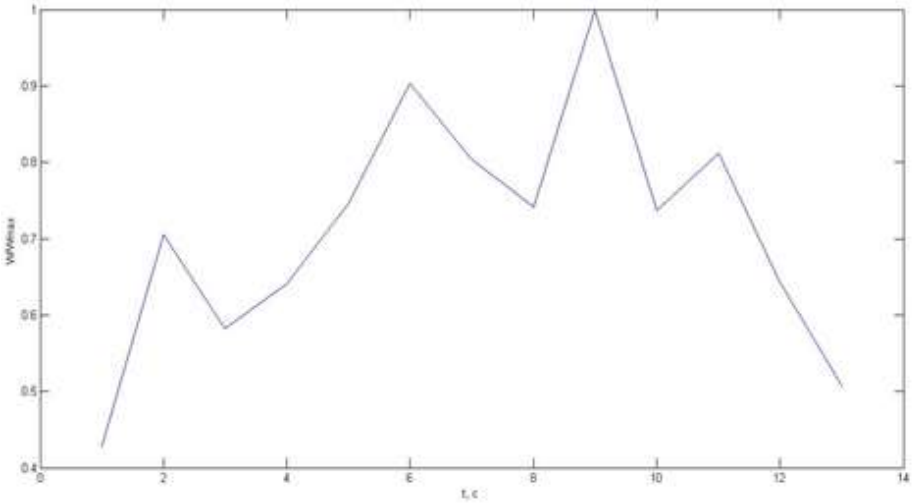


Рисунок 2 - Средний выигрыш в точности в точках интервала сглаживания (n=13)

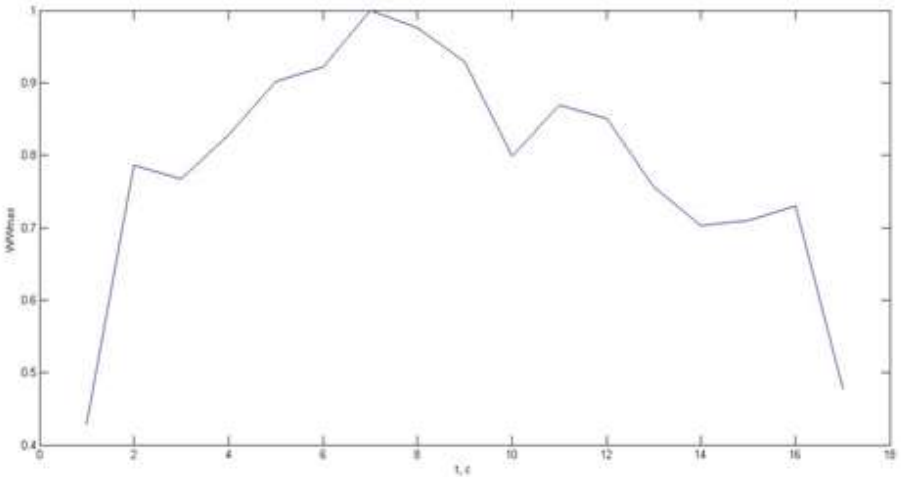


Рисунок 3 - Средний выигрыш в точности в точках интервала сглаживания (n=17)

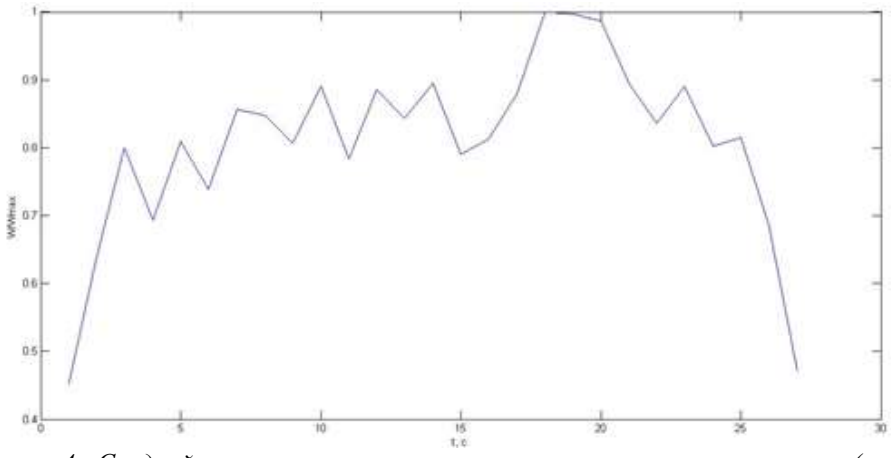


Рисунок 4 - Средний выигрыш в точности в точках интервала сглаживания (n=27)

Анализ результатов исследований позволяет сделать следующие выводы [2, 4]:

- алгоритм, с оптимизацией структуры сглаживающего полинома, более эффективен, чем алгоритм, оптимизирующий степень сглаживающего полинома;

- алгоритм, оптимизирующий структуру сглаживающего полинома по структуре I, более эффективен, чем алгоритм, оптимизирующий структуру сглаживающего полинома по структуре II;

- средний (по *N* траекториям) выигрыш в точности в точках интервала сглаживания имеет закономерный характер и сохраняется приблизительно постоянным в средней части в пределах 3/5 времени интервала сглаживания (отклонение выигрыша в точности на этом интервале от максимального выигрыша в точности не превышает 10 - 20 %;);

- при малых интервалах сглаживания (n = 5, 7, 9) средний показатель эффективности и средний выигрыш в точности исследуемых методов сглаживания в отдельных точках могут превосходить аналогичные показатели, получаемые при идеальном сглаживании вследствие того, что на коротких интервалах сглаживания отдельные коэффициенты моделирующего полинома становятся соизмеримыми с шумами (ошибками измерений) и в процессе работы адаптивного алгоритма отбраковываются; благодаря этому повышается точность вычислений;

- на краях траектории средний показатель качества и эффективности уменьшается, что является общим недостатком всех методов сглаживания;

- проигрышей в точности не наблюдалось.

Дальнейшим исследованиям подвергался алгоритм, оптимизирующий структуру сглаживающего полинома по структуре I с некоррелированными и при коррелированных ошибках измерений. Результаты исследований, алгоритма с числом точек на интервале сглаживания n = 27 представлены в табл. 1.

Таблица 1 - Эффективность алгоритма оптимизации структуры сглаживающего полинома с некоррелированными ошибками измерений

Точек на интервале	СКО R, м	СКО α, угл. мин	СКО β, угл. мин	Показатель качества и эффективности			Условия усреднения
				$\mu_{ср}$	$W_{ср}$	$W_{ис.ср}$	
27	40	7	7	0,949	4,841	5,146	по интервалу
				0,970	5,364	5,664	в ср. точке
				0,974	5,142	5,387	на 3/5 инт.
				0,931	2,356	2,524	на конц. инт.
27	40	1	1	0,967	4,716	5,059	по инт.
				0,973	5,139	5,471	в ср. точке
				0,971	4,978	5,182	на 3/5 инт.
				0,937	2,294	2,407	на конц. инт.
27	5	1	1	0,956	4,536	4,764	по интервалу
				0,978	4,982	5,107	в ср. точке
				0,976	4,806	4,916	на 3/5 инт.
				0,933	2,128	2,273	на конц. инт.
27	5	0,5	0,5	0,969	4,474	4,750	по интервалу
				0,973	4,997	5,374	в ср. точке
				0,976	4,811	5,132	на 3/5 инт.
				0,933	2,261	2,348	на конц. инт.
27	1	0,1	0,1	0,969	4,467	4,728	по интервалу
				0,973	4,896	5,346	в ср. точке
				0,976	4,749	5,072	на 3/5 инт.
				0,933	2,134	2,312	на конц. инт.
27	0,1	0,03	0,03	0,967	4,458	4,741	по инт.
				0,973	4,872	5,212	в ср. точке
				0,974	4,780	5,081	на 3/5 инт.
				0,974	2,214	2,341	на конц. инт.
27	0,1	0,01	0,01	0,969	4,473	4,749	по интервалу
				0,978	4,963	5,330	в ср. точке
				0,976	4,809	5,086	на 3/5 инт.
				0,933	2,196	4,304	на конц. инт.

Для исследования алгоритма обработке подвергались многопараметрические данные двух РЛС с некоррелированными ошибками измерений, расположенных в точках с координатами (0, 0, 0) и (0, 0, 7000). Результаты исследований с меньшим количеством точек на интервале сглаживания автором не приводятся с целью уменьшения объема статьи. Полученные результаты показывают, что:

- выводы, сформулированные применительно к данным с некоррелированными ошибками измерений, подтверждаются и в этих условиях работы алгоритма;
- для обеспечения высокой эффективности и качества сглаживания длительность интервала сглаживания необходимо выбирать в 1,5 и более раз превышающий время корреляции ошибок измерений;
- зависимость выигрыша в точности от отношения длительности интервала сглаживания ко времени корреляции носит в исследуемой области практически линейный характер.

Проведенный анализ позволяет сделать следующие выводы:

- алгоритм сохраняет работоспособность в весьма широком диапазоне задания СКО измерений (по дальности – от десятков метров до десятых долей метра, по угловым координатам – от нескольких угловых минут до нескольких угловых секунд);
- средние показатели качества и эффективности практически сохраняют свои величины при изменении СКО измерительных средств;
- проигрышей в точности не наблюдалось;
- на краях траектории по-прежнему наблюдается уменьшение среднего показателя качества и эффективности.

Для исследования алгоритма с коррелированными ошибками обработке подвергались многопараметрические данные измерений двух РЛС, расположенными в точках с координатами (0, 0, 0) и (0, 0, 7000).

Результаты исследований сведены в таблицу 2, где t – длительность интервала сглаживания, а τ – время корреляции ошибок измерений.

Таблица 2 - Эффективность алгоритма оптимизации структуры сглаживающего полинома при коррелированных ошибках измерений

Число точек на интервале сглаживания	Время корреляции, с	Выигрыш в точности			
		по интервалу	в средней точке	на 3/5 интервала	на концах интервала
9	0	2,167	2,490	2,453	1,580
	1	1,762	1,988	1,905	1,387
	3	1,520	1,591	1,546	1,405
	5	1,452	1,494	1,465	1,378
13	0	2,413	2,629	2,715	1,583
	1	1,883	2,037	2,054	1,357
	3	1,592	1,654	1,671	1,306
	5	1,524	1,560	1,580	1,305
17	0	2,725	2,958	3,017	1,596
	1	2,047	2,301	2,232	1,388
	3	1,653	1,752	1,731	1,360
	5	1,570	1,647	1,614	1,390
21	0	3,200	3,909	3,508	1,761
	1	2,200	2,320	2,340	1,512
	3	1,682	1,659	1,734	1,429
	5	1,588	1,581	1,620	1,395
25	0	3,439	3,866	3,761	1,800
	1	2,417	2,794	2,634	1,529
	3	1,782	1,937	1,874	1,387
	5	1,652	1,728	1,709	1,392

Выводы

Разработанный метод математического описания технологии совместной обработки избыточной траекторной информации и проведение анализа устойчивости аппаратно-программного комплекса послеполетной обработки траекторной информации к аномальным ошибкам данных измерений, позволяет получить независимые оценки коэффициентов сглаживающего полинома, обосновать выбор критерия проверки статистических гипотез и подтвердить устойчивость аппаратно-программного комплекса

к аномальным ошибкам данных, которые подлежат контролю.

Проведенная методом математического моделирования сравнительная оценка качества и эффективности разработанных алгоритмов адаптивного оптимального сглаживания многопараметрических данных измерений позволяет сделать вывод, что алгоритм, оптимизирующий структуру сглаживающего полинома, более эффективен, чем алгоритм оптимизирующий степень сглаживающего полинома.

Средние показатели качества и эффективности метода практически сохраняют свои величины при уменьшении СКО измерительных средств.

Литература

1. Огоднийчук, Н. Д. Обработка траекторной информации. Ч. II / Н. Д. Огороднийчук. – К.: КВВАИУ, 1986. – 224 с.

2. Щербов, И. Л. Совместная обработка данных траекторных измерений наземных и воздушных измерительных средств / И. Л. Щербов, В. В. Паслен, К. И. Мотылев, М. В. Михайлов // Наукові праці Донецького національного технічного університету. Серія: „Гірничо-геологічна”. – Донецьк: ДонНТУ, 2006. – Вип. 111. – Т. 2. – С. 55.

3. Огоднийчук, Н. Д. Обработка траекторной информации. Ч. 1 / Н. Д. Огороднийчук. – К.: КВВАИУ, 1981. – 141 с.

4. Паслен, В. В. О возможности применения базисных функций двух переменных в практике траекторных измерений / В. В. Паслен, А. В. Мильштейн, И. В. Дрозда, Я. А. Савицкая // Современные проблемы радиотехники и телекоммуникаций РТ-2012: 8-я Международная молодёжная научно-техническая конференция, 23 – 27 апреля 2012 г. – Севастополь, 2012. – С. 330.

5. Коновалов, А. А. Основы траекторной обработки радиолокационной информации. Часть 2 / А. А. Коновалов. – СПб.: СПбГЭТУ «ЛЭТИ», 2014. – 180 с.

6. Хейфец М. И. Обработка результатов испытаний: Алгоритмы, номограммы, таблицы / М. И. Хейфец. – М. : Машиностроение, 1988. – 168 с.

7. Коновалов, А. А. Основы траекторной обработки радиолокационной информации. Часть 1. / А. А. Коновалов. – СПб.: СПбГЭТУ «ЛЭТИ», 2013. – 164 с.

8. Брюс, П. Практическая статистика для специалистов Data Science: разведочный анализ данных / П. Брюс, Э. Брюс. — СПб.: БХВ-Петербург, 2018. — 304 с.

9. Тюлин, А. Е. Летные испытания космических объектов: определение и анализ движения по экспериментальным данным / А. Е. Тюлин, В. В. Бетанов; под. ред. А. Е. Тюлина. - Москва: Радиотехника, 2016. - 332 с.

10. Теория и методы инженерного эксперимента: Курс лекций / Н. Г. Бойко, Т. А. Устименко.- Донецк: ДонНТУ, 2009. – 158 с.

Щербов И. Л. Исследование алгоритма адаптивного нелинейного оптимального сглаживания многопараметрических данных измерений. *Разработанный метод математического описания технологии совместной обработки избыточной траекторной информации и проведение анализа устойчивости аппаратно-программного комплекса послеполетной обработки траекторной информации к аномальным ошибкам данных измерений, позволяет получить независимые оценки коэффициентов сглаживающего полинома, обосновать выбор критерия проверки статистических гипотез и подтвердить устойчивость аппаратно-программного комплекса к аномальным ошибкам данных, которые подлежат контролю.*

Ключевые слова: *технология совместной обработки, аппаратно-программный комплекс, сглаживающий полином, устойчивость к аномальным ошибкам.*

Shcherbov I. L. Investigation of the algorithm for adaptive nonlinear optimal smoothing of multiparameter measurement data. *The developed method of mathematical description of the technology for joint processing of redundant trajectory information and analysis of the stability of the hardware-software complex of post-flight processing of trajectory information to abnormal errors of measurement data allows obtaining independent estimates of the coefficients of the smoothing polynomial, justifying the choice of the criterion for testing statistical hypotheses and confirming the stability of the hardware-software complex to abnormal data errors that are subject to control.*

Keywords: *joint processing technology, hardware and software complex, smoothing polynomial, resistance to anomalous errors.*

Статья поступила в редакцию 19.11.2020
Рекомендована к публикации профессором Мальчевой Р. В.

УДК 004.946 : 004.932.2 : 004.921

Анализ технологий для создания дополненной реальности

М. В. Крахмаль
Донецкий национальный технический университет, г. Донецк
meri.krakhmal@gmail.com

Аннотация
В статье рассматриваются технологии виртуальной и дополненной реальности. Проанализированы способы для построения и распознавания моделей на основе маркеров и координат местоположения пользователя. Рассмотрены и выбраны наиболее подходящие алгоритмы реализации дополненной реальности для визуализации объектов. Приведена сравнительная характеристика двух технологий, показаны преимущества и недостатки. Проведен анализ текущего состояния и степени разработок технологий дополненной реальности, представлены дальнейшие перспективы развития. Поставлены задачи для дальнейшего исследования.

Введение

За последние 5 лет технологии виртуальной реальности (*ВР или VR – virtual reality*) развились от сомнительно перспективных до повсеместно используемых и внедряемых. Рынок приложений, использующий элементы *VR,* ежегодно расширяется за счет появления новых отраслей [1]. В инженерных задачах *VR*-технология применяется для улучшения уровня безопасности и условий работы с реальными техническими объектами, например, с помощью наложения текстовых инструкций и графических предупреждений [2].

Современная система образования построена таким образом, что во время обучения теоретические знания доминируют над практическими. Но хорошо известно, что знания, приобретенные практическим методом, усваиваются обучающимися лучше и сохраняются на более длительный период, чем знания, полученные только в виде теории. В некоторых образовательных организациях проведение практических занятий может быть затруднено или невозможно, так как отсутствуют необходимые приборы, стенды, материалы или реактивы для демонстрации.

Таким образом, ситуация в сфере образования, касающаяся знаний, полученных с помощью практики, обуславливает актуальность применения новых информационных технологий в виде *VR*-технологий. Одним из многообещающих направлений развития инновационных образовательных технологий является использование дополненной реальности (*ДР или AR – augmented reality*) в процессе обучения [3].

Приложения на основе дополненной реальности могут помогать человеку сфокусировать внимание на определенных элементах изображения, получаемого с камеры и улучшать понимание объектов окружающего мира путем предоставления необходимой информации (накладывается на изображение в виде текстового/звукового сообщения или визуального образа – 3D-модели).

Главным фактором востребованности виртуальных тренажеров, разработанных с помощью технологий дополненной реальности, является метод их создания, что, в свою очередь, порождает задачу формирования особой среды для их быстрого проектирования, главным образом экспертами конкретной предметной области, не имеющими глубоких навыков программирования.

Важную роль играют границы контакта человека с искусственным миром, обеспечивающие включение и погружение в его содержание, дающие возможность эффективного и безопасного взаимодействия с ним.

Исходя из этого данное исследование направлено на выявление степени актуальности и востребованности, а также определение ситуационного уровня внедрения виртуальных технологий в образовательную деятельность.

Цель исследования и постановка задачи

В настоящее время актуальной является проблема формирования содержания такого обучения, направленного на подготовку учащихся к применению не только «сегодняшних» технологий, но и технологий, которые появятся в будущем. Это будет способствовать внедрению новейших информационных технологий в процесс обучения, повседневную жизнь учащихся, повышению эффективности обучения разным учебным дисциплинам.

Современные технологии используют взаимодействие человека и компьютера. Дополненная реальность использует совершенствование интерфейса человека и

реального окружающего мира при помощи использования компьютерных технологий.

Целью данной работы является анализ методов обработки информации при помощи технологии дополненной реальности для визуализации объектов.

Для достижения цели в работе решаются следующие задачи:

– анализ развития технологии виртуальной и дополненной реальности;

– анализ методов создания и построения дополненной реальности;

– сравнительная характеристика технологий и перспективы дальнейшего развития;

– обоснование применения технологий дополненной реальности в обучении;

– определение задач для дальнейшего исследования.

Анализ предметной области

Технологии *VR/AR* предоставляют новые возможности для взаимодействия человека с цифровым миром. Обе технологии применяются на благо общественности во многих сферах жизнедеятельности. Они позволяют получать новые навыки и повышать жизненный опыт. Но необходимо различать эти технологии, так как каждая из них имеет различный подход, организацию для передачи информации и взаимодействия с пользователем (рис.1).

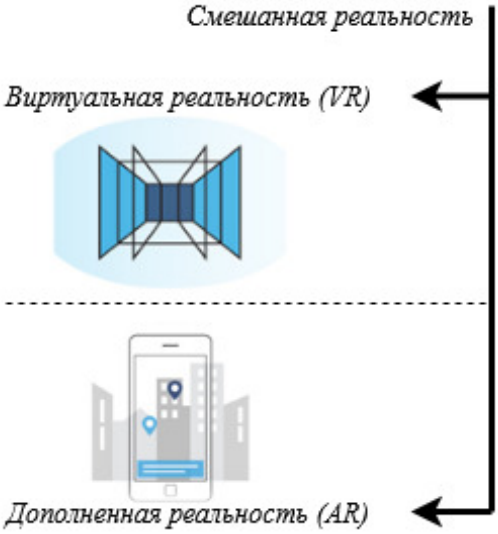


Рисунок 1 – Модель гибридной реальности

Виртуальная реальность – технология, которая погружает человека в искусственно смоделированный мир, собственную реальность. Взаимодействие с виртуальным миром происходит с помощью специальных гаджетов (очки, шлем, перчатки и контроллеры). Виртуальная реальность имитирует как воздействия, так и реакции на них [4].

VR-технология создает среду и условия реальной ситуации, чтобы человек мог приобретать или улучшать определенные навыки, знания и умения. Примером подобных систем и программ являются виртуальные тренажеры для пилотов, спасателей, врачей, водителей, военных.

Дополненная реальность – технология, которая добавляет в реальный мир виртуальные объекты, расширяя область данных, воспринимаемых пользователем. Дополненная реальность формируется путем обработки данных, полученных с камеры устройства. Программное обеспечение дополняет картинку виртуальными объектами. Объектами дополненной реальности могут быть изображения, видео, текст, объемные модели и другие (рис.2) [5].

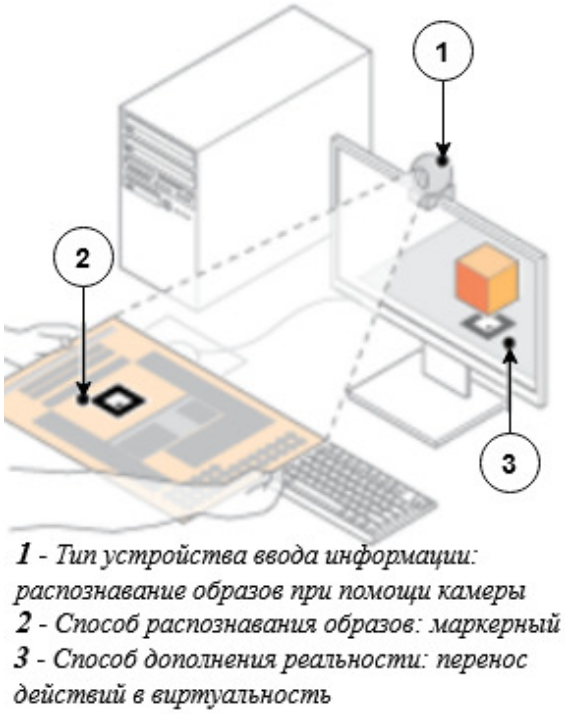


Рисунок 2 – Принцип работы технологии дополненной реальности

В настоящее время *AR*-технология предлагает уникальные возможности, сочетающие физическую и виртуальную реальность в единые миры. Это новый способ управления и взаимодействия с этим миром. Данная технология обеспечивает составное представление для пользователя с сочетанием реальной сцены, просматриваемой пользователем, и виртуальных сцен, созданных компьютером. Новый подход повышает эффективность и привлекательность преподавания и обучения.

Возможность накладывать сгенерированные компьютером виртуальные объекты на реальный мир меняет способ взаимодействия и тренинги становятся реальными, их можно увидеть в реальном времени, а не в статике. *AR*-реальность приносит виртуальную информацию или объект в любую косвенную среду мира, чтобы модели на реальных объектах или сценах максимально естественно и интуитивно взаимодействовали с пользователем в реальном времени. Это интерактивная среда, в которой реальная жизнь дополняется виртуальными вещами в реальном времени. Дополненная реальность позволяет пользователю увидеть настоящую жизнь, дополненную различными элементами, не погружаясь полностью в синтетическую среду.

Смешанная реальность (MR) – технология, которая объединяет дополненную и виртуальную реальности. Данная технология позволяет не только дополнить реальный мир виртуальными объектами, но и обеспечить взаимодействие с ними, тем самым расширяя возможности дополненной реальности. Это своего рода продвинутый тип *AR*. Другая интересная форма *MR* берет свое начало в полном погружении в виртуальную среду, где реальный мир заблокирован. Но в данном случае виртуальная среда, которую видит человек, привязана к реальной среде и как бы перекрывает её.

На сегодняшний день большинство исследований в области *AR*-технологии сконцентрировано на использовании живого видео, подвергнутого цифровой обработке и дополненного компьютерной графикой. Более серьезные исследования включают использование отслеживания движения объектов, распространение координатных меток при помощи машинного зрения и конструирование управляемого окружения, состоящего из произвольного количества сенсоров.

Технология дополненной реальности стремительно развивается и получает все большее распространение в различных областях науки и техники [6]. В военном деле дополненная реальность помогает в подготовке военнослужащих путем обучения на специальных полигонах с использованием дополненной реальности. В авиации дополненная реальность помогает обучать новых пилотов, а также служит навигатором. В сфере маркетинга *AR* помогает продавать различные товары и услуги путем показа этих товаров с помощью специальной рекламы. В сфере туризма может помогать пользователю в поиске пути, находить и выводить информацию об архитектурном здании или другой туристический объект, на который пользователь наводит свой мобильный телефон. В сфере дизайна с помощью дополненной реальности можно, например, удобно представить архитектуру разработанной здания любого размера.

В сфере образования технология дополненной реальности помогает детям познавать мир, поскольку они могут навести камеру телефона на предмет заинтересованности и увидеть подробную информацию о нем, в виде анимированных 3D-моделей.

За *AR*-учебниками будущее. С текущим внедрением мобильных технологий и недавними достижениями в области аппаратного обеспечения *AR*-технологии становятся всё более доступными и широко используемыми.

Однако у *AR/VR*-технологий есть свои сильные и слабые стороны, которые следует учитывать при интеграции этих технологий в среду обучения. Обе технологии предоставляют новую захватывающую образовательную реальность. В табл. 1 приведены основные преимущества и недостатки двух технологий, главные отличительные признаки.

Таблица 1 – Особенности технологий виртуальной и дополненной реальности

Технология	Виртуальная реальность	Дополненная реальность
Устройства визуализации	Умные очки, специальные гарнитуры, VR-шлемы	Смартфоны, планшеты, дисплеи других устройств
Источник изображения	Компьютерная графика или реальные изображения	Совмещение машинно-генерируемых изображений и реальных объектов
Ракурс объектов	Виртуальные модели меняют размер и позицию в соответствии с положением пользователя	Виртуальные модели отображаются на основании положения пользователя в реальном пространстве
Присутствие	Перенос в другую реальность без ощущения присутствия в реальном мире	Чувство присутствия в реальном мире с новыми наложенными элементами
Преимущества	• Перенос сложных инструкций в интерактивное обучение • Отсутствие реального ущерба оборудованию или здоровью в случае ошибки • Возможность повторения в формате обучения неограниченное количество раз	
Сложности	• Время на внедрение в зависимости от проекта от 3 до 6 месяцев • Стоимость технологий	

Развитие AR-технологии в настоящее время и дальнейшие перспективы

Термины *VR/AR*-технологий обрели популярность не так давно, их история начинается даже не в 20 веке. Первое устройство, объединяющее две картинки в одно трёхмерное изображение по принципу стереоскопического зрения, изобрёл английский физик Чарльз Уитстон ещё в 1838 году [7]. Однако эта технология значительно опередила время и практически на 100 лет никаких значительных изобретений в этой области не было.

В 1961 году инженерами фирмы Philco был создан первый шлем-дисплей, который получил название «Headsight» [8]. Шлем работал в реальном времени и состоял из видеоэкрана и системы слежения (рис.3). Оператор мог наблюдать среду удалённо, регулируя направление слежения поворотами головы. Позднее данная технология стала использоваться пилотами.

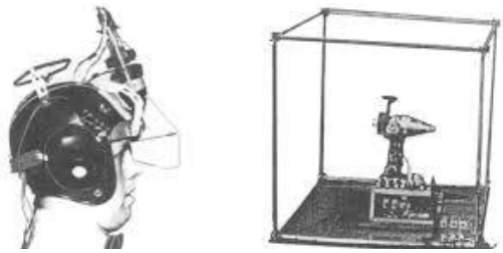


Рисунок 3 – Шлем-дисплей «Headsight»

На сегодняшний день рынок технологий виртуальной и дополненной реальности начинает развиваться и внедряться в различные сферы жизни. Одним из главных экспертов по аналитике информационных технологий является компания Gartner [9], которая ежегодно составляет график цикла зрелости технологий (Gartner Hype Cycles). Цикл зрелости технологий представляет графическую и наглядную презентацию зрелости появляющихся технологий через 5 фаз: отображаются современные технологии с их текущим положением во времени и уровнем ожиданий пользователей.

Проанализировав графики циклов зрелости технологий дополненной реальности за последние 17 лет (рис.4), можно сделать следующие выводы: по прогнозам 2006-2008 годов технология должна была достигнуть «полной зрелости» более, чем через 10 лет и после стремительного подъема по кривой опустилась к минимуму, прогнозируя при этом полное развитие в течении 5-10 лет. Начиная с 2019 года технология дополненной реальности не входит в цикл зрелости технологий. Согласно Gartner, *AR* стала настолько быстро развиваться, что больше не считается «новой технологией», она переросла из технологии, за которой «нужно наблюдать», в технологию, которую «нужно использовать». В 2020 году дополненная реальность достигла «полной зрелости» и стала проверенной отраслью технологий, которую начали массово использовать. В последнем отчете компания Gartner прогнозирует, что в 2021 году *AR*-технология станет одной из важнейших для компаний, которые вкладывают средства в разработку и использование искусственного интеллекта. Главным фактором продвижения технологии дополненной реальности является её сочетание с другими технологиями. Дополненная реальность выступает в качестве незаменимой компоненты для специализированных решений, требующих более глубокой интеграции между цифровым и реальным миром.

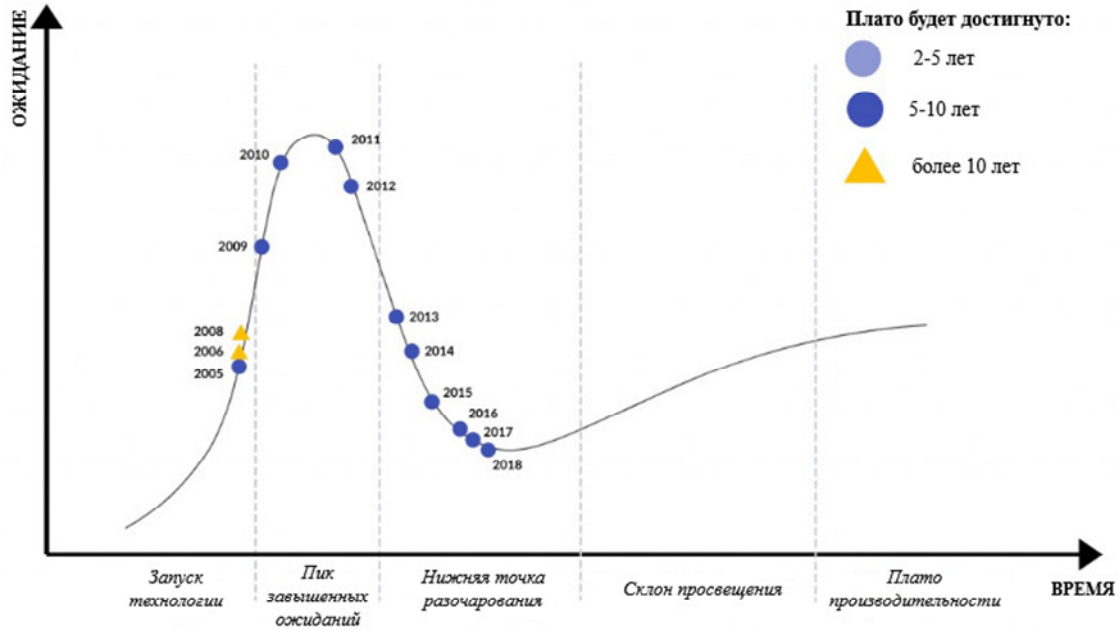


Рисунок 4 – Циклы зрелости технологий наложенной реальности (2005-2018)

В ближайшие годы *AR*-технология получит широкое распространение не только в сфере развлечений, но и в сферах здравоохранения, недвижимости, образования, коммерции, проектирования промышленности и энергетики. По данным аналитического агентства KPMG *AR*-технологию в 2019 году уже использовали 21% крупнейших отечественных компаний и это число непрерывно возрастает. Устройства дополненной реальности станут так же популярны, функциональны и востребованы, как и смартфоны.

По прогнозам компании TrendForce [10] к 2025 году технология дополненной реальности будет востребована в большем количестве областей, чем виртуальная реальность (рис.5).

Востребованность *AR*-технологий обусловлена следующими факторами: разработка не требует новейших устройств, технические средства имеют более низкие требования, чем для *VR*-технологий; применение возможно не только в специализированные комнатах, но и на смартфонах/планшетах; больше сфер для применения.

На развитие технологий активно влияют и мировые процессы. Так эпидемия COVID-19 дала толчок активному внедрению в различные компании *VR*-технологий для общения и взаимодействия сотрудников, находящихся вдали друг от друга. Так согласно данным аналитического агентства KPMG за 2019 год, *VR/AR*-технологии уже используют в 21% крупнейших отечественных компаний. И это число непрерывно растёт, о чём свидетельствуют массовые закупки шлемов *VR*. Ещё одна важная тенденция — открытие в компаниях новых отделов, специализирующихся на внедрении инновационных решений или даже отдельно *VR/AR*-технологий. Большую ставку компании делают на корпоративное обучение. В этом году на просторах интернета можно найти информацию о десятках успешных примеров внедрения *VR/AR*-проектов.

Рисунок 5 – Структура технологий дополненной и виртуальной реальности по отраслям к 2025 году

Способы построения дополненной реальности

Существует три основных принципа создания AR: технология на базе маркеров, пространственная технология и безмаркерная (на основе алгоритмов компьютерного зрения) [11].

Технология на основе *маркеров* заключается в создании специального маркера, расположенного в пространстве, который анализируется программным обеспечением для визуализации виртуальных моделей или объектов [12]. При использовании высококачественных моделей и графических фильтров, наложенный объект может иметь максимально реалистичный вид и не отличаться от окружающейся среды. При этом будет достигнут эффект физического присутствия наложенного объекта в реальном пространстве. Форма и тип маркера может варьироваться в зависимости от алгоритмов распознавания изображений: от простых геометрических фигур до человеческого лица или глаз. Данная технология является более устойчивой в работе, так как 3D-модели точно привязаны к маркеру.

Пространственная технология используется для определения объектов датчики (гироскоп, акселерометр) и данные геолокации (GPS/ГЛОНАСС). Место для визуализации виртуального объекта определяется с помощью координат в пространстве.

Безмаркерная технология широко используется при создании дополненной реальности, так как не нужно создавать специальные маркеры для распознавания объектов – объекты в реальном пространстве являются маркерами для визуализации виртуальных моделей [13]. Но существует недостаток, связанный с большими расчетами для идентификации объекта, который значительно увеличивает работу приложения. Данная технология основана на различных алгоритмах идентификации изображений и имеет три основных метода по распознаванию объектов на изображении: поиск шаблона (template matching), контурный анализ и сопоставление по ключевым точкам (feature detection). На любом изображении можно выделить особые точки, с помощью которых

можно сделать привязку к локальным особенностям изображения.

Метод *template matching* используется при поиске определенных участков на изображении, которые имеют одинаковые или похожие участки с заданным шаблоном. Поиск осуществляется путем последовательного перемещения шаблона по тестируемому изображению с интервалом в один пиксель (рис.6). После завершения работы выбирается та область исходного изображения, которая имеет наибольший коэффициент совпадения с шаблоном.

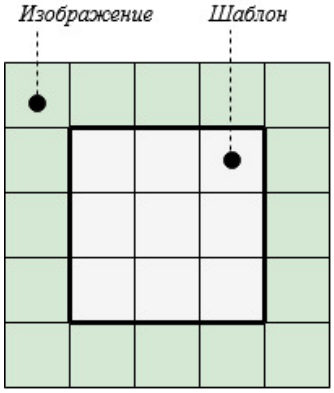


Рисунок 6 – Метод template matching

Метод имеет вероятностную характеристику, которая зависит от помех на картинке, масштаба или угла обзора, поэтому нельзя с точностью сказать был ли найдет объект-шаблон на изображении.

Метод *контурного анализа* позволяет распознавать, сравнивать и осуществлять поиск графических объектов по их контурам. Главной особенностью при проведении контурного анализа является то, что внутренние точки объекта не учитываются и контур должен содержать достаточно информации о форме объекта (рис.7).

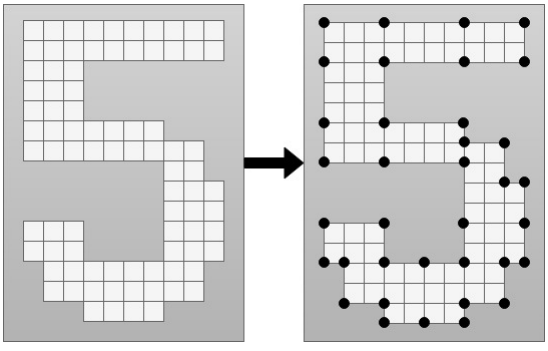


Рисунок 7 – Метод контурного анализа

В результате этого накладываются ограничения на применение данного метода, связанные с выделением контура: при перекрытии одного объекта другим, контур выделяется неправильно и невозможно распознать границы исходного объекта; из-за разной яркости фона и самого объекта контур может быть выделен нечетко; некоторые участки могут иметь помехи в виде шума, в результате чего контур невозможно определить.

Данный метод хоть и неустойчив к помехам, но его простота и быстродействие позволяют применять такой подход при распознавании печатного текста, во время перевода текста с одного языка на другой.

Метод *feature detection* [14] производит вычисление абстракций изображения, а затем выделение ключевых особенностей для сравнения двух изображений. Ключевые особенности изображений могут состоять как из изолированных точек, так и связанных областей (рис.8). Чем больше в изображении особых точек, тем качественнее будет выполняться трекинг. Одним из главных требований является то, что особые точки должны равномерно распределяться по всему изображению и быть контрастными по тону.

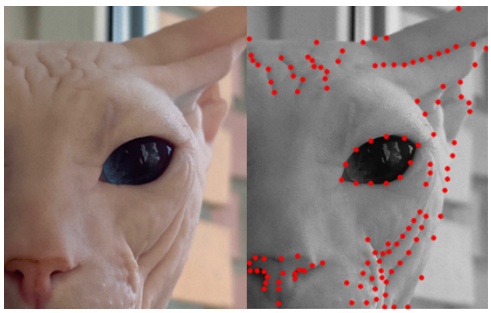


Рисунок 8 – Распределение особых точек методом feature detection

Для поиска и сопоставления ключевых точек на изображении необходимо использовать (рис.9): детектор, дескриптор и матчер.

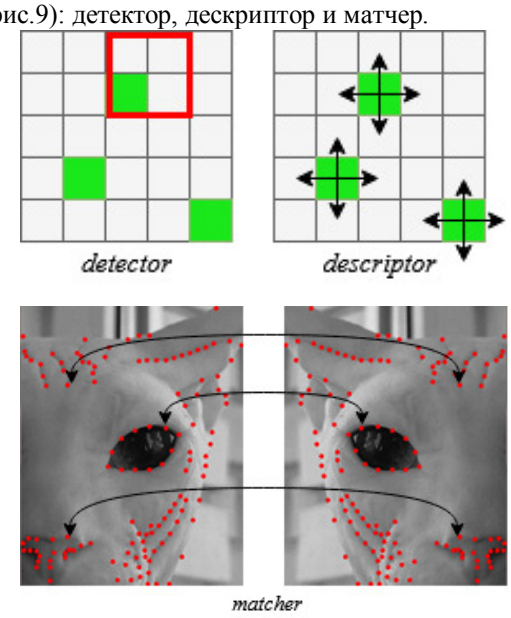


Рисунок 9 – Главные компоненты для поиска и сравнения особых точек

Детектор выполняет поиск особых точек на изображении. Дескриптор необходим для оценивания позиции найденных точек с

При использовании более быстрых детекторов, дескрипторов и матчеров можно достичь более качественного уровня отслеживания объекта или маркера. Выбор этих компонентов выполняется после проведения тестирования с замером скорости работы каждого из них. Кроме этого, для соответствующих задач необходимо разрабатывать алгоритм, который максимально точно может обнаружить нужный объект.

Практическое применение технологий дополненной реальности в образовании

В последние годы *VR/AR*-технологии стремительно развиваются. Это связано не только с популяризацией виртуальной реальности в различных сферах, но и с развитием компьютерных технологий, систем видеонаблюдения, повышением качества мониторов. Огромные вычислительные мощности современных компьютеров в сочетании с качественными full HD экранами позволяют моделировать и передавать человеку изображения поразительного качества и детализации. Это позволяет погрузить человека в виртуальную реальность, которая будет казаться реалистичной.

Изучая историю появления и развития *AR/VR*-технологий можно сделать вывод, что первоначально изобретения были направлены на развлечения. Однако со временем стали очевидны большие возможности виртуальной реальности в различных сферах, таких как медицина, образование, обучение, наука, производство.

Главным преимуществом виртуальной и дополненной реальности является возможность получать информацию, взаимодействовать с оборудованием находясь в безопасном месте и не подвергая риску себя и технику. Так на сегодняшний день существуют различные симуляторы, позволяющие обучать специалистов, операторов дорогостоящей и опасной техники, находясь в одном месте. Обучение пилотов самолётов гражданской и военной авиации, водителей специализированных автомобилей, моряков выполняется на тренажёрах-симуляторах, имитирующих кабину транспортного средства с полным погружением.

Система обучения, как и, собственно, технологии постоянно изменяются и часто проблема заключается в невозможности обеспечить всех физическими материалами для обучения. Кроме этого, существует ряд специальностей, обучение которым практически невозможно без систем виртуальной реальности.

помощью окружающих областей. Матчер осуществляет построение соответствий между двумя наборами точек изображения.

Технологии виртуальной и дополненной реальности дают ученикам и студентам возможность глубже изучать предметы, анализировать последствия мировых событий, участвовать в археологических экспедициях и многое другое, а главное — в развлекательной форме. *AR* и *VR* позволяют приобрести опыт, к которому у обучающихся обычно нет доступа.

Также новейшие технологии играют важную роль в обучении детей с физическими, социальными или когнитивными нарушениями. Ведь с помощью таких технологий можно создать инклюзивную учебную среду с учетом потребностей и возможностей каждого.

Успешные примеры внедрения *AR/VR*-технологий известны по всему миру:

– пекинские учёные провели исследование на тему влияния виртуальной реальности на процесс обучения. Разным группам детей преподавали одну и ту же дисциплину, но одной группе с использованием *VR*, а другой без. После тестирования знаний по пройденному материалу группа, использовавшая *VR,* показала результаты лучше на 20% успешней и закрепила материал лучше, о чём свидетельствовали повторные тесты спустя время;

– компания «Simtars» в Австралии предлагает вводный учебный курс для обеспечения безопасности работы персонала в шахтах. Путём виртуальной симуляции опасных ситуаций стажёры идентифицируют степень опасности и учатся применять методы её контроля и устранения, не выходя за пределы учебного класса;

– компания «AR production» выпустила мобильное приложение, которое позволяет видеть дополненную реальность в обычных школьных учебниках;

– компания «Виртуальные пространства» выпустила приложение «Our minds AR», которое визуализирует для учителя сообщения в мессенджере от учеников во время занятия в виде облачка над их головой в дополненной реальности.

На сегодняшний день проводится множество конференций по разработкам систем виртуальной, дополненной и смешанной реальностей. С каждым годом растет число публикаций, посвященных новым алгоритмам и способам создания таких систем. В этом направлении науки работают многие российские и зарубежные ученые. Большая часть материалов является англоязычными.

С развитием технологий в Донецком национальном техническом университете осуществлялось трехмерное моделирование

городских ландшафтов и университетской сетевой инфраструктуры в режиме реального времени [15, 16]. Разработки тренажеров проводились также для угольных шахт с целью визуализации опасных участков [17], что использовалось в первую очередь для тестирования знаний студентов по технике безопасности в шахте (рис.10).

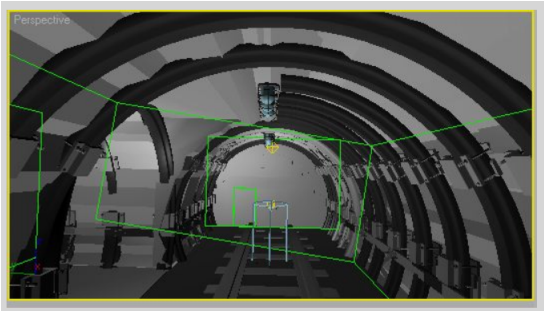


Рисунок 10 – Тренажер для тестирования знаний: модель фрагмента шахтной выработки

Кроме этого проводились работы по исследованию методов интеграции трехмерных объектов в видеоизображении в реальном времени (рис.11) [18], разработка систем расширенной реальности для моделирования трехмерных сцен.

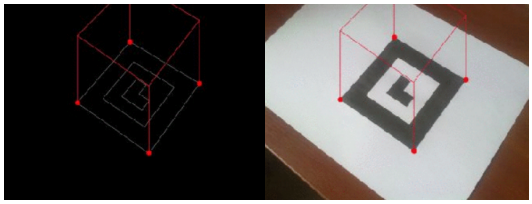


Рисунок 11 – Поэтапная обработка изображения

На портале магистров ДонНТУ имеется множество студенческих работ, связанных с трехмерным моделированием, движением 3D-объектов, с использованием дополненной реальности.

Выводы

Результаты проведенного исследования позволяют говорить о высокой степени актуальности и востребованности в сфере высшего образования применения новейших технологий дополненной и виртуальной реальности, а также констатировать довольно активный процесс внедрения *AR/VR* технологий в российских и зарубежных высших учебных заведениях.

Процесс внедрения новых форматов обучения является не только востребованным, но и взаимообусловленным. С одной стороны, студенты с интересом относятся к любым технологическим новинкам в обучении, и в рамках научной работы самостоятельно разрабатывают предложения и проекты для повышения эффективности систем коммуникации «преподаватель – студент». С другой стороны, преподаватели хорошо осведомлены о новейших виртуальных технологиях, используют онлайн-системы в своей работе и готовы к дальнейшему внедрению актуальных информационно-коммуникационных технологий в образовательный процесс высших учебных заведений.

Дальнейшие исследования направлены на тестирование методов безмаркерной технологии для выявления более точного трекинга объектов. Также для выделения особых точек на изображении необходимо проанализировать алгоритмы по детекторам углов. Для программной разработки будет использоваться библиотека компьютерного зрения OpenCV и уже имеющийся движок дополненной реальности (библиотека Vuforia).

Литература

1. Линовес, Д. Виртуальная реальность в Unity / Д. Линовес ; Пер. с англ. Рагимов Р.Н. – Москва : ДМК Пресс, 2016. – 316 с. : ил.

2. Henderson, S. Evaluating the Benefits of Augmented Reality for Task Localization in Maintenance of an Armored Personnel Carrier Turret / S. Henderson, S. Feiner ; Proceeding of IEEE International Symposium on Mixed and Augmented Reality (ISMAR '09). – 2009. – P. 135–144.

3. Бойченко, И. В. Дополненная реальность: состояние, проблемы и пути решения / И. В. Бойченко, А. В. Лежанкин // Доклады ТУСУР. – 2010. – № 1(21). – Ч.2. – С. 161–165.

4. Мурашов, А. А. Виртуальная реальность и дополненная реальность. Взгляд на будущее / А. А. Мурашов, Л. В. Смоленцева // Сборник трудов молодых ученых УВО «Университет управления «ТИСБИ». – Казань: 2016. – С. 91-96.

5. Благовещенский, И. А. Технологии и алгоритмы для создания дополненной реальности / И. А. Благовещенский, Н. А. Демьянков // Моделирование и анализ информационных систем. – 2013. – Т.20. – №. 2. – С. 129-138.

6. Дрокина, К. В. Анализ возможностей применения технологии дополненной реальности в современных условиях / К. В. Дрокина, Н. В. Григорьева // Инновационная наука. – 2016. – №2-1(14). – С. 114-116.

7. Виртуальная реальность врывается в жизнь современного человека // Роспатриот. – URL: https://rospatriotcentr.ru/news/264/ (дата обращения: 13.12.2020).

8. Развитие систем VR | Мечта о виртуальной реальности // tom's hardware guide. –

URL:http://www.thg.ru/business/razvitiye_sistem_vr/print.html (дата обращения: 13.12.2020).

9. 5 Trends Drive the Gartner Hype Cycle for Emerging Technologies, 2020 // Gartner. – URL: https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020/ (дата обращения: 13.12.2020).

10. AR/VR Devices Shipment Projected to Reach 43.2 Million Units in 2025, Driven by Glasses-Like AR/VR Devices, Says TrendForce // Gartner. – URL:https://www.trendforce.com/presscenter/news/20200723-10401.html (дата обращения: 13.12.2020).

11. Прокопов, С. А. Основы и принципы работы технологии дополненной реальности / С. А. Прокопов, Н. А. Соколовский // Информационно-управляющие системы. – 2018. – №2. – С. 201-203.

12. Абилмажинова, Б. С. Детекторы углов или как происходит распознавание маркеров дополненной реальности / Б. С. Абилмажинова, В.О. Андреев // Инновации в науке. – 2016. – № 2 (51). – С. 156-162.

13. Барышок, Е. С. Сравнительный анализ технологий построения дополненной реальности / Е. С. Барышок, В. В. Хорошева, А. А. Тропченко // Международный научно-исследовательский журнал. – 2017. – № 05(59). – Ч.3. – С. 6-9.

14. Awad, A. I. Image Feature Detectors and Descriptors. Foundations and Applications. Springer / A. I. Awad, M. Hassaballah. – 2016. – 438 p.

15. Аноприенко, А. Я. Моделирование университетской сетевой инфраструктуры / А. Я. Аноприенко, С. Н. Джон, С. В. Рычка // Вестник Кременчугского государственного политехнического университета. – Кременчуг, 2001. – Вып. 2(11). – С. 306–308.

16. Грищенко, А. В. Создание фотореалистичных трехмерных моделей / А. В. Грищенко, А. Я. Аноприенко // Информатика и компьютерные технологии : материалы II междунар. науч.–техн. конф., 13 дек. 2006 г., г. Донецк. – Донецк, 2006. – С. 390–391.

17. Аноприенко, А. Я. Трехмерное интерактивное моделирование угольной шахты на базе системы «Blender» / А. Я. Аноприенко, Е. В. Бабенко, Е. А. Навка, О. М. Оверчик // Моделирование и компьютерная графика : материалы IV междунар. науч.–техн. конф., 05–08 окт. 2011 г., г. Донецк. – Донецк, 2011. – С. 279–285.

18. Грищенко, С. В. Исследование методов интеграции трехмерных объектов в видеоизображения в реальном времени // Портал Магистров ДонНТУ, 2013 г. – URL: http://masters.donntu.org/2013/fknt/gryshchenko/diss/index.htm (дата обращения: 13.12.2020).

Крахмаль М. В. Анализ технологий для создания дополненной реальности. *В статье рассматриваются технологии виртуальной и дополненной реальности. Проанализированы способы для построения и распознавания моделей на основе маркеров и координат местоположения пользователя. Рассмотрены и выбраны наиболее подходящие алгоритмы реализации дополненной реальности для визуализации объектов. Приведена сравнительная характеристика двух технологий, показаны преимущества и недостатки. Проведен анализ текущего состояния и степени разработок технологий дополненной реальности, представлены дальнейшие перспективы развития. Поставлены задачи для дальнейшего исследования.*

Ключевые слова: *дополненная реальность, распознавание объектов, метод контурного анализа, метод feature detection.*

Krakhmal M. Analysis of technologies for creating augmented reality. *The virtual and augmented reality technologies are discusses in the article. Methods for constructing and recognizing models based on markers and coordinates of the user's location are analyzed. The most appropriate algorithms for the implementation of augmented reality for visualizing objects are considered and selected. Comparative characteristics of the two technologies are given, advantages and disadvantages are shown. The analysis of the current state and degree of development of augmented reality technologies is carried out, further development prospects are presented. Tasks for further research are set.*

Keywords: *augmented reality, object recognition, contour analysis method, feature detection method.*

Статья поступила в редакцию 09.12.2020
Рекомендована к публикации профессором Аноприенко А.Я.

УДК 004.048, 004.912

Обобщенная модифицированная модель представления текстовых информационных ресурсов

Н.К. Андриевская
Донецкий национальный технический университет
nataandr@yandex.ru

Андриевская Н.К. Обобщенная модифицированная модель представления текстовых информационных ресурсов. *Данная статья посвящена разработке новых эффективных подходов к формированию представлений текстовой информации при реализации задач информационного поиска в различных информационных системах (ИС). Общей проблемой является разреженность и большие объёмы входных данных, что приводит к необходимости уменьшить вычислительную сложность и повысить производительность методов основных процедур обработки данных. Для решения проблемы авторами предлагается использовать обобщенную модифицированную модель представления текста, а также предварительно выполнить редукцию по тематическим векторам, что одновременно уменьшает размерность и увеличивает семантику представления текста. Кроме этого, для усиления семантики предлагается использовать предметную онтологию системы. Это дает возможность многократно использовать ранее обработанные документы в структурированном виде и использовать уже существующие контекстные вектора в процедурах обработки.*

Ключевые слова: *векторная модель, сингулярное разложение, факторизация, модель bag-of-words, модель bag-of-concepts, LSA, NMF, семантические связи, RDF-граф, тензор, онтология.*

Введение

В последнее время все более возрастает интерес к системам и методам извлечения знаний для использования в самых различных сферах при решении конкретных практических задач. Переход на электронный документооборот, активное использование в профессиональной деятельности веб-ресурсов специалистами различного профиля, привело к резкому росту объёма информационных ресурсов (ИР) различного вида, в том числе и неструктурированных. Обмен информацией в ИС происходит преимущественно с помощью текстовых документов, текстовых сообщений и данных. Поэтому не случайно, что весьма значительную долю ИР современных ИС составляет информация текстового вида. Слабая структурированность информационных фондов осложняет управление информацией и работу пользователей с ней. В связи с этим созданию эффективных технологий хранения, обработки и поиска текстовой информации уделяется большое внимание.

Определенным образом это относится и к потокам информации, с которыми встречаются сотрудники вузов и других научных и научно-образовательных организаций. Развитие систем информационного поиска стимулировалось в значительной мере потребностями информационной поддержки научных исследований и образования, разработками автоматизированных библиотечных систем, систем управления знаниями, которые все активнее используются в различных сферах деятельности.

К примеру, в статье [1] была представлена функциональная структура и архитектура системы управления профессиональными знаниями сотрудников вуза для информационной поддержки преподавательской и исследовательской работы. В работе [2] был предложен базирующийся на онтологии подход к структуре такой системы, обеспечивающий представление и интерпретацию информации в виде знаний. При создании системы управления информационными ресурсами научно-образовательных организаций (СУИР) одним из основных модулей был описан блок приобретения знаний. В связи с этим, актуальной становится задача поиска интересующих пользователя ИР как среди огромного количества информации в сети Интернет, так по хранилищу коллекции документов. Формирование представлений ИР требует решения не только при поиске, но и в задачах автоматического аннотирования документа, индексации ИР, а также тематической рубрикации ИР.

Постановка проблемы

Для определения семантического соответствия необходимо, чтобы семантические значения запроса и документа были известны.

Определение семантических связей можно выполнить разными способами. Одни из них основаны на анализе текста и методах обработки естественного языка, другие используют различные методы индексации векторного пространства и словари типа WordNet [3], третьи используют явно определенную семантику различных онтологических ресурсов, в том числе и полновесные онтологии предметных областей.

В случае неструктурированных документов необходимы передовые методики работы с онтологиями для выделения их семантик и их использования для выявления зависимых документов. Общая производительность ИС в данном случае будет сильно зависеть от качества ранее разработанной онтологической модели и от качества наполнения онтологии экземплярами. Для СУИР была разработана базовая онтологическая модель, которая при онтологическом подходе к проектированию системы является «ядром системы» [4]. Онтология системы используется каждым разработанным модулем системы, особенно на этапах поиска [5] и извлечения информации.

Подведя итог, обобщим некоторыми требования, которые следует учитывать при разработке модели представления текста:

– так как информационные ресурсы в большинстве случаев редко изменяются, тем более значительным образом, построение представления каждого имеющегося в системе документа необходимо осуществлять однократно на этапе его ввода (импорта) в систему;

– так как ИР могут быть значительных объёмов, а коллекции документов, хранимых в системе, могут быть довольно большого размера, то возникает проблема производительности, в некоторых случаях даже критичная, что следует учитывать при разработке алгоритмов;

– следует избегать методов, обладающих большой вычислительной сложностью;

– необходимо использовать все достоинства онтологического подхода.

Таким образом, необходимо разработать такую эффективную модель представления текста, основанную на использовании онтологических ресурсов СУИР и традиционных моделей представления текста, которая обеспечивала бы экономию вычислительных мощностей системы.

Обзор моделей представления текста

В системах информационного поиска используются различные подходы к построению представлений хранимых документов. От характера используемых представлений документа существенно зависит качество поиска, в том числе точность, полнота, производительность и другие характеристики.

Модель представления текста – это формализованная математическая структура, построенная по неструктурированному тексту. Существуют следующие модели: векторная, языковая, модель скрытых тематик и др. В настоящее время наиболее часто используются векторная модель и модель на основе латентных семантик, а также их модификации.

В векторной модели Vector Space model (VSM) каждый терм представляет собой вектор в пространстве слов [6]. Суть метода отображения текста в вектор заключается в том, что каждому слову соответствует определенная координата в пространстве признаков или вес в соответствии с выбранной весовой функцией.

Для полного определения векторной модели текста необходимо выбрать саму весовую функцию. Главным образом, оценочные весовые функции строятся на частотных и вероятностных характеристиках текста, а также на дистрибутивной семантике.

Среди частотных методов используются следующие функции взвешивания:

– двоичные — вес равен 1, если терм встречается в документе и 0 если не встречается;

– TF (term frequency) — функция веса вычисляется от количества вхождений терма в документе [7];

– TF-IDF (term frequency — inverse document frequency) — функция веса вычисляется как произведение функции TF и функции IDF, благодаря которой можно снизить весомость наиболее широко используемых слов (предлогов, союзов, общих терминов и понятий) [7].

Существуют и другие весовые функции, например, логарифмическая, скорректированных весов Гаусса, вероятностная, которая основана на понятии условной вероятности и теория Байеса, а на моделях дистрибутивной семантики, например, word2vec [8].

Кроме всего, Vector Space model (VSM) модель имеет ряд расширений, среди которых наиболее известна Bag-of-Words (BOW). В модели BOW текст представляет собой набор неупорядоченных и несвязанных слов. Считается, что тексты, не имеющих общих слов или имеющих их небольшое количество, являются неблизкими по смыслу (семантически и тематически), что в целом неверно.

Для определения подобия между двумя документами используются различные меры сходства между двумя векторами, определяющими эти документы. Наиболее часто рассчитывают меру косинусной меры близости на основе скалярного произведения векторов, позволяющей провести оценку смысловой близости.

Достоинствами базовой векторной модели BOW является простота, возможность использования операций линейной алгебры для определения сходства между текстами и при ранжировании слов в поисковых запросах.

Недостатками является низкая степень сходства для более длинных текстов, отсутствует учет синонимии и полисемии. Главный недостаток заключается в том, что множество слов проецируется в пространство высокой размерности и разреженности и появляется эффект так называемого «проклятия размерности», когда сложность вычислений растет экспоненциально из-за увеличения размерности данных.

На преодоление этих моментов проводится много исследований и написано немало работ. Вариант решения проблемы многим авторам видится в необходимости перехода от традиционного представления текста в разреженном пространстве к новым, семантическим пространствам. Среди широко известных рассмотрим три подхода: на базе концептов, на базе контекстных векторов, на базе латентных семантических связей [9].

Модель bag of concepts (BOC) построена на основе BOW, позволяет учитывать скрытые смысловые связи. Концептами будем считать наборы синонимичных и семантически связанных слов с одинаковыми и похожими смыслами. Каждый концепт представляется не только с помощью определяющих его слов, но и с помощью контекстного вектора, содержащего веса слов в концепте.

Укрупненный алгоритм формирования контекстного вектора D(k) следующий:

1. Из модуля извлечения информации передаются все необходимые данные после автоматического разбора текста.
2. Формируется вектор из ключевых слов, извлеченных с помощью частотного анализа.
3. Вектор дополняется термами из онтологии.
4. Поисковый запрос расширяется терминами со сторонних ресурсов (WordNet, Wiktionary) [10].
5. В случае необходимости корректируется экспертом или автором ресурса полученный вектор.
6. Для каждого термина вычисляется TF (term frequency).
7. Элементы вектора сортируются в зависимости от значения TF.
8. Вектор ограничивается по количеству элементов k (параметр k можно изменять).
9. Полученный контекстный вектор сохраняется в онтологии в дальнейшем является векторным представлением документа.
9. Попутно наполняется онтология новыми терминами из массива ключевых слов.

Для получения набора концептов, близких к данному тексту, достаточно иметь вектор концепта и вектор текста, содержащий веса этих же слов в тексте. Два вектора сравниваются по какой-либо мере семантической близости и отбираются несколько наиболее близких, которые и образуют после сортировки вектор, представляющий текст в новом пространстве.

В качестве оператора перехода из одного пространства дескрипторов в другое используется матрица семантических связей «термина-на-термины» размерности kxk, состоящая из контекстных векторов k-мерного пространства.

Пусть текст представлен вектором D(k). Матрица семантических связей R (k, k) позволяет отобразить D(k) - исходное представление текста, в новое представление S(k), уже отражающие семантические связи между словами. Математически модель можно представить в виде выражения:

$$S_k = D_k * R_{kk} \quad (1)$$

В отличие от модели BOC, модель «контекстных векторов» учитывает зависимости между всеми словами текста, а не только концептами и формирует контекстные вектора по всем словам документа.

Среди моделей также интересным представляется подход на базе латентно-семантического анализа (LSA/LSI). Каждый набор документов (коллекция) имеет неявную, латентную семантическую структуру, представляющую собой объединение отдельных терминов документов. В результате количественного анализа латентных факторов можно получить новое семантическое представление нового документа, в результате которого веса терминов могут быть скорректированы, и поисковый образ документа станет более адекватным его содержанию.

Рассмотрим модель LSA/LSI подробнее. Формирование представлений осуществляется путем факторизации матрицы «Документы-термины» [11]. Существуют несколько способов определения тематик, например, LSA/LSI использует сингулярное разложение (SVD) в качестве факторизации (см. рис. 1).

Рисунок 1 – SVD разложение для LSA/LSI

Извлечение скрытых тематик также можно произвести на основе метода неотрицательной матричной факторизации

(NMF), который вместо SVD использует разложение на две матрицы (см. рис. 2) [12].

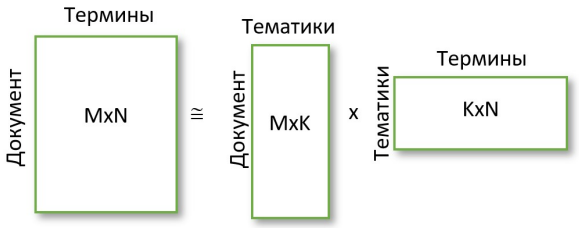


Рисунок 2 – Метод NMF

Главным недостатком всех моделей на базе BOW является огромная разреженность данных и отсутствие семантики. Для устранения этих недостатков модель BOW часто комбинируется с другими моделями.

В данной работе для повышения эффективности обработки данных и преодоления ограничений модели Bag-of-words, предлагается модифицировать модель Bag-of-words, комбинируя модели Bag-of-concepts и модель контекстных векторов, а также предметную онтологию.

Разработка модифицированной модели представления текста

Для представления многомерных семантических данных мы использовали формализм RDF семантической сети, где отношения моделируются как тройки (субъект, предикат, объект) и где предикат либо определяет отношения между двумя сущностями, либо отношения между сущностью и значением атрибута [13]. Трехмерный тензор является удобной структурой представления многомерных семантических данных (см. рис. 3) [14,15].

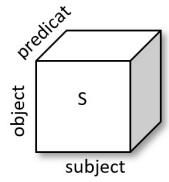


Рисунок 3 – Тензорное представление RDF-графа знаний

При этом понятия “объект” и “субъект” могут быть заменены понятием “концепт” или “термин”. При формировании тензора семантической близости для разных типов связей меры семантической близости были рассчитаны по-разному. Среди мер были как меры, вычисляемые по онтологии (таксономические и атрибутные), так и меры, вычисляемые по векторной модели представления текста.

Работа с тензорами не вызывает особых затруднений. Тензорная алгебра представлена большим количеством различных операций для работы с тензорами, среди которых выделим сложение, вычитание, свертка, тензорное произведение и тензорное разложение различных видов [16].

Существует и достаточное количество инструментальных средств и библиотек, предназначенных для работы с тензорной алгеброй, в том числе и для Python [17].

Поскольку вместо матрицы семантических связей «термина-на-термины» размерности kxk, состоящей из контекстных векторов k-мерного пространства мы используем трехмерное представление в виде семантического тензора «концепт-концепт-предикат» размерности kxkxp, то необходимо модифицировать стандартную модель Bag-of-concepts, обобщив ее для использования в пространствах третьего порядка.

В качестве оператора перехода из одного пространства дескрипторов в другое используется 3-х мерный тензор семантических связей $T_p^{кк}$.

Тензорное представление графа знаний позволяет эффективным образом вычислить обобщенную оценку семантической близости между объектами через факторизацию срезов тензора. Выполним аддитивную свертку тензора $T_p^{кк}$ с вектором коэффициентов значимости каждого типа отношения:

$$R^{kk} = \sum_{p=1}^{n} W^p T_p^{kk}, \text{где} \qquad (2)$$

$T_p^{кк}$ – тензор семантических связей;

W^p - вес, который определяет относительную важность каждого типа отношения;

n- число отношений;

k- число концептов;

$R^{кк}$-обобщенная матрица семантических связей концептов.

Таким образом, получаем значение обобщенной семантической матрицы путем аддитивной свертки семантического тензора по третьему измерению(p).

Пусть текст представлен ковариантным тензором первого ранга D_k. Матрица семантических связей позволяет отобразить D_k - исходное представление текста, в новое представление контравариантного тензора первого ранга S^k, уже отражающее семантические связи между словами. Применим формулу 1.

Математически модель получения векторного семантически обогащенного представления можно представить в виде выражения:

$$S^k = D_k * R^{кк} \qquad (3)$$

Но проблему большой вычислительной сложности это не устранило, так как в больших

онтологиях количество концептов чрезвычайно велико и загрузить в память исходный тензор для выполнения дальнейших операций будет проблематично.

Чтобы справиться с большим количеством RDF-троек, а также с разреженностью данных, необходимо использовать технику редукции признакового пространства, т.е. переход к пространствам более низких порядков.

Редукция признакового пространства

Так как многие алгоритмы поиска и классификации информации чрезвычайно чувствительны ко времени вычисления, то проблеме уменьшения размерности пространства и получения качественной онтологической модели уделялось внимание еще на этапе извлечения знаний с помощью модуля автоматического разбора текста.

Для снижения размерности были использованы следующие алгоритмы:

1. Удалены все ненужные для анализа символы (пунктуационные знаки, цифры, т.д.).
2. Выполнялась токенизация (текст разбивался по пробелам для получения слов по отдельности).
3. Удалялись стоп-слова (междометия, союзы, вводные слова и т.д.).
4. Проверялась орфография (если возможно, исправлялась ошибка).
5. Проверялась часть речи слова (отбирались только существительные).
6. Все остальные слова приводились к единой форме (лемме).

Но эти меры не привели к ощутимым результатам. Поэтому актуальной является задача оптимизации признакового пространства, которая состоит в том, чтобы представить исходную информацию, заданную в виде большого количества признаковых описаний, в пространстве меньшей размерности, и, по возможности, минимизировать потери информации.

Данные в системе управления информационными ресурсами научно-образовательных организаций имеют вполне “разумную” размерность (1000-2000 текстовых документов). Несмотря на незначительное количество документов, размерность матрицы, полученной после векторизации текстов, может быть довольно значительной для хранения её в памяти ЭВМ (7000-50000 индексированных термов). Следует предусмотреть и увеличение размерности задачи хотя бы до 5000 документов.

Снижение размерности следует проводить до размерности, позволяющей осуществлять обработку без существенного ухудшения качества поиска.

Стандартные подходы снижения размерности исходного признакового пространства могут быть разделены на два больших класса:

– с трансформацией признакового пространства (PCA, SVD, NMF);

– без трансформации исходного пространства с исключением неинформативных признаков.

PCA (анализ главных компонент) – один из самых популярных методов сокращения размерности. Однако признаки PCA трудно интерпретировать и метод PCA непредсказуем в выборе координат объектов в новом пространстве.

Для уменьшения размерности также часто применяется сингулярное матричное разложение SVD и неотрицательное матричное разложение.

NMF предусматривает разложение в произведение матриц меньшего размера, но при этом элементы U и V неотрицательны. Результат – признаки, которые лучше интерпретируются и могут иметь больше смысла по сравнению с SVD. Проблема самих методов SVD и NMF заключается в том, что это очень медленные и ресурсоемкие алгоритмы. Среди подходов с уменьшением признаков без трансформации пространства весьма перспективным представляется использование тематических разделов и контекстных векторов онтологии для решения задачи снижения размерности.

Тематическая редукция на базе предметной онтологии системы

Часть проблем, связанных с производительностью и ресурсоемкостью системы, решается с помощью эффективно полученных представлений. В этом случае в процессе обработки мы работаем не с самими, иногда очень объёмными ИР, а с их структурированными представлениями, сохраненными в онтологии. Поскольку при создании СУИР был разработан целый блок для автоматического разбора и структурирования документов (см. рис. 4), а также блок для работы с онтологиями, то мы можем в полном объёме воспользоваться преимуществами уже структурированного представления документов для разработки эффективных моделей представления ИР текстового вида [18].

Рисунок 4 – Укрупненная структура СУИР

Использование ранее сохраненного представления документа вместо непосредственно самого документа позволяет избежать трудоемких процессов просмотра и анализа полного его содержания при каждой операции с ИР.

Пусть пользователь сформулировал некоторый запрос $Q=\{q_1, q_2, .. q_к\}$.

Семантический образ тематического раздела формируется на базе семантических образов отдельных концептов, принадлежащих данному разделу, и, соответственно, представляет собой частоты слов TF в ИР.

Тензор 2-ранга Z_t^k задает соответствие между концептами, являющимися ключевыми словами и темами. Редукцию будем выполнять по тематике из тензора Z_t^k, наиболее соответствующей пользовательскому запросу Q_k.

Для определения подходящей тематики, сначала выполним аддитивную свертку тензора Z_t^k с тензором первого ранга поискового запроса Q_k. Тема с максимальным значением веса из вектора G_t и будет наиболее подходящей, чтобы участвовать в редукции.

$$G_t \cong Z_t^k * Q_k \Rightarrow \max \quad (4)$$

Далее формируем тензор 3-го ранга T_p^{kk}, добавляя в него данные из RDF – хранилища по отношениям, в которые входили ключевые концепты выбранной темы. Таким образом получаем тематически редуцированный тензор, который содержит только семантически важные термины (см. рис. 5).

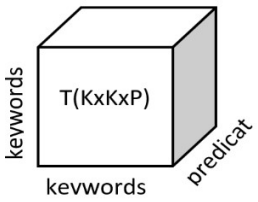


Рисунок 5 – Тематический редуцированный тензор

Существенное снижение размерности (до порядка k - размера контекстного вектора) было достигнуто за счет использования обобщенного модифицированного векторного представления документа на базе модели Word-of-Concepts и онтологической модели, а также за счет тематической редукции семантического тензора.

Редукция методом тензорных разложений

Использование тензоров может представлять структуру данных более естественно, чем матрицы. Например, трехмерный тензор является более удобной структурой представления семантических данных графа знаний, чем матрица.

Факторизация N-мерного тензора генерирует N матриц, состоящих из к - векторов для каждого измерения скрытого семантического пространства, и служит уникальным средством моделирования и выявления взаимосвязей и совместного поведения n переменных в массиве n-мерных данных [19].

Информационное пространство изобилует работами по технологии факторизации, включая латентный семантический анализ (Deerwester et al., 1990), вероятностные варианты LSA (Hofmann, 1999), Анализ основных компонентов и вероятностные и полиномиальные версии PCA (Buntine & Perttu, 2003; Tipping & Bishop, 1999) и совсем недавно неотрицательная матричная факторизация (NMF) (Paatero & Tapper, 1994; Lee & Seung, 1999)[20].

Для редукции признакового пространства семантического тензора будем использовать подход, основанный на тензорной факторизации. Факторизация данных в более низкое размерное пространство вводит компактный базис, который при соответствующей настройке может описать исходные данные в сжатой форме, ввести некоторую невосприимчивость к шуму.

Для случая NMF существует несколько алгоритмов и реализаций, но не существует эффективных реализаций, которые были бы распространены на более общий тензорный случай [21]. В некотором смысле метод неотрицательной факторизации тензоров можно назвать также n-мерным обобщением SVD, так как SVD матричное разложение для тензора второго порядка является частным случаем тензорного разложения [22]. В работе [22] также отмечается, что единого способа обобщения SVD на тензоры 3-го и более порядка не существует.

В работе [23] приведено большое количество типов тензорных разложений - CANDECOMP/PARAFAC, TUCKER, DEDICOM, INDSCAL, PARAFAC2, CANDELINC, PARATUCK2. Наиболее часто используется каноническое разложение (CP) и разложение Такера, а также их вариации.

Декомпозиция Tucker2 [23] тензора третьего порядка устанавливает одну из факторных матриц в качестве единичной матрицы. Аналогично, декомпозиция Tucker1 [23] устанавливает две факторные матрицы в качестве тождественной матрицы.

Очевидно, что существует множество вариантов тензорных разложений, что может привести к путанице в выборе модели для конкретного приложения. На базе разложения Такера описаны различные варианты разложений, в том числе и HOSVD, и представляет собой один из вариантов обобщения SVD. Но мы воспользуемся вариацией разложения Такера и получаем (см. рис. 6).

$$X^{oo}{}_{p} \cong D_p^{kk} * A^{o}{}_{k} * A^{o}{}_{k} * E, \quad \text{де} \qquad (5)$$

X(OxOxP) - тензор 3-го ранга, хранящий концепты и связи межу ними;

D(KxKxP) - тензор 3-го ранга, хранящий вектора ключевых слов по темам и связи межу ними;

A(OxK) - тензор 2-го ранга, хранящий вектора ключевых слов каждого концепта;

E - единичный тензор;

O – количество терминов;

P – количество отношений;

K – количество ключевых терминов.

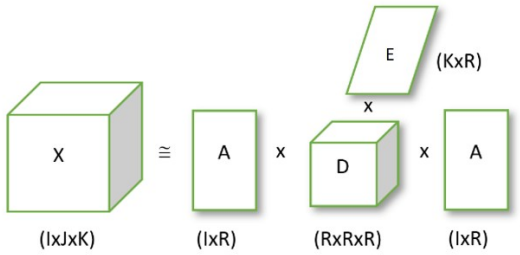


Рисунок 6 – Тензорное разложение Tucker2

Поскольку K гораздо меньше по значению, чем O, то тензор D можно рассматривать как сжатую версию X. Таким образом, мы осуществляем редукцию тензора X, что существенно уменьшает размерность пространства и увеличивает семантику подобно методу сингулярного разложения матриц SVD, используемых в LSA/LSI.

Эксперименты

Перед реализацией программных модулей был выполнен ряд экспериментов на прототипах для тестирования основных алгоритмов и моделей. Для тестирования использовался ПК со средними на текущий момент параметрами: Процессор Intel(R) Core(TM) i3-4010U CPU @ 1.70GHz, 1700 МГц, ядер: 4, логических процессоров: 4, оперативная память (RAM) 8,00 ГБ. Прототипы были разработаны в программной среде Python 3.8, на базе нескольких стандартных библиотек. Для проведения экспериментов были использованы собственные ИР - материалы конференций факультета за последние 10 лет в объеме более 1000 документов. С помощью модуля автоматической обработки тексты документов были распарсены, сохранены в онтологии, а также были получены и сохранены векторные представления документов.

Сравнивалась обработка различных моделей BOW: ONTO – тематически редуцированная модель и TEN – обобщенная на трехмерный случай SVD модель на базе факторизации и тензорных разложений.

Основные параметры модели опишем в табл. 1.

Таблица 1 - Параметры и переменные модели

№	Пар-р	Описание параметров
1	o	Количество объектов, извлеченных из RDF- хранилища
2	s	Количество субъектов, извлеченных из RDF- хранилища
3	p	Количество типов предикатов, извлеченных из RDF- хранилища
4	k	Длина контекстного вектора
5	w	Контекстный вектор
6	l	Семантическая оценка концепта
7	X	Исходный тензор модели
8	A, B, C, D	Тензоры, полученные в результате разложения Таккера

В процессе экспериментов изменялись такие параметры, как количество концептов модели, количество семантических мер модели (пока реализовано 8 мер СБ), размер контекстного вектора (от 0 до 100) с целью оценить, насколько эти параметры влияют на производительность системы.

Результаты эксперимента, как влияют на производительность операции с многомерным тензором были сведены в табл. 2.

Таблица 2 - Результаты экспериментов

p	**Время выполнения алгоритмов для o=5000 к=30**	
	Ввод данных TEN	Обработка данных TEN
	T1, *s*	**T2, *s***
1	0,1456	0,076
2	8,456	0,876
3	10,8002	1,652
4	10,8972	1,887
5	20,8992	2,09
6	20,0997	2,431
7	20,4991	2,786
8	30,399	3,876
Σ	122,0504	15,598

Можно сделать вывод, что изменение параметра p от 1 до 8 хотя и замедляет обработку в среднем на два порядка, но не влияет критично на производительность системы в заданных размерах тензора.

Проведем исследования, как влияет количество входных термов, а, следовательно, и размеры исходной матрицы, на скорость выполнения базовых операций: ввода данных и обработки данных (см. табл. 3).

Таблица 3 - Результаты экспериментов

P=8 k=20	**Время выполнения функциональных блоков алгоритмов**			
	Процедура ввода данных		Процедура обработки данных	
	ONTO	**TEN**	**ONTO**	**TEN**
Кол.	**T1, s**	**T2, s**	**T3, s**	**T4, s**
0	0,005	0,002	0,000100	0,09495
750	0,00899	0,1679	0,000001	0,01099
1500	0,01299	0,61916	0,000099	0,006
2250	0,02399	1,56565	0,000001	0,008
3000	0,02699	6,27798	0,001004	0,08495
3750	0,03998	8,40323	0,001004	0,09495
4500	0,03598	13,1394	0,000999	0,64206
5250	0,004	11,9704	0,001000	0,49979
6750	0,006	Ошиб-ка!	0,020998	---------
7500	0,008		0,000001	
8250	0,00899		0,003004	
9000	0,008		0,001004	
9750	0,00899		0,000999	
Σ	0,16392	42,14572	0,025206	1,44169

Как показали результаты анализа процедур ввода, алгоритм на базе модели ONTO за счет предварительно выполненной редукции работает быстро даже при больших размерах входного тензора.

Совсем другая ситуация наблюдается для алгоритма TEN, когда при выполнении операций загрузки с объёмом тензора 6000x6000x8 при тестировании происходит аварийное завершение работы, связанное с переполнением памяти (см.рис.6).

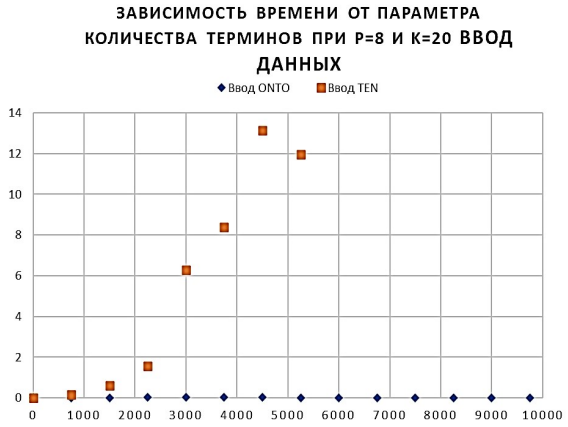


Рисунок 6 – Время выполнения процедур ввода при увеличении количества терминов

Сравним отдельно скорость выполнения процедур ввода данных при использовании моделей TEN и ONTO. При выполнении процедуры ввода для алгоритма TEN среднее время ввода начинает резко увеличиваться, начиная где-то при количестве терминов более 2500 (см. рис. 7).

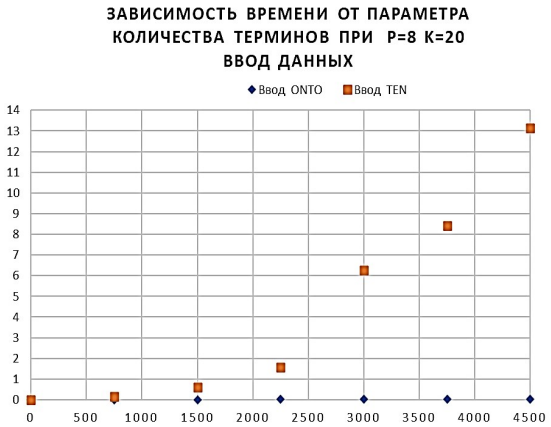


Рисунок 7 – Время выполнения процедур ввода при увеличении количества терминов

Сравним отдельно алгоритмы обработки данных TEN и ONTO по скорости выполнения. Среднее время обработки по алгоритму с использованием модели TEN на несколько порядков больше, чем при использовании модели ONTO, для которой тестирование выполнялось практически мгновенно (см. рис. 8).

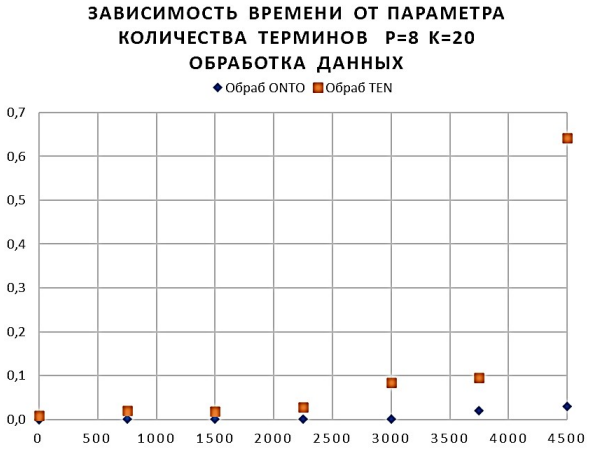


Рисунок 8 – Время выполнения процедур обработки при увеличении количества терминов

Конечно, при использовании более мощной вычислительной техники результаты будут улучшены, но для большинства организаций этот вариант не является приемлемым из-за удорожания стоимости системы в целом.

Выводы

В ходе исследований была разработана обобщенная модифицированная модель на базе известных подходов bag-of-words и bag-of-concepts, улучшенная за счет использования онтологической модели предметной области и

техники редукции по тематическим векторам, что одновременно снизило размерность задачи и увеличило семантику.

Исследования показали, что для больших текстовых массивов ИР организаций за счет высокой требовательности к объёмам оперативной памяти не могут быть использованы обобщенные на трехмерный случай модели на базе NMF и SVD в “чистом” виде. Только применение в качестве представлений тематически редуцированных тензоров позволяет значительно сократить затраты памяти и вычислительных ресурсов и обойти сложности обработки исходного тензора.

Полученные результаты подтверждают оправданность применения разработанной модели представления текстовых документов.

Литература

1. Андриевская Н.К. Основные принципы и подходы при разработке системы управления профессиональными знаниями ВУЗа / Н.К. Андриевская // Информатика и кибернетика. 2019. № 4 (18), с 49-56.

2. Андриевская Н.К. Онтологический подход в системах обработки данных научных и научно-образовательных организаций / Н.К. Андриевская // Проблемы искусственного интеллекта. 2020. № 1 (16), С. 23-36.

3. WordNet. Лексическая база данных для английского языка. - Режим доступа: https://wordnet.princeton.edu/ (дата обращения 28 декабря 2020).

4. Андриевская Н.К. Разработка прикладной онтологии в системах обработки данных научных и научно-образовательных организаций / Н.К. Андриевская // Вестник Донецкого национального университета. Серия Г: Технические науки. 2020. №3. – С.43-51.

5. Канатуш С. В. Онтологический подход к веб-поиску / С.В. Канатуш, Н.К. Андриевская // Проблемы радиоэлектроники и телекоммуникаций : сб. науч. тр. / под ред. Ю. Б. Гимпилевича. — Москва-Севастополь : Изд-ва : РНТОРЭС им. А.С. Попова, СевГУ, 2020. — № 3. — 247 с., С. 205— ISSN 2658-6347.

6. Melucci, Massimo. (2009). Vector-Space Model. Encyclopedia on Database Systems.

7. TF-IDF [Электронный ресурс]: Википедия. Свободная энциклопедия. – Режим доступа: https://ru.wikipedia.org/wiki/TF-IDF (дата обращения 28 декабря 2020).

8. Word2vec [Электронный ресурс]: Википедия. Свободная энциклопедия. – Режим доступа: https://ru.wikipedia.org/wiki/Word2vec (дата обращения 28 декабря 2020).

9. Нугуманова А.Б. Обогащение модели Bag of words семантическими связями для повышения качества классификации текстов предметной области / А.Б. Нугуманова, И.А. Бессмертный, П. Пецина, Е.М. Байбурин // Программные продукты и системы. 2016. № 2, с. 89-99.

10. Андриевская Н.К. Анализ возможностей использования существующих словарей для пополнения онтологии / Н.К. Андриевская, А.И. Секирин, С.В. Канатуш // Информатика и кибернетика. 2020. №2(20), с13-21.

11. Адамов Б.И. Применение основных матричных разложений в задачах механики и робототехники / Б.И. Адамов, А.Н. Маслов, Н.В. Осадченко. - М.: Издательство МЭИ, 2019. - 84 с.

12. Silva C., Ribeiro B. (2009) Knowledge Extraction with Non-Negative Matrix Factorization for Text Classification. In: Corchado E., Yin H. (eds) Intelligent Data Engineering and Automated Learning - IDEAL 2009. IDEAL 2009. Lecture Notes in Computer Science, vol 5788. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04394-9_37

13. RDF - Semantic Web Standards [Электронный ресурс]: w3.org. – Режим доступа: https://www.w3.org/RDF / (дата обращения 28 декабря 2020).

14. Nickel, M. et al. A review of relational machine learning for knowledge graphs / M. Nickel et al. // Proceedings of the IEEE. – 2015. – Vol.104. No. 1. - P.11-33.

15. Nickel M., Tresp V., Kriegel H. P. A Three-Way Model for Collective Learning on Multi-Relational Data // ICML. 2011. Vol. 11.

16. Вильчевская Е.Н. Тензорная алгебра и тезорный анализ: учеб. пособие / Е.Н. Вильчевская. — СПб. : Изд-во Политехн. ун-та, 2012. 4 с.

17. Kossaifi, Jean; Panagakis, Yannis; Anandkumar, Anima; Pantic, Maja (2019). "TensorLy: Tensor Learning in Python". JMLR. 20 (26): 1–6.

18. Андриевская Н. К. Разработка архитектурной модели системы управления информационными ресурсами организаций / Н.К. Андриевская, А.И Секирин, О.В. Ченгарь // Программная инженерия: методы и технологии разработки информационно-вычислительных систем (ПИИВС-2020): сборник научных трудов III Международной научно-практической конференции, Том. 1. 25-26 ноября 2020 г. – Донецк, ГОУВПО «Донецкий национальный технический университет», 2020. – 169 с. – С. 46-54.

19. Lathauwer, L& DeMoor, Bart& Vandewalle, Joos. (2000). Multilinear Singular Value Tensor Decompositions.SIAM J.Matrix Anal. Apl.24.

20. Amnon Shashua and Tamir Hazan. 2005. Non-negative tensor factorization with applications to statistics and computer vision. In Proceedings of the 22nd international conference on

Machine learning (ICML '05). Association for Computing Machinery, New York, NY, USA, 792–799. DOI:https://doi.org/10.1145/1102351.1102451

21. Department of Computer Science Technical Report TR-2006-21 October 2006, University of British Columbia Computing nonnegative tensor factorizations Michael P. Friedlander* Kathrin Hatz† October 19, 2006 for computing the NTF of a dataset.

22. Ma, S., Jiang, B., Huang, X., & Zhang, S. (2016). Tensor models: Solution methods and applications. In S. Cui, A. Hero, III, Z. Luo, & J. Moura (Eds.), Big Data over Networks (pp. 3-36). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316162750.002

23. Kolda Tamara G., Bader Brett W. Tensor Decompositions and Applications // SIAM Rev., 51(3), 455–500.

Андриевская Н.К. Обобщенная модифицированная модель представления текстовых информационных ресурсов. *Данная статья посвящена разработке новых эффективных подходов к формированию представлений текстовой информации при реализации задач информационного поиска в различных информационных системах (ИС). Общей проблемой является разреженность и большие объёмы входных данных, что приводит к необходимости уменьшить вычислительную сложность и повысить производительность методов основных процедур обработки данных. Для решения проблемы авторами предлагается использовать обобщенную модифицированную модель представления текста, а также предварительно выполнить редукцию по тематическим векторам, что одновременно уменьшает размерность и увеличивает семантику представления текста. Кроме этого, для усиления семантики предлагается использовать предметную онтологию системы. Это дает возможность многократно использовать ранее обработанные документы в структурированном виде и использовать уже существующие контекстные вектора в процедурах обработки.*

Ключевые слова: *векторная модель, сингулярное разложение, факторизация, модель bag-of-words, модель bag-of-concepts, LSA, NMF, семантические связи, RDF-граф, тензор, онтология.*

Andrievskaya N. K. Generalized modified model of the text word embedding for the textual information resources. *This article is devoted to the development of new effective approaches to the formation of words embedding of text information in the implementation of information search tasks in different information systems (IS). Sparsity and large amounts of input data is a common problem that needs to be solved to reduce computational complexity and improve the performance of methods of basic data processing procedures. The authors propose to use a generalized modified model of text word embedding to solve the problem, as well as to perform a reduction by thematic vectors, which will simultaneously reduce the dimension and increase the semantics of text representation. In addition, to strengthen the semantics, it is proposed to use the subject ontology of the system. This will make it possible to reuse previously processed documents in a structured form and use already existing context vectors in processing procedures.*

Keywords: *vector model, singular value decomposition, factorization, bag-of-words model, bag-of-concepts model, LSA, NMF, semantic relations, RDF graph, tensor, ontology.*

Статья поступила в редакцию 05.12.2020
Рекомендована к публикации профессором Павлышом В. Н.

УДК 004.7

Математическая модель перегрузки компьютерной сети

Д. В. Бельков, Е. Н. Едемская
Донецкий национальный технический университет
belkovdv@list.ru

Аннотация
Важной проблемой в компьютерных сетях является перегрузка. Это состояние, при котором сетевые ресурсы на протяжении достаточно длительного интервала времени не способны обрабатывать поступающие пакеты. Перегрузка в сетях на основе протокола TCP/IP возникает из-за отсутствия централизованного управления сетевыми ресурсами. При передаче трафика в сети можно выделить три фазы: фаза без перегрузки, фаза управляемой перегрузки, фаза серьезной перегрузки. Целью данной статьи является уменьшение времени доставки пакетов за счет регулирования ширины фазы управляемой перегрузки. В работе предложена математическая модель перегрузки компьютерной сети, построенная на основе волнового процесса. Для определения ширины фазы управляемой перегрузки использован алгоритм RED и уравнение Бюргерса.

Введение

В компьютерных сетях важной проблемой является перегрузка. Это состояние, при котором сетевые ресурсы на протяжении достаточно длительного интервала времени не способны обрабатывать поступающие задания [1,2]. В сетях на основе протокола TCP/IP перегрузки возникают из-за отсутствия централизованного управления сетевыми ресурсами. Так как пакеты обрабатываются последовательно, то в случае одновременного поступления нескольких пакетов необходимо применение механизма, определяющего порядок их обработки и передачи, например, пакеты, ожидающие обработки, могут помещаться в буферную память маршрутизатора и далее обрабатываться в соответствии с дисциплиной FIFO.

Применение механизма буферной памяти в сетевых маршрутизаторах и серверах позволяет обрабатывать весь сетевой трафик при незначительном увеличении его интенсивности. Однако, поскольку, буферная память конечна, существенный рост интенсивности передачи приводит к потерям пакетов.

Проблема перегрузки не может быть разрешена реализацией принципа неограниченной буферной памяти. В таком случае неограниченно возрастает длина очереди пакетов, и, как следствие, неограниченно увеличиваются задержки передачи пакетов от отправителей к получателям. Кроме того, время, выделяемое на обработку одного пакета в компьютерной сети, ограничено, и при перегрузке сети часть пакетов будет потеряна, что потребует их повторной передачи [3]. Таким образом, существенное увеличение объема буферной памяти является неэффективным.

К перегрузкам в компьютерных сетях могут приводить следующие ситуации [4]:

– существенное увеличение объема передачи данных по одному каналу при одновременной работе нескольких пользователей);

– если время передачи превышает некоторое критическое значение TCP-протокол повторно передает пакеты, которые уже находятся в процессе передачи или приняты получателем;

– потеряны пакеты, которые некоторое время передавались по сети, но не были доставлены получателю;

– фрагментация пакетов, когда в сети передаются фрагменты, которые будут в дальнейшем удалены, так как из них невозможно восстановить исходный пакет.

Результатом перегрузок в компьютерной сети является высокий уровень потери пакетов, существенное увеличение времени их доставки от отправителя к получателю, временная недоступность сетевых ресурсов. Таким образом, актуальной является разработка и применение механизмов управления перегрузкой в компьютерных сетях, которые позволят избежать потенциальных перегрузок.

Система динамического управления потоками данных является обязательным элементом сетевого оборудования. Под управлением потоками понимается совокупность механизмов, управляющих доступом к сетевым ресурсам и обеспечивающих согласование параметров сети и пользовательского трафика.

Известно [5], что при передаче трафика в сети можно выделить три фазы: фаза без перегрузки, фаза управляемой перегрузки, фаза серьезной перегрузки.

Целью данной статьи является уменьшение времени доставки пакетов за счет

регулирования ширины фазы управляемой перегрузки. Задачи работы:

– разработка математической модели перегрузки компьютерной сети на основе волнового процесса;

– определение ширины фазы управляемой перегрузки с использованием уравнения Бюргерса;

– определение ширины фазы управляемой перегрузки при различных моделях алгоритма RED.

Перегрузка компьютерной сети

Если количество пакетов, одновременно передаваемых по сети, превысит определенный пороговый уровень, производительность сети начинает снижаться. Такая ситуация называется перегрузкой. В случае низкой нагрузки трафик и коэффициент обслуживания сети растут с увеличением нагрузки. Количество доставленных пакетов пропорционально количеству посланных. Это равновесный режим. При повышении нагрузки достигается точка А, показанная на рис. 1, после которой сетевой трафик растет медленнее, чем нагрузка на входе. Это связано с тем, что сеть входит в состояние управляемой перегрузки, когда сеть справляется с нагрузкой с увеличенной задержкой (интервал между точками А и В на рис. 1).

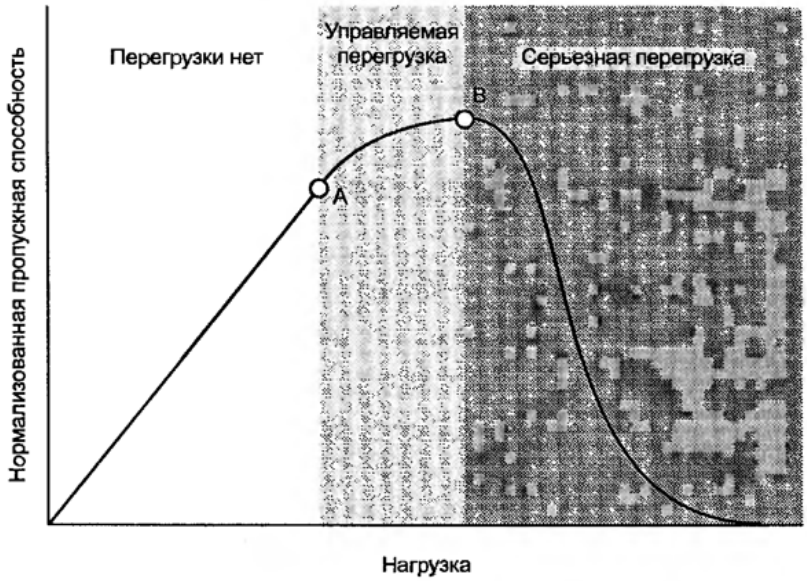


Рисунок 1 – Перегрузка сети [5].

Отклонение от равновесия вызвано следующими причинами. Во-первых, нагрузка распределяется по сети неравномерно. В то время, как отдельные узлы испытывают управляемую перегрузку, другие могут быть серьезно перегружены и перестать передавать трафик. Во-вторых, при увеличении нагрузки сеть пытается ее сбалансировать, направляя пакеты через участки с меньшей нагрузкой. Чтобы изменение маршрутов было возможно, узлы должны обмениваться большим количеством управляющих сообщений. Эти накладные расходы снижают пропускную способность.

При дальнейшем увеличении нагрузки очереди в узлах продолжают удлиняться. В точке В на рис. 1 достигается состояние перегрузки. С ростом нагрузки пропускная способность сети падает. Причиной этого является ограниченный объем буферов в узлах. Если буфер заполняется, узел вынужден терять пакеты. В таком случае отправитель должен передавать потерянные пакеты повторно, из-за чего нагрузка на систему возрастает. С увеличением количества переполненных буферов эффективная пропускная способность стремиться к нулю [5].

Идеальный случай работы сети показан на рис. 2. Предполагается наличие бесконечных буферов и отсутствие накладных расходов.

В идеальном случае перегрузки нет. Как показано на рис. 3, сеть передает пакеты с максимальной скоростью равной пропускной способности.

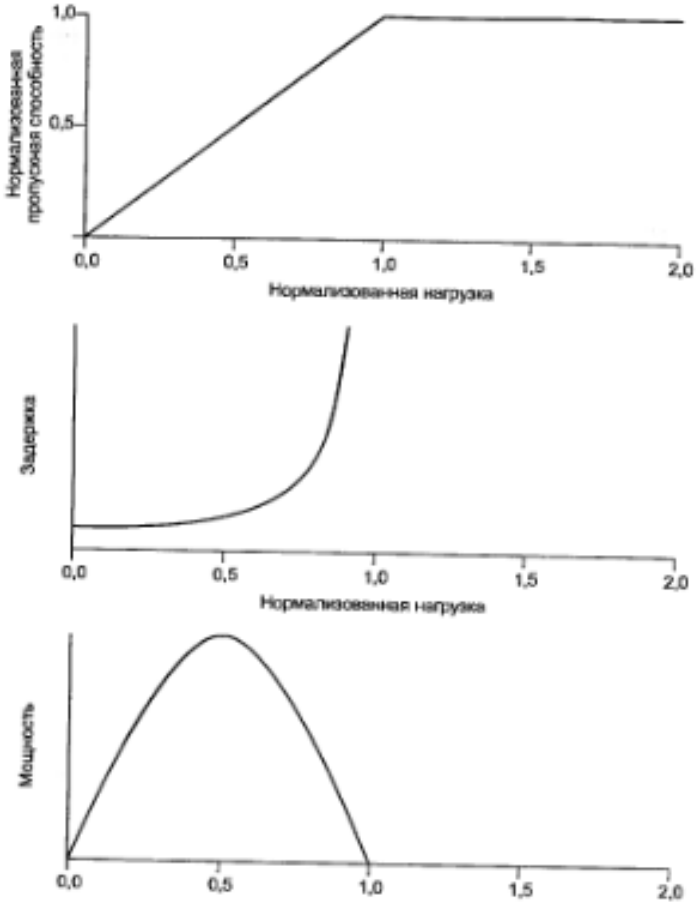


Рисунок 2 – Идеальный случай работы сети [5].

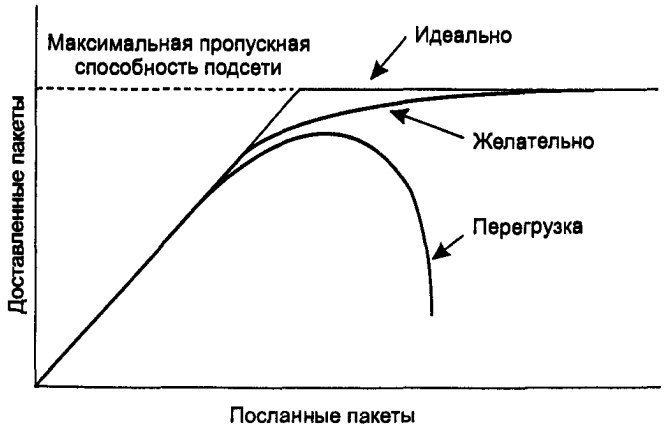


Рисунок 3 – Варианты работы сети [6].

Разработка модели перегрузки

В данном разделе рассматривается движение среды из невзаимодействующих частиц без потерь и появления новых частиц [7]. Пусть *n(x,t)* – плотность частиц в точке *x* в момент *t*, *v* – скорость среды: $v = dx/dt$. В таком случае движение подчиняется уравнению (1).

$$\partial n / \partial t + v \cdot \partial n / \partial x = 0 \qquad (1)$$

Если $v = v_0 = const$ – скорость волны, то решение уравнения (1) имеет вид: $n = n(x - v_0 t)$. Начальное условие *n(x,0)=F(x)* определяет профиль волны, который движется без искажений вдоль *x* со скоростью v_0 .

Пусть среда является нелинейной и скорость зависит от плотности: *v=v(n)*. В таком случае решение имеет вид простой волны:

$$n(x,t) = F(x - v(n)t) \qquad (2)$$

В волне (2) скорость движения разных точек профиля неодинакова. Если $v'(n) > 0$, то фронт волны становится все более крутым и волна опрокидывается. Момент опрокидывания определяется условиями (3).

$$t_0 = 1/\max(F'v'), F'v' < 0 \qquad (3)$$

Для уравнения свободного движения несжимаемой среды (4) эти условия имеют вид (5).

$$\partial v/\partial t + v \cdot \partial v/\partial x = 0 \qquad (4)$$
$$t_0 = 1/\max(v'), v' < 0 \qquad (5)$$

При опрокидывании производные n_x, n_t и v_x, v_t обращаются в бесконечность, т.е. наклон профиля становится перпендикулярным к оси x.

Пусть среда является вязкой с коэффициентом вязкости ν. В таком случае существует фактор противодействия опрокидыванию волны. Если уравнение (4) дополнить вязким членом, то получим уравнение Бюргерса (6).

$$v_t + v \cdot v_x = \nu v_{xx} \qquad (6)$$

Если значение t приближается к t_0, крутизна фронта волны растет и, следовательно, увеличивается производная $v_x(x,t)$. В результате вязкий член νv_{xx} станет большим и может компенсировать нелинейный член v_x. Появляется конкуренция противоположных процессов: увеличения крутизны фронта из-за нелинейности и уменьшения крутизны из-за вязкости. Как следствие конкуренции может возникнуть стационарное движение. При малой вязкости ($\nu << 1$) стационарный профиль устанавливается через большое время: $t \to \infty$.

Известно [8,9], что для уравнения Бюргерса асимптотическая форма профиля волны формируется инвариантом (7)

$$\int_{-\infty}^{\infty} dx v(x,t) = inv = I \qquad (7)$$

При малой величине ν интеграл (7) можно вычислить методом перевала. Если y_0 – точка перевала, то решение имеет вид:

$$v(x,t) \sim (x - y_0)/t \qquad (8)$$

Пусть x_0 – значение x, при котором v=0. Для вычисления x_0 подставим (8) в (7) и получим:

$$x_0 \sim \sqrt{2It}; \max(v(x,t)) \sim x_0/t = \sqrt{2I/t} \to 0 \qquad (9)$$

На рис. 4а показано асимптотическое решение уравнения Бюргерса в виде треугольной волны при $\nu \to 0$. При конечных значениях вязкости, как показано, на рис. 4б, имеется переходный слой шириной Δx.

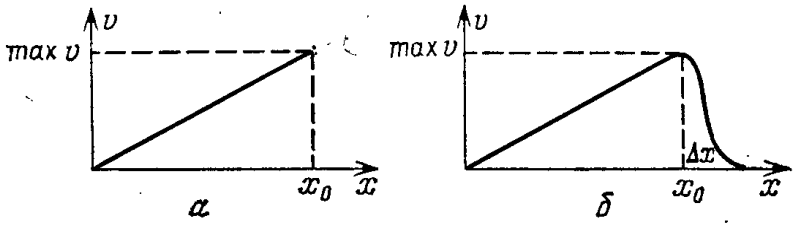


Рисунок 4 – Асимптотическое решение уравнения Бюргерса

Идея данной статьи состоит в том, что профиль волны, показанный на рис. 4б, отождествляется с перегрузкой компьютерной сети, показанной на рис. 3. В таком случае переходный слой Δx соответствует фазе управляемой перегрузки, а максимальная скорость волны – максимальной пропускной способности.

Решение уравнения Бюргерса, в котором опрокидывание не происходит, является примером образования ударной волны. Этот случай соответствует идеальной работе компьютерной сети.

Процесс движения волны определяется конкуренцией нелинейности и вязкости. Взаимоотношение этих факторов учитывается с помощью числа Рейнольдса, равного отношению нелинейного члена к вязкому: $R = vv_x/\nu v_{xx}$. Известно [9], что

$$\Delta x/x_0 \sim 1/R \qquad (10)$$

Формула (10) имеет практический смысл для предлагаемой аналогии между волновым процессом и перегрузкой компьютерной сети,

если параметрам $\Delta x, x_0$ найти аналогии в механизмах управления перегрузкой. В данной статье величине Δx ставится в соответствие рабочий интервал алгоритма RED [10,11].

Принцип действия алгоритма RED показан на рис. 5.

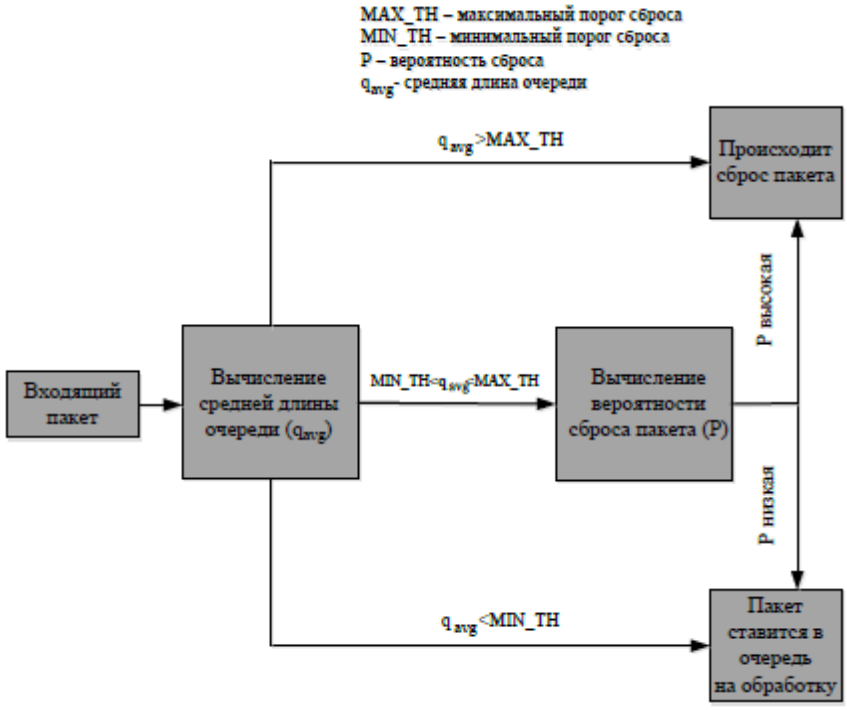


Рисунок 5 – Принцип действия алгоритма RED

Реализация RED является практически стандартной функцией маршрутизаторов TCP/IP, например, Cisco. Алгоритм состоит из двух частей: оценка среднего размера очереди и принятие решения о сбросе вновь поступившего пакета. При поступлении нового пакета, используя фильтр высоких частот совместно с «экспоненциально взвешенной скользящей средней» EWMA, RED определяет среднюю длину очереди avg. Затем величина avg сравнивается со значениями границ: верхней r_2 и нижней r_1, которые определяют степень нагрузки в очереди.

Если значение avg принадлежит интервалу $[0;\ r_1]$, то поступающий пакет помещается в очередь, если значение avg принадлежит интервалу $[r_1; r_2]$, то вычисляется вероятность сброса пакета P_a. С вероятностью P_0 пакет сбрасывается, либо с вероятностью $1-P_0$ он помещается в очередь. Если значение avg превышает r_2, то поступающий пакет сбрасывается. Таким образом, вероятность сброса пакета является функцией от переменной avg, т.е. с ростом нагрузки повышается вероятность сброса пакета.

Алгоритм RED имеет возможность работать в режиме маркировки пакетов. Маркировка заключается в изменении бита опционального поля заголовка пакета с целью извещения протокола транспортного уровня о надвигающейся перегрузке. В таком случае протокол приемника должен понизить размер окна и, тем самым, сообщить источнику о необходимости сокращения нагрузки. Источник должен понизить размер окна и, тем самым, понизить количество пакетов, передаваемых в сеть за некоторый промежуток времени.

Реализация маркировки пакета в RED более эффективна, чем реализация сброса в случае функционирования на транспортном уровне протокола TCP. Если осуществляется управление потоками UDP, то RED должен работать в режиме сброса.

Таким образом, RED состоит из двух частей. С помощью алгоритма вычисления среднего размера очереди фактически вычисляется степень пачечности нагрузки, которая может быть принята буфером. С помощью алгоритма подсчета вероятности маркировки или сброса пакета на основе известного значения нагрузки (т.е. значения средней длины очереди) определяется как часто маршрутизатор маркирует или сбрасывает

пакеты. Целью функционирования RED в маршрутизаторе является такая реализация маркировки/сброса пакетов, при которой будет соблюдаться принцип «справедливого распределения ресурсов».

В данной статье считается, что ширина фазы управляемой перегрузки и, соответственно, ширина переходного слоя Δx, совпадают с величиной $r_2 - r_1$, показанной на рис. 6. В таком случае загрузка сети измеряется средней длиной очереди. С учетом формулы (10), имеем

$$(r_2 - r_1)/r_1 = \Delta x / x_0 \sim 1/R \qquad (11)$$

Рекомендация [11], что значение r_2 должно быть в два раза больше, чем r_1, приводит к условию R=1, т.е. равенству нелинейного и вязкого членов и, соответственно, стационарному движению без опрокидывания (отсутствию фазы серьезной перегрузки). Таким образом, предлагаемая модель подтверждает

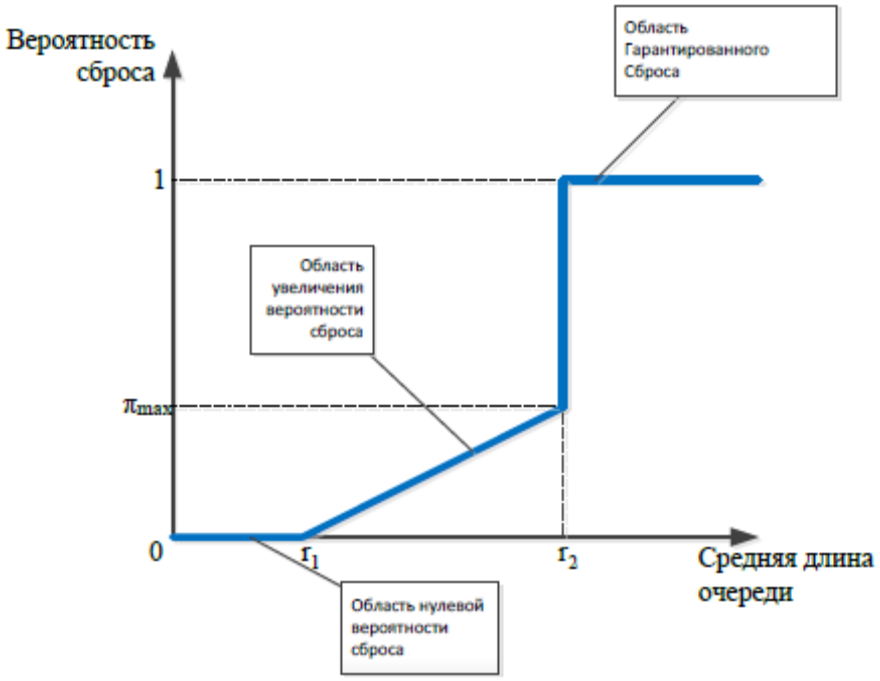


Рисунок 6 – Алгоритм RED [11]

Базовый алгоритм RED, описанный выше, не способен обеспечить обработку пакетов в соответствии с их приоритетами. Для маршрутизаторов, поддерживающих архитектуру DiffServ и AF PHB, было предложено модифицировать алгоритм RED для того, чтобы гибко управлять нагрузкой. Одним из основных свойств архитектуры DiffServ является возможность назначения пакетам различных приоритетов. Пакеты классифицируются на базе информации уровня IP по значению поля «тип услуги», находящегося в заголовке. В терминах архитектуры DiffServ данное поле называется DS. Назначение уровня приоритета осуществляется при поступлении пакета в домен DiffServ пограничным входящим узлом, в результате чего каждый пакет, находящийся в домене DiffServ, имеет свой приоритет и при поступлении в маршрутизатор в рамках этого домена должен быть обработан в соответствии со значением поля DS.

Текущие спецификации AF PHB определяют четыре класса с тремя уровнями приоритета пакета для каждого. Пакеты, принадлежащие одному классу, распределяются независимо от пакетов, принадлежащих другому классу. При этом в маршрутизаторе, находящемся в домене DiffServ, для каждого класса PHB AF должны быть назначены ресурсы. Таким образом в DiffServ-маршрутизаторах необходимо реализовывать алгоритмы активного управления очередями, которые способны обрабатывать пакеты в соответствии с их приоритетами.

Алгоритмы, основанные на RED и реализующие поддержку приоритетов пакетов, имеют общее название «многоуровневые RED» (MRED). Классификация MRED показана на рис. 7. Алгоритмы WRED и RIO показаны на рис. 8 и рис. 9.

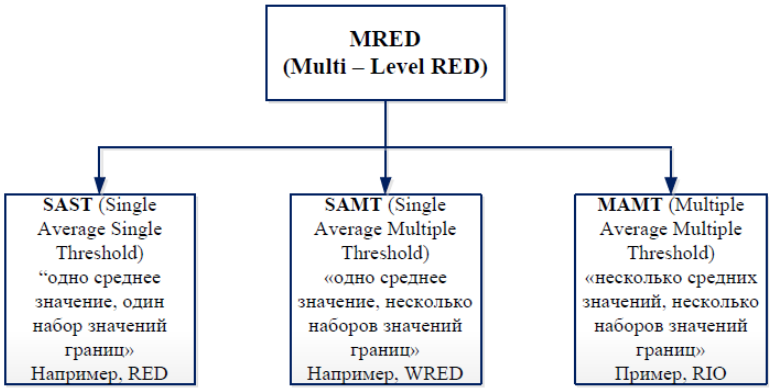


Рисунок 7 – Классификация MRED

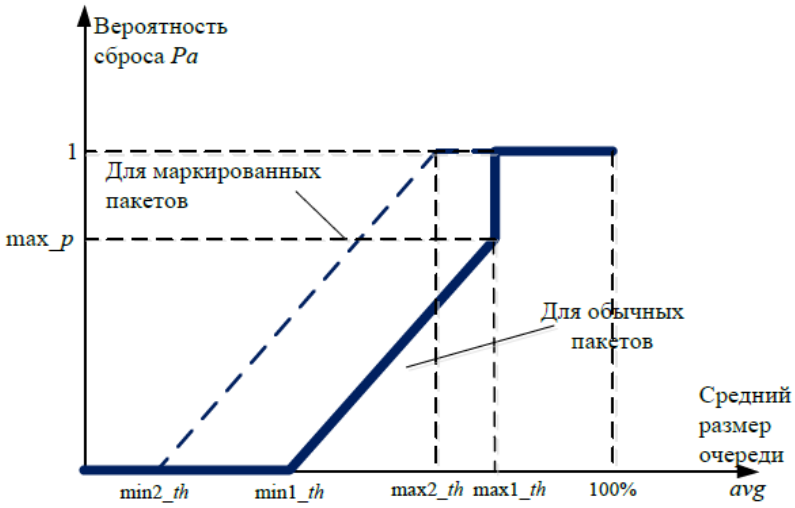


Рисунок 8 – Алгоритм WRED

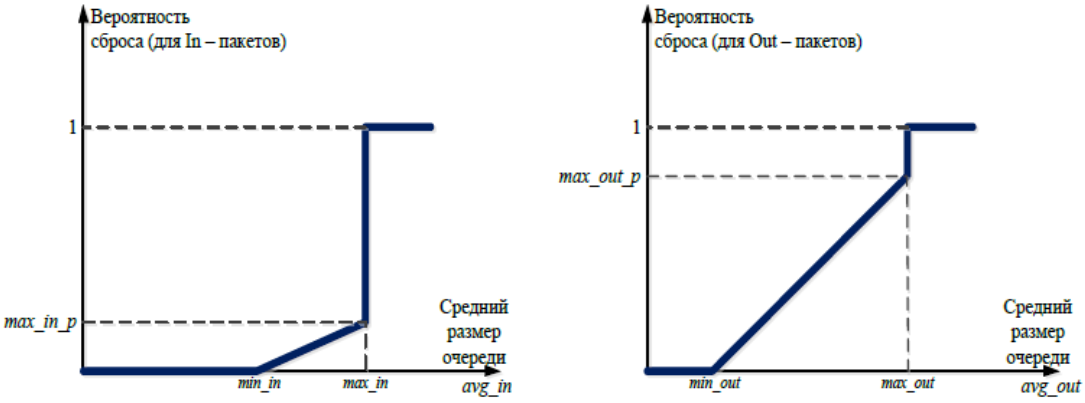


Рисунок 9 – Алгоритм RIO

Отличие WRED от RED состоит в том, что для пакетов каждого приоритета определен отдельный набор параметров. На рис. 8 показан случай, когда в очередь маршрутизатора поступают пакеты с двумя приоритетами. Пусть маркированный пакет является пакетом с низким приоритетом, а немаркированный пакет имеет высокий приоритет.

В соответствии с предлагаемой в статье моделью, для алгоритма WRED нужно рассматривать две волны с разными значениями числа Рейнольдса. Для немаркированных пакетов число Рейнольдса определяется по формуле (12), для маркированных пакетов - по формуле (13).

$$R_1 = (\max 1\_th - \min 1\_th)/\min 1\_th \quad (12)$$

$$R_2 = (\max 2\_th - \min 2\_th)/\min 2\_th \quad (13)$$

При использовании алгоритма управления очередью RIO (RED In&Out) предполагается, что существует два класса пакетов: немаркированные (in-profile) и маркированные (out-profile. Алгоритм RIO выполняет сброс маркированных пакетов с большей вероятностью, чем немаркированных.

Для алгоритма RIO, в соответствии с предлагаемой в статье моделью, рассматривается две волны с разными значениями числа Рейнольдса. Для in-пакетов число Рейнольдса определяется по формуле (14), для out-пакетов – по формуле (15).

$$R_{in} = (\max\_in - \min\_in)/\min\_in \quad (14)$$

$$R_{out} = (\max\_out - \min\_out)/\min\_out \quad (15)$$

В алгоритме DSRED, предложенном в работе [11] вводится дополнительное пороговое значение qmid между минимальным qmin и максимальным qmax порогами, которое вычисляется по формуле qmid=0,5 (qmax-qmin). Как показано на рис. 10, функция сброса описывается двумя линейными сегментами с углами наклона α и β соответственно, регулируемыми задаваемым селектором режимов γ. Для алгоритма DSRED в предлагаемой модели рассматриваются две волны с числами Рейнольдса, которые определяются по формулам (16) и (17).

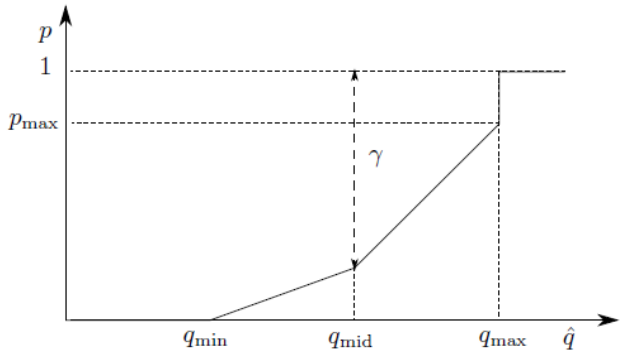


Рисунок 10 – Алгоритм DSRED

$$R_{q1} = (q_{mid} - q_{\min})/q_{\min} \quad (16)$$

$$R_{q2} = (q_{\max} - q_{\min})/q_{\min} \quad (17)$$

В алгоритме GRED, показанном на рис. 11, в рабочем диапазоне находится стандартная функция сброса RED. Однако существует дополнительный сегмент для обеспечения плавного поведения функции сброса за пределами рабочего диапазона.

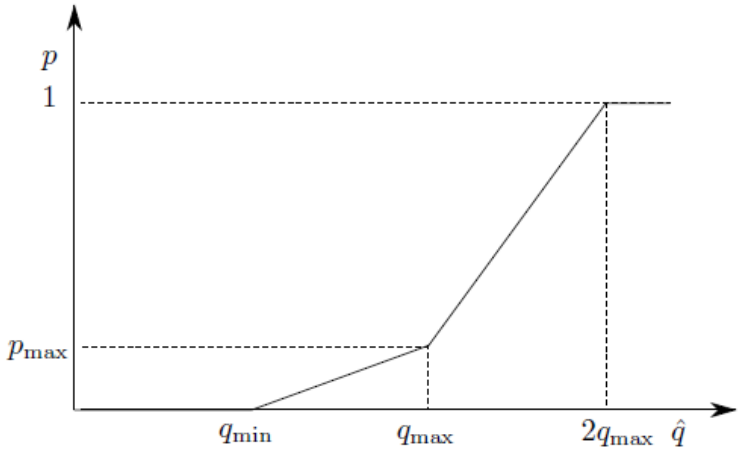


Рисунок 11 – Алгоритм GRED

Для алгоритма GRED в предлагаемой модели числа Рейнольдса для двух волн определяются по формулам (18), (19).

$$R_{q1} = (q_{\max} - q_{\min})/q_{\min} \quad (18)$$

$$R_{q2} = (2q_{\max} - q_{\min})/q_{\min} \quad (19)$$

Выводы

В работе получены следующие результаты:

– разработана математическая модель перегрузки компьютерной сети на основе волнового процесса;

– предложена формула для определения ширины фазы управляемой перегрузки с использованием уравнения Бюргерса;

– предложены формулы для определения ширины фазы управляемой перегрузки при различных моделях алгоритма RED.

Литература

1. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы. – СПб.: Питер, 2010. – 944 с.

2. Интернет: перегрузка. [Электронный ресурс], 2020. – Режим доступа: http://www.nonlin.ru/articles/podlazov/soc

3. Кучерявый Е.А. Управление трафиком и качество обслуживания в сети Интернет. [Электронный ресурс], 2020. – Режим доступа: http://litvik.ru/2/13/uchebniki_manuals/6575-upravlenie-trafikom-i-kachestvo-obsluzhivanija-v.html

4. Технология организации сетей. [Электронный ресурс], 2020. – Режим доступа: https://pandia.ru/text/78/654/44278-4.php

5. Столлингс В. Современные компьютерные сети. [Электронный ресурс], 2020. – Режим доступа: https://booksee.org/book/488197

6. Таненбаум Э. Компьютерные сети. Санкт-Петербург: Питер, 2012. – 991 с.

7. Заславский Г.М., Сагдеев Р.З. Введение в нелинейную физику: от маятника до турбулентности и хаоса. [Электронный ресурс], 2020. – Режим доступа: http://bookre.org/reader?file=543079

8. Рыскин Н.М., Трубецков Д.И. Нелинейные волны. Москва. – 2013. – 306 с.

9. Асимптотические решения уравнения Бюргерса. Ширина фронта ударной волны. [Электронный ресурс], 2020. – Режим доступа: https://studfile.net/preview/7395358/page:4/

10. Королькова А.В., Кулябов Д.С., Черноиванов А.И. К вопросу о классификации алгоритмов RED. [Электронный ресурс], 2020. – Режим доступа: https://www.researchgate.net/publication/235974583_K_voprosu_o_klassifikacii_algoritmov_RED

11. Киреева Н.В., Буранова М.А., Поздняк И.С. Изучение алгоритма RED в среде NS-2. [Электронный ресурс], 2020. – Режим доступа: https://www.sibsau.ru/sveden/edufiles/127542/

Бельков Д. В., Едемская Е. Н. Математическая модель перегрузки компьютерной сети. *Важной проблемой в компьютерных сетях является перегрузка. Это состояние, при котором сетевые ресурсы на протяжении достаточно длительного интервала времени не способны обрабатывать поступающие пакеты. Перегрузка в сетях на основе протокола TCP/IP возникает из-за отсутствия централизованного управления сетевыми ресурсами. При передаче трафика в сети можно выделить три фазы: фаза без перегрузки, фаза управляемой перегрузки, фаза серьезной перегрузки. Целью данной статьи является уменьшение времени доставки пакетов за счет регулирования ширины фазы управляемой перегрузки. В работе предложена математическая модель перегрузки компьютерной сети, построенная на основе волнового процесса. Для определения ширины фазы управляемой перегрузки использован алгоритм RED и уравнение Бюргерса.*

Ключевые слова: *перегрузка, волна, уравнение Бюргерса, ширина фазы управляемой перегрузки, алгоритм RED.*

Belkov D., Edemskaya E. Mathematical model of computer network congestion. *An important problem in computer networks is overload. This is a state in which network resources are unable to process incoming packets for a long enough period of time. Network overload based on the TCP/IP protocol is due to a lack of centralized network management. When transmitting traffic to the network, you can highlight three phases: phase without overload, phase of managed congestion, phase of serious overload. The purpose of this article is to reduce the delivery time of packets by regulating the width of the managed congestion phase. The paper offers a mathematical model of computer network congestion based on the wave process. The RED algorithm and the Burgers equation used to determine the width of the controlled congestion phase.*

Keywords: *congestion, wave, Burgers equation, part-managed overload width, RED algorithm.*

Статья поступила в редакцию15.09.2020
Рекомендована к публикации профессором Павлышом В. Н.

УДК 004.65

Исследование технологии шардинга для повышения эффективности обработки данных программного комплекса приемной комиссии ДонНТУ

О. Ю. Чередникова, С. В. Щедрин, А. Ю. Исаков, Е. А. Ногтев
Донецкий Национальный Технический университет
andreyyisakov@gmail.com, cherednikova@donntu.org, do010575ssv@gmail.com

Аннотация

В статье рассмотрено понятие шардинга, некоторые из его основных преимуществ и недостатков для решения проблемы масштабируемости АСУ приемной комиссии, связанной с ростом количества обрабатываемых данных. Также рассматриваются основы и принципы организации секционирования в объектно-реляционной СУБД PostgreSQL, а именно, что такое секции и секционированные таблицы, какие спецификации секционирования существуют и общий метод его реализации. Проведен анализ эффективности использования секционирования таблиц, в целях увеличения скорости выполнения запросов к ним, и возможности применения полученных результатов к таблицам базы данных АСУ приемной комиссии. Проведены эксперименты по установлению влияния на скорость выполнения SQL запроса выборки записей из таблицы в зависимости от количества записей в ней и от того, происходит ли выборка в пределах одной секции или на пересечении двух секций.

Ключевые слова: *SQL, СУБД PostgreSQL, масштабирование, шардинг, секционирование*

Введение

В условиях постоянно возрастающей нагрузки, связанной с увеличением объемов обрабатываемых АСУ приемной комиссии данных, затрудняется организация хранения и получения информации из базы данных. В связи с этим, падает уровень производительности, доступности и отказоустойчивости сервера БД. Одним из способов решения данной проблемы является масштабирование. Существует две основные стратегии — репликация и шардинг.

Репликация — это техника копирования данных на один или несколько других (называемые репликами) серверов. Ведущие сервера называют мастерами (master), ведомые - слейвами (slave). Сама по себе репликация не очень удобный механизм масштабирования и чаще всего используется из соображений надежности. Главным недостатком репликации, безусловно, является то, что если реплицируемый объект обновляется, то и все его копии должны быть обновлены (проблема распространения обновлений) [1]. В случае очень больших наборов данных или объемов обрабатываемой информации этого недостаточно: необходимо разбить данные на секции, иначе говоря, выполнить шардинг данных [2].

Шардинг — это шаблон масштабирования, суть которого заключается в разбиении больших таблиц базы данных на несколько меньших таблиц по выбранным критериям.

На каждом из шардов, хранится некоторая часть всего множества имеющихся данных, благодаря этому и реализуется масштабирование. Они могут размещаться как на одной, так и на нескольких машинах для балансировки нагрузки. Шардинг обычно начинают с наиболее крупных и нагруженных таблиц. Стоит отметить, что шардирование данных носит главным образом логический характер и оно используется в целях оптимизации выполнения запросов и для увеличения быстродействия сервера базы данных. Например, повышение скорости выполнения операции чтения происходит за счет того, что данные читаются из более узкого сегмента, а не из всей таблицы целиком. Шардинг бывает вертикальным и горизонтальным.

Вертикальный шардинг — это разделение одной таблицы на несколько меньших поколоночно (см.рис.1).

Горизонтальный шардинг — это разделение одной таблицы на несколько меньших построчно (см.рис.2).

Горизонтальный шардинг имеет одно явное преимущество перед вертикальным - он бесконечно масштабируем. Особенно о нём следует задумываться в тех случаях, когда в исходной таблице количество записей превышает нескольких миллионов или даже десятков, сотен миллионов.

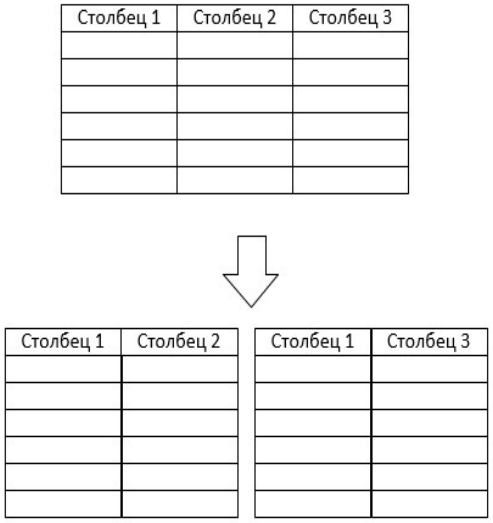


Рисунок 1 – Схема вертикального шардинга

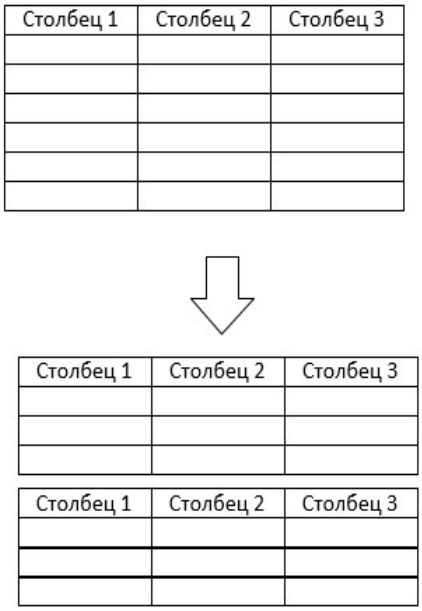


Рисунок 2 – Схема горизонтального шардинга

В данной статье будет исследован только горизонтальный шардинг.

Шардинг в PostgreSQL

В качестве системы управления базами данных, в программном комплексе приемной комиссии, используется объектно-реляционная СУБД PostgreSQL 9.5. В настоящее время PostgreSQL считается наиболее совершенной свободно распространяемой СУБД, которая конкурирует с лучшими из проприетарных.

Сильными сторонами PostgreSQL считаются:

– высокопроизводительные и надёжные механизмы транзакций и репликации;

– расширяемая система встроенных языков программирования;

– легкая расширяемость;

– поддержка наряду с базовыми типами данных, композитных типов данных, массивов и сложных типов (hstore, xml, json, jsonb);

– наследование.

В PostgreSQL наследование можно использовать для определения таблиц [3]. Именно на механизме наследования таблиц в PostgreSQL 9.5 и реализован горизонтальный шардинг [4]. PostgreSQL предлагает метод разделения, основанный на секционировании таблиц [5]. Система секционирования в PostgreSQL была впервые добавлена в версии PostgreSQL 8.1 основателем 2ndQuadrant Саймоном Риггсом.

В PostgreSQL 9.5 секционирование можно реализовать по следующим двум критериям:

– секционирование по диапазонам значений;

– секционирование по точному списку значений.

При секционировании по диапазону значений (range), строки таблицы распределяются по секциям в зависимости от значения какого-либо ключа [6]. Таблица разбивается по ключевому столбцу или набору столбцов, без перекрытия диапазонов значений, отнесенных к различным секциям. Наиболее очевидными примерами является секционирование данных по диапазонам идентификаторов (первый миллион записей, второй миллион записей, и так далее) или по дате (основываясь на годе, в котором была создана запись).

При секционировании по списку значений (list), явно указываются значения ключа, которые должны относиться к каждой секции.

Секционирование таблицы позволяет базе данных делать интеллектуальную выборку - сначала СУБД уточнит, какой секции соответствует запрос и только потом выполнит этот запрос, применительно к одной или нескольким нужным секциям [7]. Это позволяет значительно увеличить быстродействие, особенно когда большой процент часто запрашиваемых строк таблицы относится к одной или лишь нескольким секциям. Секционирование может сыграть роль ведущих столбцов в индексах, что позволит уменьшить размер индекса и увеличить вероятность нахождения наиболее востребованных частей индексов в памяти. Грамотно спроектированные секции в дальнейшем позволят осуществлять массовую загрузку и удаление данных, добавляя и удаляя секции, если это было предусмотрено.

Например, массовое удаление может быть произведено путем удаления одной или нескольких секций (DROP TABLE гораздо быстрее, чем массовый DELETE), а редко используемые данные могут быть перенесены в другое хранилище.

Постановка задачи

В связи с ростом объема данных, возникает необходимость в реализации и исследовании эффективности горизонтального шардинга, применительно к АСУ приемной комиссии. В базе данных АСУ имеются таблицы содержащие большое количество строк, хранящихся в произвольном порядке. Поиск в них нужных данных, по заданному критерию путем последовательного просмотра таблицы строка за строкой, может занимать много времени. При этом запросы на чтение, в большинстве случаев приходятся только на самую последнюю их часть (т.е. активно читаются те данные, которые недавно появились).

Примером тому может служить таблица личных данных абитуриентов. В ней хранится информация обо всех, подавших документы, абитуриентах, как в текущем году, так и в предыдущих. Но большинство запросов будет приходиться только на работу с данными за последний год, так как сотрудникам приемной комиссии чаще приходится работать с данными в рамках текущей приемной компании. Следовательно, в рассмотренном случае, секционирование позволит распределить нагрузку на таблицу по её секциям. И таким образом, можно будет добиться повышения скорости выполнения запросов к последней секции, которая значительно меньше основной таблицы, что дает ряд преимуществ и позволит улучшить производительность системы в целом.

В рамках данного исследования используется стратегия разбиения данных по диапазонам дат, так как её довольно легко понять и количество строк в таблице будет достаточно стабильным, что позволяет более гибко распределять нагрузку. Для проведения исследования создадим таблицу in_attr, от которой будут наследоваться все секции. Столбцы таблицы приведены на рис. 3.

```
CREATE TABLE public.in_attr
(
    id integer NOT NULL DEFAULT nextval('in_attr_id_seq'::regclass),
    lastname character varying(255),
    name character varying(255),
    facultid integer,
    priemid date,
    CONSTRAINT in_attr_pkey PRIMARY KEY (id)
)
```

Рисунок 3 – Столбцы таблицы in_attr

В таблице in_attr пять столбцов:

– id – идентификатор записи в таблице;

– lastname – фамилия абитуриента;

– name – имя абитуриента;

– facultid – идентификатор факультета;

– priemid – дата подачи документов абитуриентом.

После чего добавим в таблицу 1000000 записей и выполним запрос на выборку с условиями отбора по дате (рис.4).

```
explain analyze select * from in_attrr where priemid >= '2019-01-01';
----------------------------------------
"Seq Scan on in_attrr (cost=0.00..22810.00 rows=99497 width=54) (actual time=129.031..148.385 rows=100000 loops=1)"
"  Filter: (priemid >= '2019-01-01'::date)"
"  Rows Removed by Filter: 900000"
"Planning time: 0.100 ms"
"Execution time: 153.608 ms"
```

Рисунок 4 – Запрос на выборку с условиями отбора по дате

Анализ планировщика показал следующие значения, полученные в результате исполнения запроса:

– Seq Scan – таблица in_attr сканировалась последовательно;

– cost – условная величина, предназначенная для оценки времени, которое затрачивается на извлечение одного блока размером 8 Кб. Первое значение (0.00) – затраты времени до момента начала возврата первых результатов. Второе (22810.00) – затраты времени до того, как будет возвращен весь результат;

– rows (99497) – предполагаемое количество возвращаемых строк;

– width (54) – средний размер одной строки в байтах;

– actual time (126.031..145.385) – реальное время потраченное на выполнение запроса. В этом же блоке присутствуют значения rows (100000) и loops (1) – это реальное количество строк, которое вернул запрос, и количество итерации, которое было проделано в ходе его выполнения.

– Filter (priemid >= '2019-01-01'::date) – условие выборки для неиндексированных полей в таблице;

– Rows Removed by Filter (900000) – количество отброшенных строк в результате действия фильтра;

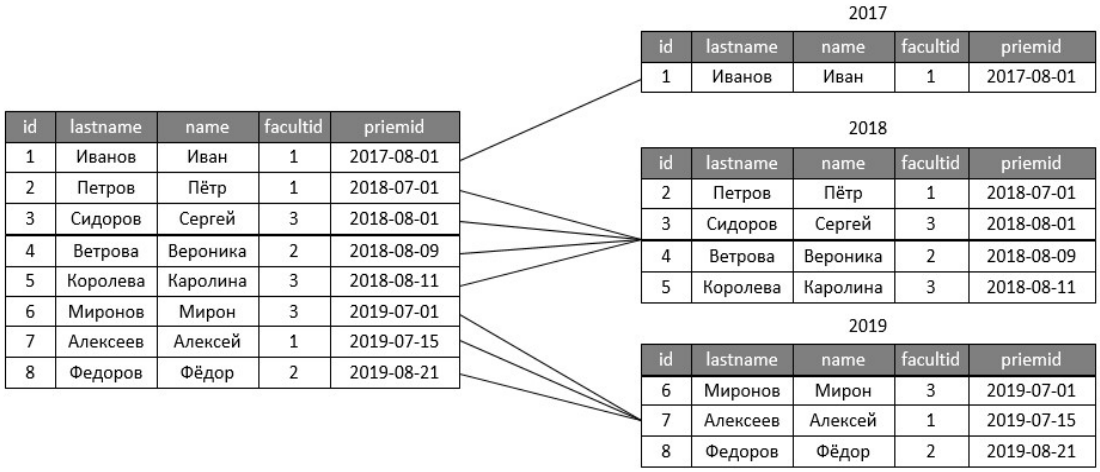
– Planning time (0.100 ms) – время планирования запроса;

– Execution time (149.608 ms) – время и выполнения запроса.

Секционирование

Теперь создадим десять дочерних таблиц, в которых не будет никаких дополнительных столбцов, в ней будут только столбцы, унаследованные от родительской таблицы in_attr. И добавим дочерним таблицам ограничения, по которым значения секций будут фильтроваться при выполнении запроса к ним. Стоить заметить, что диапазоны значений секций не должны пересекаться. В данном случае, секционирование производим по дате, а именно по годам приема документов от абитуриентов, с 2010 по 2019 включительно, в результате получаем десять диапазонов значений. Поле priemid является секционным ключом.

Для удобства заполнения секций данными, создадим триггера для главной таблицы и триггерную функцию, которая позволит вставлять данные в родительскую таблицу in_attr, и при этом они будут автоматически перенаправляться в нужные секции, в зависимости от значений секционного ключа. При попытке вставить запись, значение секционного ключа которой не попадает ни в один из существующих диапазонов, будет выведено сообщение об ошибке. На рисунке 5, приведен принцип разбиения тестовой таблицы на секции по годам.



id	lastname	name	facultid	priemid
1	Иванов	Иван	1	2017-08-01
2	Петров	Пётр	1	2018-07-01
3	Сидоров	Сергей	3	2018-08-01
4	Ветрова	Вероника	2	2018-08-09
5	Королева	Каролина	3	2018-08-11
6	Миронов	Мирон	3	2019-07-01
7	Алексеев	Алексей	1	2019-07-15
8	Федоров	Фёдор	2	2019-08-21

Рисунок 5 – Принцип разбиения тестовой таблицы

Теперь повторяем аналогичный запрос уже к секционированной таблице in_attr (рис.6).

```
$ explain analyze select * from in_attr where priemid
= '2019-07-08';
------------------------------------------------------------
"Append  (cost=0.00..2281.00 rows=3584 width=54)
(actual time=0.010..16.105 rows=3448 loops=1)"
"  ->  Seq Scan on in_attr  (cost=0.00..0.00 rows=1
width=54)  (actual  time=0.001..0.001  rows=0
loops=1)"
"        Filter: (priemid = '2019-07-08'::date)"
"  ->  Seq Scan on in_attr_10  (cost=0.00..2281.00
rows=3583  width=54)  (actual  time=0.008..15.792
rows=3448 loops=1)"
"        Filter: (priemid = '2019-07-08'::date)"
"        Rows Removed by Filter: 96552"
"Planning time: 0.434 ms"
"Execution time: 16.289 ms"
```

Рисунок 6 – Запрос на выборку к секционированной таблице

Из полученных результатов следует, что при выполнении запроса просматриваются только 2 таблицы: главная (в ней не содержатся данные и она служит только для связи секций) и подходящая секция in_attr_10. Append – запускает множества субопераций и возвращает все возвращенные ими строки в виде общего результата. В данном случае append запустил сканирования по двум таблицам и вернул все строки вместе.

Результаты

Эффективность оцениваем с помощью сбора временных показателей выполнения SQL-запросов к тестовым таблицам разного размера. Каждый эксперимент проводился для 100 000, 500 000, 1 000 000 и 5 000 000 записей с вычислением среднего времени выполнения по десяти попыткам. В качестве среды управления базой данных используем – pgAdmin III. Программа упрощает основные задачи администрирования, отображает объекты баз данных, позволяет выполнять запросы SQL [8]. Время выполнения запроса оценивается с помощью команды EXPLAIN. Она может в точности рассказать, что происходит, когда выполняется запрос. Прежде чем приступить к непосредственному выполнению запроса, PostgreSQL формирует план его выполнения [9]. При наличии опции ANALYSE, команда EXPLAIN планировщика покажет какие именно

операции производились запросом, реальное и планируемое время выполнения запроса и количество возвращаемых строк [10]. Подсчеты и построение гистограмм осуществлялись в программном продукте Microsoft Excel. Полученные результаты представлены в виде графиков для более удобного сравнения.

Измерим время выполнения запроса с условием отбора по конкретной дате, результат приведен на рисунке 7.

explain analyse select * from in_attr where priemid = '2019-07-08';

Количество возвращаемых записей - 17241.

Рисунок 7 – Сравнение времени выполнения операций выборки с условиями отбора по конкретной дате

Измерим время выполнения запроса с условием отбора по диапазону дат в пределах одной секции, результат приведен на рисунке 8.

explain analyse select * from in_attr where priemid >= '2019-01-01' AND priemid <= '2019-12-31';

Количество возвращаемых записей - 500000.

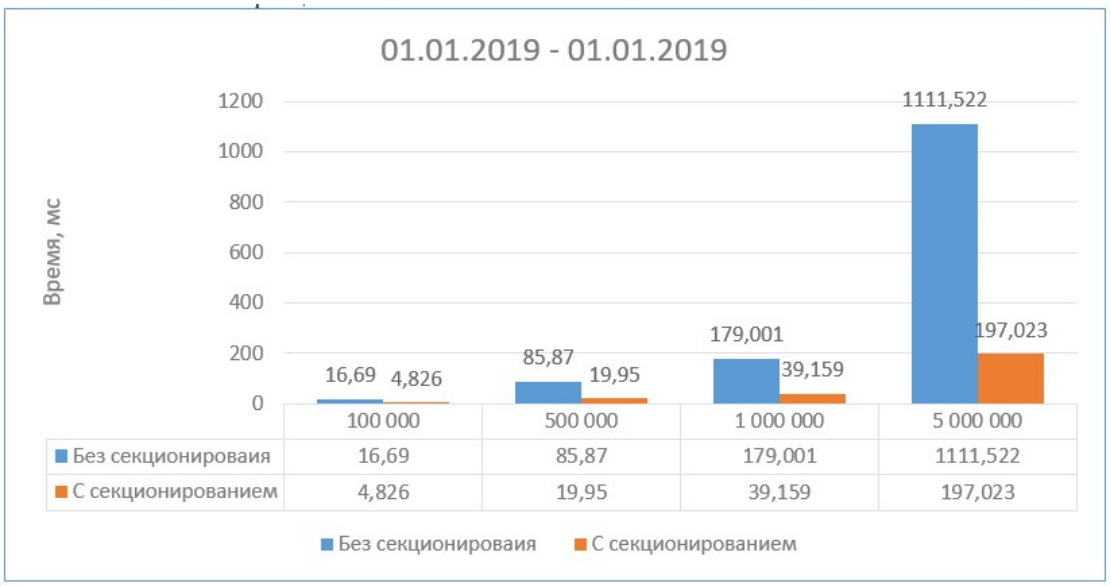


Рисунок 8 – Сравнение времени выполнения операций выборки с условиями отбора по диапазону дат в пределах одной секции

Измерим время выполнения запроса с условием отбора по диапазону дат на пересечения двух секций, результат приведен на рис. 9.

explain analyse select * from in_attr where priemid >= '2018-07-31' AND priemid <= '2019-07-30';

Количество возвращаемых записей - 517242. На рис. 10 приведено процентное изменение скорости выполнения запросов.

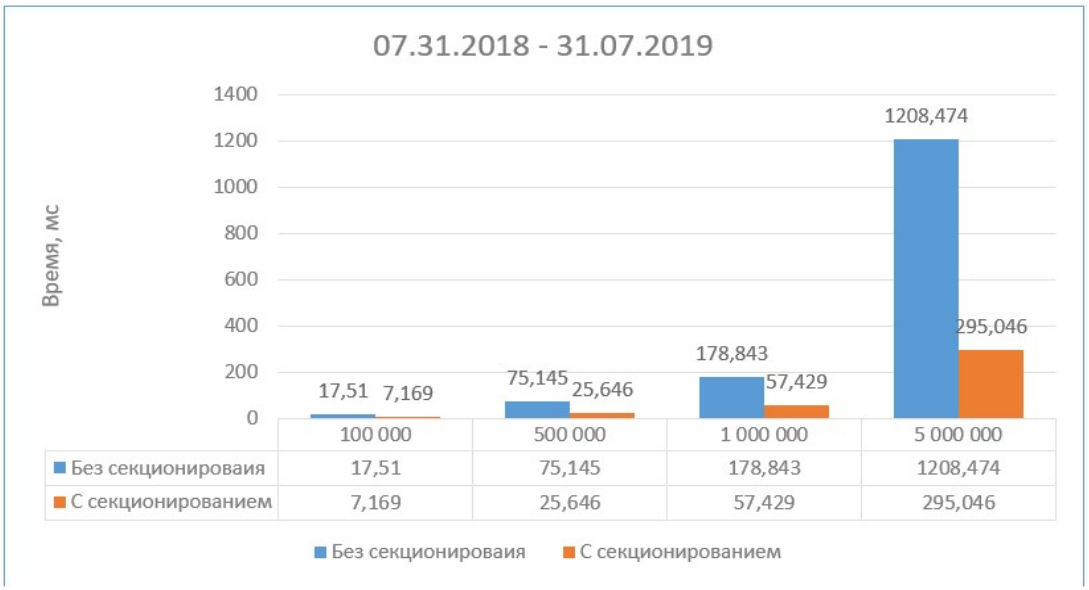


Рисунок 9 – Сравнение времени выполнения операций выборки с условиями отбора по диапазону дат на пересечении двух секции

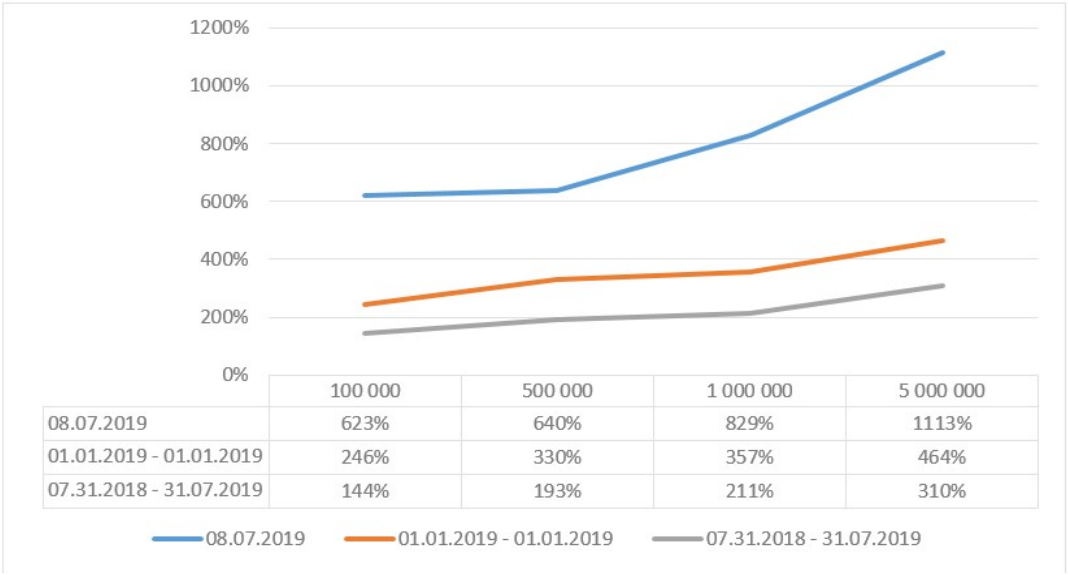


Рисунок 10 – Процентное изменение скорости выполнения

Заключение

В работе исследована технология горизонтального шардинга, с учетом особенностей организации таблиц приемной комиссии в СУБД PostgreSQL.

Для оценки эффективности технологии горизонтального шардинга по диапазонам дат, был выполнен сравнительный анализ времени планирования и выполнения запроса SELECT. Запросы выполнялись к различным по размеру таблицам, для определения временных зависимостей от количества строк.

Проведенное исследование показало существенное сокращение времени выполнения запросов к секционированным таблицам, по сравнению с обычными, что позволяет заявлять о целесообразности использования технологии секционирования в контексте повышения эффективности работы АСУ приемной комиссии.

Направление дальнейших разработок связано с размещением секций на разных серверах, а также с исследованием возможности оптимизации структуры базы данных.

Литература

1. Дейт, К.Дж. Введение в системы баз данных / К.Дж. Дейт. - К.: Диалектика; Издание 6-е, 2015. - 784 с.

2. Клеппман, М. Высоконагруженные приложения. Программирование, масштабиро-

вание, поддержка. / Мартин Клеппман. — СПб.: Питер, 2018. – 740 с.

3. Новиков, Б. А. Основы технологий баз данных: учеб. пособие / Б. А. Новиков, Е. А. Горшкова; под ред. Е. В. Рогова. — М.: ДМК Пресс, 2019. — 240 с.

4. Коротков, А. Концепции PostgreSQL. PG Day'15 Russia. – Режим доступа: http://pgday.ru/files/papers/12/pgday.2015.alexander.korotkov.pg.concepts.pdf – Заглавие с экрана.

5. Джуба, С., Волков, А. Изучаем PostgreSQL 10/ пер. с англ. А.А. Слинкин. –М:. ДМК Пресс, 2019. – 400с.

6. Пирогов, В. Ю. Информационные системы и базы данных: организация и проектирование: учеб. пособие / В. Ю. Пирогов. СПб.: БХВ-Петербург, 2009. – 528 с.

7. Васильев, А. Ю. Работа с PostgreSQL : настройка и масштабирование / А. Ю. Васильев – 4-е изд. – Электрон. дан. – Б. м.. : Б. и., 2010–2014. – Режим доступа: http://postgresql.leopard.in.ua/ – Заглавие с экрана.

8. Лузанов, П. PostgreSQL для начинающих / П. Лузанов, Е. Рогов, И. Лёвшин ; Postgres Professional. – М., 2017. – 146 с.

9. Моргунов, Е. П. PostgreSQL. Основы языка SQL: учеб. пособие / Е. П. Моргунов; под ред. Е. В. Рогова, П. В. Лузанова. — СПб.: БХВ-Петербург, 2018. — 336 с.

10. Саймон, Ригс Администрирование PostgreSQL 9. Книга рецептов / Ригс Саймон. - М.: ДМК Пресс, 2018. - 806 с.

Чередникова О.Ю., Щедрин С.В., Исаков А.Ю., Ногтев Е.А. Исследование технологии шардинга для повышения эффективности обработки данных программного комплекса приемной комиссии ДонНТУ. *В статье рассмотрено понятие шардинга, некоторые из его основных преимуществ и недостатков для решения проблемы масштабируемости АСУ приемной комиссии связанной с ростом количества обрабатываемых данных. Также исследуются основы и принципы организации секционирования в объектно-реляционной СУБД PostgreSQL, а именно, что такое секции и секционированные таблицы, какие спецификации секционирования существуют и несколько общих методов его реализации. Проведен анализ эффективности использования секционирования таблиц, в целях увеличения скорости выполнения запросов к ним, и возможности применения полученных результатов к таблицам базы данных АСУ приемной комиссии. Проведены эксперименты по установлению влияния на скорость выполнения SQL запроса выборки записей из таблицы, в зависимости от того, происходит ли выборка в пределах одной секции или нет и от количества записей в таблице.*

Ключевые слова: *SQL, СУБД PostgreSQL, масштабирование, шардинг, секционирование*

Cherednikova O.Y., Shchedrin S.V., Isakov A.Y., Nogtev E.A. Research of sharding technology to improve the efficiency of data processing of the software complex of the selection committee of DonNTU. *This article discusses the concept of sharding, some of its main advantages and disadvantages for solving the scalability problem of ACS of the selection committee, associated with an increase in the number of processed data. It also discusses the basics and principles of organizing partitioning in the PostgreSQL object-relational database management system, namely, what are partitions and partitioned tables, what partitioning specifications exist, and a general method for its implementation. The analysis of the effectiveness of the use of table partitioning, in order to increase the speed of query execution to them, and the possibility of applying the results to the tables of the ACS database of the admission committee. Experiments have been carried out to establish the effect on the speed of execution of an SQL query of selecting records from a table, depending on the number of records in it and on whether sampling occurs within one section or at the intersection of two sections.*

Key words: *SQL, PostgreSQL DBMS, scaling, sharding, partition*

Статья поступила в редакцию 19.11.2020
Рекомендована к публикации профессором Мальчевой Р. В.

УДК 004.056.52

Обзор алгоритмов защиты авторского права на программное обеспечение стеганографическими и криптографическими средствами

Н. И. Полуденный, А. В. Чернышова
Донецкий национальный технический университет
nik.poludennyy@mail.ru, chernyshova.alla@rambler.ru

Аннотация
В данной статье рассматриваются методы стеганографических и криптографических методов внедрения цифровых водяных знаков и цифровых отпечатков в программные продукты. Выделены проблемы встраивания таких меток в программное обеспечение, существующие методы внедрения меток в исходный и объектный код продукта и рассмотрены самые распространённые методы атак средств защиты для дальнейших поисков путей ликвидации уязвимостей существующих решений. Отмечена необходимость усовершенствовать алгоритмы защиты авторского права на программное обеспечение с использованием стеганографических и криптографических средств, что позволит повысить их эффективность.

Введение

Авторское право — это право интеллектуальной собственности. Каждый автор, вложивший в создание чего-либо (объекта права) время, силы и другие ресурсы, хочет защитить своё творение от всякого рода посягательств на свою интеллектуальную собственность со стороны третьих лиц (субъектов права). И если защита права владения физическим объектом может быть подкреплена юридически с помощью документов, то программное обеспечение – это, по сути, набор цифровой информации, а информация – объект абстрактный, и при извлечении фрагмента кода ПО и использовании его в другом программном обеспечении очень сложно это проконтролировать и пресечь. Поэтому сегодня активно развиваются методы защиты авторского права на программный продукт.

Под понятие программного продукта попадает исходный код, объектный код и иные материалы, полученные в процессе производства программного обеспечения. Признание исключительного права позволяет установить правообладателя произведения, признание права авторства направлено на разрешение конфликта по поводу личных неимущественных прав, если нарушитель оспаривает существование авторских прав или их принадлежность определенному лицу.

Авторское право не позволяет третьим лицам использовать продукт творческой деятельности автора без его на то разрешения. Оно позволяет автору требовать плату (или заключать иные соглашения) за право использования продукта.

Система защиты от копирования или система защиты авторских прав — комплекс средств, обеспечивающих затруднение или запрещение нелегального распространения, использования и/или изменения программных продуктов.

Под нелегальным распространением понимается продажа, обмен или бесплатное распространение программного продукта, авторские права на который принадлежат третьему лицу, без его согласия.

Нелегальное использование — использование программного продукта без согласия владельца авторских прав.

Нелегальное изменение — внесение в код программы или внешний вид (интерфейс) изменений, не оговоренных с владельцем авторских прав с тем, чтобы измененный продукт не подпадал под действие авторских прав их владельца[1].

Сегодня проблема авторского права стоит очень остро, так как программные продукты, являются, по сути, информацией, а потому не могут быть защищены от кражи такими же методами как физические объекты. Стеганографические методы позволяют внедрять информацию, доказывающую авторство, но при этом не посвящать третьих лиц о существовании таковой. На сегодняшний день цифровая стеганография активно развивается и является отличным способом для создания незаметных меток в разного рода цифровых продуктах.

Эффективной считается защита, для взлома которой требуется больше ресурсов, чем на приобретение копии продукта.

Виды технических средств защиты

Технические средства защиты можно разделить на программные, аппаратные и программно-аппаратные.

Программными являются средства защиты, реализованные программным образом. Это наиболее доступные средства.

Относительно программной защиты необходимо принять во внимание следующее.

Любая программная защита может быть раскрыта в конечное время. Известный специалист по программной защите от копирования А.Щербаков разъясняет: «Верность этого утверждения следует из того, что команды программы, выполняющей защиту, достоверно распознаются компьютером и в момент исполнения присутствуют в открытом виде как машинные команды. Следовательно, достаточно восстановить последовательность этих команд, чтобы понять работу защиты. А таких инструкций, очевидно, может быть лишь конечное число» [1].

Аппаратными называются средства защиты, использующие специальное аппаратное оборудование.

Аппаратная защита является самой надежной, но и стоимость её может превышать программную в несколько раз. В настоящее время многие фирмы и организации во всем мире работают над тем, чтобы сделать аппаратную защиту удобной и дешевой, а также рассчитанной и на массового индивидуального пользователя.

К программно-аппаратным средствам относятся средства, комбинирующие программную и аппаратную защиту.

Это наиболее оптимальные средства защиты. Они обладают преимуществами как аппаратных, так и программных средств.

К программным средствам относят разного рода средства защиты кода криптографическими и стеганографическими средствами, средствами защиты от дизассемблирования и т.д.

Проблемы защиты программных продуктов

В вопросе о защите программного продукта может встать два основных вопроса: «кому принадлежит авторство данного продукта?» и «получена ли данная копия продукта с соблюдением соглашений?» (например, была ли копия куплена или же была получена посредством «пиратства»). Первый вопрос может быть решён простым внедрением информации об авторе в продукт. Второй же вопрос может быть решён внедрением информации о копии (номер копии или данные о субъекте, который приобрёл экземпляр). Но если такую информацию разместить без попытки её сокрытия, то злоумышленники могут удалить или модифицировать эту информацию в своих целях. Во избежание таких ситуации эту информацию (метку) можно попытаться скрыть или зашифровать. С этой целью применяются различного рода методы стеганографических и криптографических внедрений меток. В случае возникновения вопроса о принадлежности авторства при наличии спора между потенциальными авторами можно будет подтвердить авторство, раскрыв метку авторства, внедрённую в спорный экземпляр продукта. Такой тип метки называется стеганографическими водяными знаками (ЦВЗ)[2].

В случае возникновения вопроса о том, получена ли копия с соблюдением соглашений, может быть извлечена метка с данными о копии, идентифицирующими копию. Такие метки называются цифровыми отпечатками (ЦО)[3].

Таким образом, для предварительной защиты программного продукта автору необходимо решить следующие вопросы:

– какую информацию будет содержать метка;

– способ внедрения метки в экземпляр продукта;

– метод сокрытия метки.

Проблемы содержимого ЦВЗ и ЦО

Содержанием ЦВЗ может быть любая информация, идентифицирующая автора. Например, ФИО, адрес, телефон, возможно, даже, какое-то кодовое слово.

Вопрос о данных, которые будут идентифицировать пользователя и приобретённую им копию продукта, несколько сложнее. Дело в том, что для идентификации пользователя, лучше всего использовать относительно неизменяемую информацию, потому что, допустим, пользователь ввёл в качестве ЦО свои ФИО и номер телефона, но через некоторое время сменил номер телефона или фамилию, тогда информация в метке не будет соответствовать действительности, и идентифицировать пользователя будет сложно, даже если он фактически санкционированно приобрёл программный продукт. Также не рекомендуется использовать для ЦО имя пользователя, имя компьютера, версию операционной системы и т.д., так как эти параметры могут часто изменяться в процессе эксплуатации ПК.

Более подходящими для таких целей служат данные об информации о дисковых устройствах, параметрах центральных процессоров, серийные номера аппаратных составляющих, MAC-адрес сетевой платы и т.д.

Однако и эти данные могут изменятся при замене аппаратных составляющих, поэтому привязка содержимого ЦО к параметрам ПК имеет свои недостатки[4].

Методы сокрытия метки в программном продукте

Скрыть метку в продукте можно, как и любую другую информацию, двумя способами. Первый – скрыть содержание информации, другими словами зашифровать её, в таком случае факт наличия информации будет известен, и хоть изменить злоумышленнику её будет проблематично, однако испортить или удалить труда не составит, для этого даже не нужно будет вскрывать алгоритм или ключ шифрования.

Второй способ – скрыть сам факт наличия метки или сделать неизвестным место, в котором метка размещена. Её содержимое в таком случае может быть не зашифровано, тогда злоумышленник может даже не догадываться о наличии метки, а потому и не попытается её удалить или изменить[5]. Однако эффективнее будет использование обоих способов – и шифрования, и стеганографии – для внедрения метки.

Однако применение методов цифровой стеганографии несёт в себе некоторые неудобства для внедрения ЦО и ЦВЗ в программные продукты: если оцифрованные изображение или звук могут быть до некоторой степени изменены без риска обнаружения сенсорными органами человека, то код программного продукта требует абсолютной точности, так как он предназначен для «чтения» компьютером, и изменение даже одного бита может привести к полной потере работоспособности программного обеспечения[6].

На рис. 1 представлено изображение-контейнер, то есть изображение, в которое не встроено никакой скрытой информации, а на рис. 2 – изображение, которое будет встроено в изображение-контейнер.

Рисунок 1 – Изображение-контейнер

Рисунок 2 – Скрываемое изображение

В результате встраивания скрываемого изображения в пустое изображение-контейнер получается изображение, представленное на рис. 3, которое практически неотличимо человеческим глазом от пустого контейнера (к тому же несовершенство аппаратных составляющих системы может даже не отобразить разницу).

Рисунок 3 – Заполненный контейнер

Обзор стеганографических способов внедрения метки в программный продукт

Способ внедрения метки зависит, во-первых, от типа и формата продукта, во-вторых, от возможности скрыть информацию в продукте определённого формата.

Одним из способов внедрения метки может быть внедрение некой информации на этапе реализации программного кода продукта, например, прописать некий незадокументированный функционал к продукту. Допустим, автор приложения сделал так, чтобы выводилась метка при вводе команды, которая не была описана в документации приложения.

Преимущество лёгкая реализация такого метода внедрения ЦВЗ и ЦО.

Недостатки: не универсально, так как для различных программных продуктов может быть

различный интерфейс; без защиты от отладки и дисассемблирования метка может быть обнаружена и удалена/модифицирована.

На сегодняшний день одним из самых распространённых расширений исполнимых файлов является EXE. Их структура не привязана к языку программирования, на котором был написан программный продукт. Поэтому внедрение меток в объектные коды имеет большую универсальность и независимость от программного интерфейса.

Один из таких способов основан на наличии свободных участков в объектных кодах программ, хранящихся в исполнимых файлах, там могут содержаться полностью или частично свободные секторы файла. Данный метод опирается на то, что секции в PortableExecutable(PE) файле после процедуры выравнивания заполняют недостающие байты нулевым значением, чтобы длина секции была кратна определённому значению (см. рис. 4).

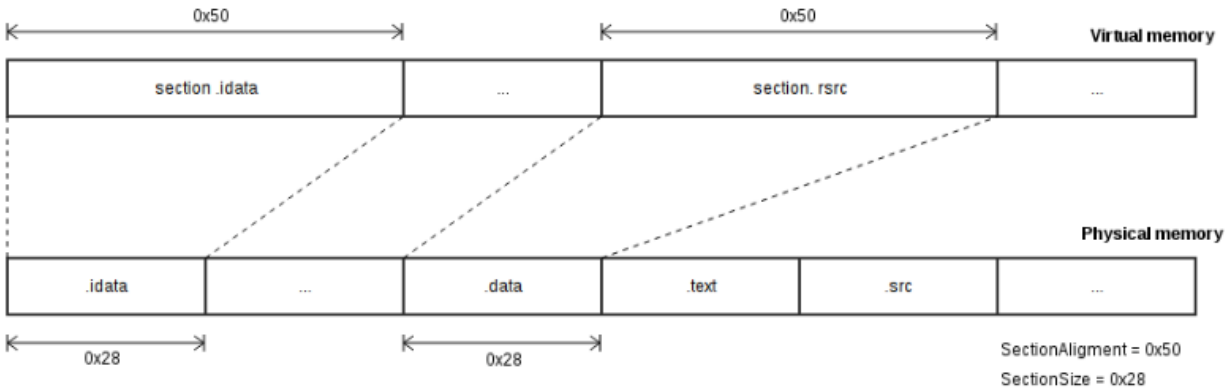


Рисунок 4 – Принцип выравнивания

На рисунке 4 видно, что размер секций в физической памяти = 0x28 байт, тогда как в виртуальной памяти, размер должен быть кратен 0x50. Для этого байты с адресами 0x28-0x50 заполнятся нулевыми байтами 0x00, в таких участках и могут быть спрятаны метки [7].

Внедрение авторской информации в свободные участки, во-первых, гарантирует правильную работу объектного кода, который был изменён, потому что не внедряется в значимую часть кода, а во-вторых, не изменяет размер файла после внедрения метки.

Преимущества:

– размер объектного файла не изменится после внедрения, что снижает вероятность обнаружения метки злоумышленником;

– внедрение метки не влияет на работоспособность программного продукта;

– метод прост в реализации.

Недостатки:

– при обнаружении метки злоумышленником она может быть удалена/модифицирована без потери работоспособности программного продукта.

Вторым способом внедрения в объектный код является изменение участков объектных кодов, которые не влияют на работоспособность программного продукта. Например, структура файлов исполнимых PE-файлов такова, что изменение значений некоторых полей не скажется на функциональности программного продукта (см. рис. 5 и рис. 6) [8].

```
Offset(h)  00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F  Текст декодирован
00000000   4D 5A 90 00 03 00 00 00 04 00 00 00 FF FF 00 00  MZђ.........яя..
00000010   B8 00 00 00 00 00 00 00 40 00 00 00 00 00 00 00  ё.......@.......
00000020   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00000030   00 00 00 00 00 00 00 00 00 00 00 00 80 00 00 00  ............Ъ...
00000040   0E 1F BA 0E 00 B4 09 CD 21 B8 01 4C CD 21 54 68  ..є..ґ.H!ё.LH!Th
00000050   69 73 20 70 72 6F 67 72 61 6D 20 63 61 6E 6E 6F  is program canno
00000060   74 20 62 65 20 72 75 6E 20 69 6E 20 44 4F 53 20  t be run in DOS
00000070   6D 6F 64 65 2E 0D 0D 0A 24 00 00 00 00 00 00 00  mode....$.......
00000080   50 45 00 00 4C 01 03 00 7C EA 2C 5B 00 00 00 00  PE..L...|к,[....
00000090   00 00 00 00 E0 00 22 00 0B 01 30 00 00 2E 00 00  ....а."...0.....
```

Рисунок 5 – Фрагмент объектного кода без внедрённой метки

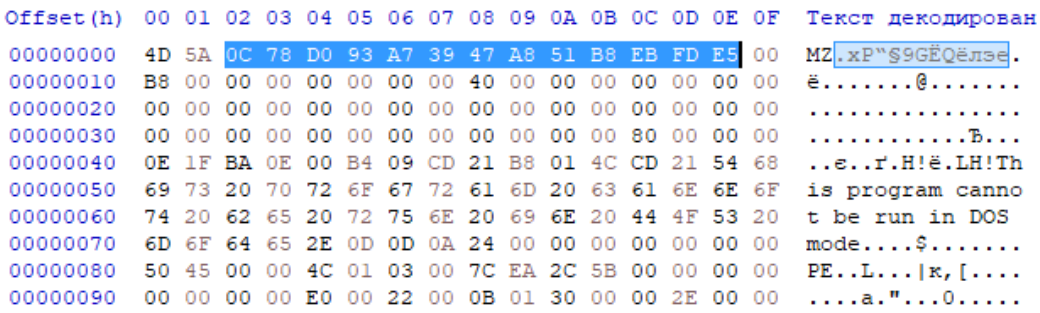


Рисунок 6 – Фрагмент объектного кода с внедрённой меткой

К примеру, DOS-заголовок имеет размер 64 байта, из которых 6 байт являются критически важными: 2 байта отведены для сигнатуры «MZ» в начале каждого PE-файла, без которого файл просто не запустится, и 4 байта для хранения смещения до PE-заголовка. Остаётся 58 байт для произвольного сообщения [7].

Преимущества:

– внедрение метки не влияет на работоспособность программного продукта;

– не влияет на размер исполнимого файла;

– метод прост в реализации.

Недостатки:

– при обнаружении злоумышленником, метка может быть удалена/модифицирована без нарушения работоспособности программного продукта.

Третий способ внедрения метки в объектный код основывается на том, чтобы в случае её изменения или удаления, нарушалась работоспособность программного продукта.

Такой метод может быть реализован на этапе кодирования продукта или реализован как отдельный модуль, который, если будет обнаружено нарушение целостности метки в случае ЦВЗ, и если обнаружится несовпадение данных и контрольных данных в случае ЦО, предпримет соответствующие меры [7].

Преимущества:

– удаление/модификация метки напрямую в объектном коде нарушает работоспособность программного продукта;

Недостатки:

– без защиты от дизассемблирования может быть модифицирована;

– сложнее в реализации в сравнении с предыдущими способами, так как внедрение происходит в значимую часть кода.

Обзор криптографических способов внедрения метки в программный продукт

В дополнение к стеганографическим методам сокрытия, метка, как и любая другая информация, может быть зашифрована некоторыми криптографическими алгоритмами. И, естественно, использование и криптографического, и стеганографического подходов внедрения метки увеличат её стойкость к модификации.

Многие специалисты в качестве средства защиты прав автора на компьютерные программы называют электронную цифровую подпись. Электронная цифровая подпись используется при передаче информации в компьютерных сетях для подтверждения автора (создателя) передаваемой информации и, кроме того, служит для доказательства (проверки) того факта, что подписанное сообщение или данные не были модифицированы.

Электронная цифровая подпись строится на основе двух компонент: во-первых, содержания информации, которая подписывается, во-вторых, личной информации (код, пароль, ключ) того, кто подписывает. Очевидно, что изменение каждой компоненты приводит к изменению электронной цифровой подписи. Алгоритмы создания электронной цифровой подписи основаны на шифровании с открытым ключом [9]. При использовании электронной цифровой подписи для защиты авторских прав на компьютерные программы необходимо учитывать, что электронная цифровая подпись только тогда может выступать в качестве информации об авторском праве на компьютерную программу, если она приложена к каждому экземпляру (копии) программы [8].

На основе шифрования на практике используются следующие механизмы защиты программ [8]:

– шифрование кода программы;

– шифрование фрагмента программы.

В первом случае шифруется код программы, и расшифровывается только во время выполнения. Во втором случае, обычно, выбирается критический участок программы.

В обоих случаях может применяться как статическое, так и динамическое шифрование.

При статическом шифровании весь код (фрагмент кода) один раз шифруется/дешифруется. В зашифрованном виде код постоянно хранится на внешнем носителе.

В расшифрованном виде присутствует в оперативной памяти.

При динамическом шифровании последовательно шифруются/дешифруются процедуры (или отдельные фрагменты) программы или критического участка программы [7]. Такие же методы могут применятся и для шифрования ЦО и ЦВЗ.

Методы вскрытия защиты

Выполним анализ наиболее распространенных методов вскрытия защиты[10].

Прямое построение по загрузочному модулю текста программы на ассемблере или другом языке. После этого из текста удаляются подпрограммы проверки, и строится заново загрузочный модуль.

Недостаток – если исполняемые коды зашифрованы и расшифровываются лишь в момент исполнения, по загрузочному модулю практически невозможно восстановить его первоначальный вид на ассемблере.

В загрузочном модуле отладчиком прослеживается система защиты, а затем в модуль вносятся такие изменения, чтобы соответствующая подпрограмма уже не смогла активизироваться.

Недостаток – существуют приемы противодействия, прерывающие процесс дизассемблирования. На ПЭВМ большинство отладчиков используют для отладки стандартные прерывания. Если изменить векторы этих прерываний, исследование программы отладчиком становится затруднительным. Второй путь – передача управления в программе с помощью изменения указателя стека так, что отладчик не будет знать адрес следующей исполняемой программы.

С помощью отладчика программа загружается в ОЗУ, проходятся все ступени защиты и в момент завершения работы программы защиты на диск записывается копия ОЗУ, чтобы в дальнейшем можно было загрузить с диска программу с пройденным этапом защиты.

Недостаток – этот метод имеет смысл применять для программ небольшого размера и только в том случае, если защита проверяется сразу, а не в середине работы и нет проверок в течение всего сеанса. Как правило, необходима полная информация о работе защиты (в частности долговременное наблюдение).

Аппаратная трассировка с помощью внутрисхемных эмуляторов или логических анализаторов с блоками для трассировки.

Недостаток – нужна специальная дорогостоящая аппаратура и специалисты, умеющие с ней работать. Противодействие аналогично борьбе против отладчиков и дизассемблирования.

Выводы

В данной работе рассмотрены существующие методы защиты авторского права на программное обеспечение с использованием стеганографических и криптографических средств. Выявлены достоинства и недостатки существующих подходов. Рассмотрены методы вскрытия защиты.

В дальнейшем планируется разработать усовершенствованный алгоритм защиты авторского права на программное обеспечение с использованием стеганографических и криптографических средств, повысив эффективность существующих средств защиты авторского права на программное обеспечение.

Литература

1. Щербаков, А. Защита от копирования / А. Щербаков. – М.: ЭДЭЛЬ, 1992. – 79 с.

2. Мельников, Ю. Цифровые водяные знаки — новые методы защиты информации / Ю. Мельников, А. Теренин, В. Погуляев. - PC Week N48, 2007.

3. Орешин, Е. Эффективные способы защиты авторских и смежных прав в Интернете / Е. Орешин // Журнал Суда по интеллектуальным правам, 2015. - N9. - С. 48-55.

4. Андерсон, Р. Security Engineering / Р. Андерсон. – N.Y.JohnWiley&Sons, 2001. – 612 с.

5. Грибунин, В. Цифровая стеганография / В. Грибунин, И. Оков, И. Туринцев.– М.: СОЛОН-ПРЕСС, 2009. – 277 с.

6. Коханович, Г. Компьютерная стеганография: теория и практика / Г. Коханович, А. Пузыренко. – Киев: МКПресс, 2006. – 286 с.

7. Peering Inside the PE: A Tour of the Win32 Portable Executable File Format. – Режим доступа: https://docs.microsoft.com/ru-ru/previous-versions/ms809762(v=msdn.10)?redirectedfrom=MSDN

8. Монахов, М. Защита авторских прав на программное обеспечение / М. Монахов, В. Ташмухамедова. - Владимирский государственный университет, 2009. – 58 с.

9. Чмора, А. Современная прикладная криптография. 2-е изд., стер. / А. Чмора. – М.: Гелиос АРВ, 2004. – 256 с.

10. Середа, С. Оценка эффективности систем защиты программного обеспечения / С. Середа. — Режим доступа: http://www.security.ase.md/publ/ru/pubru30.html

Полуденный Н.И., Чернышова А.В. Обзор алгоритмов защиты авторского права на программное обеспечение стеганографическими и криптографическими средствами. *В данной статье рассматриваются методы стеганографических и криптографических методов внедрения цифровых водяных знаков и цифровых отпечатков в программные продукты. Выделены проблемы встраивания таких меток в программное обеспечение, существующие методы внедрения меток в исходный и объектный код продукта и рассмотрены самые распространённые методы атак средств защиты для дальнейших поисков путей ликвидации уязвимостей существующих решений. Отмечена необходимость усовершенствовать алгоритмы защиты авторского права на программное обеспечение с использованием стеганографических и криптографических средств, что позволит повысить их эффективность.*

Ключевые слова: *стеганография, криптография, защита, шифрование, авторское право, программное обеспечение*

Poludenny N. I., Chernyshova A.V. Review of steganographic and cryptographic software protection tools. *This article discusses the methods of steganographic and cryptographic methods for introducing digital watermarks and digital fingerprints into software products. The problems of embedding such tags in software, existing methods of embedding tags in the source and object code of the product are highlighted, and the most common methods of security attacks for further searching for ways to eliminate vulnerabilities of existing solutions are considered. It is noted that it is necessary to improve the algorithms of copyright protection for software using steganographic and cryptographic means, which will increase their efficiency.*

Keywords: *steganography, cryptography, protection, encryption, copyright, software.*

Статья поступила в редакцию 19.07.2020
Рекомендована к публикации профессором Феляевым О. И.

УДК 004.7

Метод прогнозирования оценки пропускной способности

В.В. Климов

Донецкий институт железнодорожного транспорта

boban_cc@mail.ru

Аннотацияи

Необходимость прогнозирования определенных оценок эффективности работы сети, таких как пропускная способность, обусловлена сложностью сбора и обработки соответствующих данных из-за их динамически изменяющегося характера. Целью данной статьи является разработка метода оценки пропускной способности, учитывающего фрактальные свойства поступающего трафика при обеспечении заданных параметров качества обслуживания. В качестве базовой исследуется классическая модель экспоненциального сглаживания, основными временными параметрами которой являются интервалы статистики и регулирования. Решение об их величинах принимались с учетом оценки вероятности потерь и эффективности прогнозирования. Под последним понимается процентное соотношение между спрогнозированными оценками трафика и пропускной способности. Данное условие представлено в аналитическом описании классической модели. Сделан результат о низком уровне оценок эффективности прогнозирования пропускной способности для классической модели и ее модификации. Предложено использовать в качестве модели предсказания трафика ARFIMA-модель. Это позволило повысить эффективность предсказания. Для улучшения это тенденции предложена модель, позволяющая «отслеживать» резкие фрактальные выбросы трафика. Это дает возможность регулировать пропускную способность только по «необходимости». Разработанный метод оценки пропускной способности, учитывающий фрактальные свойства поступающего трафика при обеспечении заданных параметров качества обслуживания.

Введение

Качество работы любой сети передачи данных оценивается исходя из таких факторов как: качество обслуживания абонентов, эффективность использования сетевых ресурсов, эффективность управления сетью с целью перераспределения имеющихся сетевых ресурсов. Для того, чтобы это оценить, необходим анализ текущего состояния транспортной сети оператора мобильной связи, включая оборудование и линии связи. Так как это довольно нетривиальная задача сбора и обработки информации, то целесообразнее прогнозировать определенные характеристики. Одной из таких является оценка необходимой пропускной способности.

Постановка задачи и цели исследования

Особенностью прогнозирования оценки необходимой пропускной способности является ее зависимость от поступающего трафика. При этом необходимо выполнять требования по обеспечению параметров качества обслуживания. Существующие методы прогнозирования учитывают или тип поступающего трафика или возможность обеспечения требуемого качество обслуживания. Следовательно, целью статьи является разработка метода оценки пропускной способности, учитывающего фрактальные свойства поступающего трафика при обеспечении заданных параметров качества обслуживания.

Изложение основного материала

Так как трафик современных телекоммуникационных сетей – самоподобный [1], на данный момент не существует единой универсальной математической модели трафика ТКС. В работе [2] приведен обзор моделей фрактальных точечных процессов таких как: фрактальный ON/OFF источник, фрактальный дробовый точечный процесс, фрактальный биноминальный процесс. Данные модели позволяют создавать реализации самоподобных потоков в процессе имитационного моделирования.

В работах [3, 4] на основе анализа большого числа работ по исследованию трафика в IP сетях приведена классификация трафика, и каждому виду трафика сопоставлен закон распределения. В этой же работе утверждается, что для описания процессов в современных ТКС наиболее широко применяются три вида распределений: Парето, Вейбулла и логнормальное. Рассмотрим прогностический

поход к определению характеристик поступающего трафика.

Итак, согласно работам [5, 6], в общем виде прогностическая модель трафика выглядит следующим образом:

$$Y(t) = T(t) + S(t) + C(t) + \varepsilon(t), \quad (1)$$

где $T(t)$ – тренд, который представляет собой плавно изменяющуюся составляющую; $S(t)$ – сезонная составляющая; $C(t)$ – циклическая составляющая; $\varepsilon(t)$ – случайная составляющая.

Тренд, сезонная и циклическая составляющие могут быть определены методами статистического анализа. Определенную сложность вызывает случайная составляющая, т.к. именно она формирует динамический характер трафика современных сетей.

Производительность предложенных моделей будет оцениваться использованием реальных сетевых трассировок. Исходная трассировка содержит время поступления каждого пакета. Рассчитывается скорость трафика для каждого временного интервала и прогнозируется скорость передачи для следующего интервала. Трассировка делится на две части. Первые 25 % трассы составляют тренировочный набор, а оставшиеся используются для проверки точности прогноза. На этапе обучения определяются оптимальные параметры модели, чтобы ошибка между прогнозируемыми и фактическими значениями сводилась к минимуму. На этапе вычисления, выходные данные прогнозируются с использованием входных (тестовых) данных, которые не использовались для обучения, и вычисляется ошибка между прогнозируемыми и фактическими значениями.

В качестве исходной, взята модель, реализующая простой метод экспоненциального сглаживания, реализованный в аналитической модели прогнозирования.

Итак, в классической прогностической модели, определены следующие основные параметры:

- оценка величины пропускной способности $(b_j(\Delta\tau))$;
- частота сбора статистических данных;
- размер интервала статистики (Δt);
- исходная трассировка;
- время, необходимое для проверки точности прогноза $(\Delta\tau)$;
- размер интервала перераспределения.

Алгоритм процедуры решает следующую задачу: для некоторого потока трафика $tr_j(i)$, где j – сервис, а i – последовательность временных отсчетов, в которые производится оценка трафика, необходимо оценить пропускную способность $(b_j(\Delta\tau))$, с учетом параметров качества обслуживания. Последовательность $\lambda_j(i)$ формируется путем мониторинга поступающего трафика и обрабатывается следующим образом:

$$c_j(i\Delta t) = \frac{1}{n}\sum_{l=1}^{n}\lambda_j((i(n-1)+l)\Delta t), \quad (2)$$

где $c_j(n)$ – оценка интенсивности j-го потока трафика i-ой исходной трассировки;

n – число измерений трафика за статистический интервал Δt.

Значение оценки пропускной способности прогнозируется, исходя из максимальных значений оценок интенсивностей в пределах интервала перераспределения:

$$b_j(m\Delta\tau) = \max_{i=1..k}(c_j(m(k-1)+i)\Delta t), \quad (3)$$

где k – число статистических интервалов в одном интервале регулирования;

m – номер интервала перераспределения.

Полученное значение является оценкой пропускной способности для j-го потока трафика на следующий интервал перераспределения, при выполнении следующих условий:

$$b_j(m\Delta\tau),\ b_j(m\Delta\tau) > b_j((m-1)\Delta\tau), \quad (4)$$

$$b_j((m-1)\Delta\tau),\ b_j(m\Delta\tau) \le b_j((m-1)\Delta\tau), \quad (5)$$

$$P \le P_{QoS},\ utl \to 1, \quad (6)$$

где P – доля потерь трафика исследуемой математической модели;

P_{QoS} – требования к качеству обслуживания для поступающего трафика;

utl – показатель оценки качества прогнозирования пропускной способности.

При этом выполняется условие отсутствия перегрузок в физическом канале:

$$\sum_{j=1}^{N} b_j(m\Delta\tau) < C, \quad (7)$$

где C – пропускная способность физического канала;

N – количество виртуальных каналов с заданной пропускной способностью, в одном физическом.

Значения $b_j(m\Delta\tau)$ рассчитываются для всех каналов с заданной пропускной способностью, принадлежащих одному логическому каналу. Для упрощения расчетов, согласно разным моделям, каждый физический

канал содержит один логический. Выражение (5) учитывает требования к качеству обслуживания, которые можно оценить согласно:

$$Pl = \frac{1}{M}\sum_{l=1}^{M} P_j((n-l)\Delta\tau), \qquad (8)$$

где M − количество интервалов перераспределения в реализации поступающего трафика;

$P_j(\Delta\tau)$ − доля потер трафика за интервал перераспределения.

Коэффициент использования канала, если его оценка будет принята, как предложенная оценка пропускной способности:

$$utl = \frac{1}{M}\sum_{l=1}^{M} utl_j((n-l)\Delta\tau), \qquad (9)$$

где $utl_j(\Delta\tau)$ − коэффициент использования рассчитанной оценки пропускной способности в рамках интервала перераспределения.

В зависимости от того, будет ли интенсивность поступающего трафика больше или меньше, чем рассчитанная оценка, вычисляется доля потерь трафика и коэффициент использования рассчитанной оценки пропускной способности в рамках интервала перераспределения. Для $\lambda_j(\Delta\tau) > tun_j(\Delta\tau)$,

$$P_j(\Delta\tau) = 0, \;\; utl_j(\Delta\tau) = \left[\frac{\int_0^{\Delta\tau}\lambda_j(\Delta\tau)d\Delta\tau}{\int_0^{\Delta\tau} b_j(\Delta\tau)d\Delta\tau}\right]; \qquad (10)$$

для $\lambda_j(\Delta\tau) < b_j(\Delta\tau)$,

$$utl_j(\Delta\tau) = 1,$$

$$P_j(\Delta\tau) = \left[\frac{\int_0^{\Delta\tau}\lambda_j(\Delta t)d\Delta t - \int_0^{\Delta\tau} b_j(\Delta t)d\Delta t}{\int_0^{\Delta\tau}\lambda_j(\Delta t)d\Delta t}\right]. \qquad (11)$$

Для величин интервалов статистики и перераспределения, равным 5 минутам и нескольким часам, соответственно, данные значения необходимо уточнять. Это связано с высокой скоростью изменения трафика. Для этого проведены исследования зависимостей показателей качества обслуживания поступающего трафика от величин интервалов статистики и перерегулирования. Приемлемое значение величин интервалов статистики и перераспределения можно определить, исходя из условий обеспечения требуемых параметров качества обслуживания (8) и (9).

В качестве исследуемого трафика использовалась последовательность интенсивности http-запросов к серверу NASA [7] в течение месяца с периодичностью в 1 с. Для выбранной последовательности проведен анализ на самоподобность. Выявлено, что она обладает свойствами медленно убывающей зависимости с коэффициентом Херста равным 0,78. На основе анализа высказано предположение о том, что последовательность обладает «краткосрочной памятью». Это дает возможность для ее прогнозирования.

Для пакетных сетей передачи данных, для голосового трафика величина вероятности потерь не должна превышать 0,25%. Согласно полученным зависимостям, величина интервала статистики, удовлетворяющая необходимым значениям вероятности потерь и максимальной эффективности прогнозирования составляет 6 отсчетов, при фиксированной величине интервала регулирования (20 отсчетов). А величина интервала перераспределения – 10, при фиксированном размере интервала статистики в 2 отсчета. Т.о., чем чаще снимается статистика, тем выше качество работы алгоритма, при этом увеличение размера интервала перераспределения может привести к увеличению размера ошибки прогнозирования. Можно сделать вывод о предпочтительности использования значений меньших величин интервалов статистики регулирования.

К основным недостаткам классической модели относятся несовершенство оценки прогнозируемой величины пропускной способности и медленная скорость адаптации к интенсивности поступающего трафика. Каждый из вышеперечисленных недостатков вносит свою долю погрешности в оценку пропускной способности. Все это приводит к основному недостатку классической модели – высокому объему используемого канального ресурса. Ослабление либо устранение приведенных недостатков возможно путем модифицирования исходной модели.

Ошибка прогнозирования, вносимая в расчеты оценки пропускной способности, может быть уменьшена за счет использования более точных прогностических моделей. В представленной выше классической математической модели, в качестве прогностической, используется упрощенный алгоритм экспоненциального сглаживания. Применение этого алгоритма для процессов с высокой скоростью развития приводит к большим погрешностям прогнозирования.

Для уменьшения погрешностей прогнозирования предложены следующие модификации.

Математическая модель с увеличением интервала статистики. Формирование статистических данных происходит в моменты времени, равные этим отсчетам:

$$c_j(i\Delta t) = \lambda_j(n \cdot \Delta t), \quad n = \Delta t, 2 \cdot \Delta t, \ldots, \qquad (12)$$

т.е. оценка значения в интервале статистики фиксируется каждые Δt отсчетов.

Расчет оценки пропускной способности в рамках интервала перераспределения производится согласно формуле (2).

Математическая модель с увеличением интервала перераспределения. Оценка пропускной способности вычисляется согласно следующему выражению 13, оценка значений трафика в рамках интервала статистики производится согласно (2):

$$b_j(m\Delta\tau) = \frac{1}{k}\sum_{i=1}^{k} c_j((m(k-1)+i)\Delta t). \qquad (13)$$

Математическая модель максимальных значений осуществляет выбор максимальных значений из отсчетов в рамках интервала статистики и усреднение значений по интервалу перераспределения. Математическое выражение для оценки интенсивностей поступающего трафика в интервале статистики:

$$c_j(i\Delta t) = \max_{l=1..n}(\lambda_j((i(n-1)+l)\Delta t)). \qquad (14)$$

Расчет оценки пропускной способности при этом будет соответствовать (3).

Для всех модифицированных алгоритмов условия выделения необходимой пропускной способности для дальнейшей передачи трафика соответствуют выражениям (4) – (7). Затем, согласно (8) − (9), произведен расчет оценок полученных значений пропускной способности по предложенным моделям, рисунок 1. На рис. 1 приведены графические зависимости полученных оценок расчета: Модель 1 – математическая модель с увеличением интервала статистики; Модель 2 – математическая модель с увеличением интервала регулирования; Модель 3 – математическая модель максимальных значений.

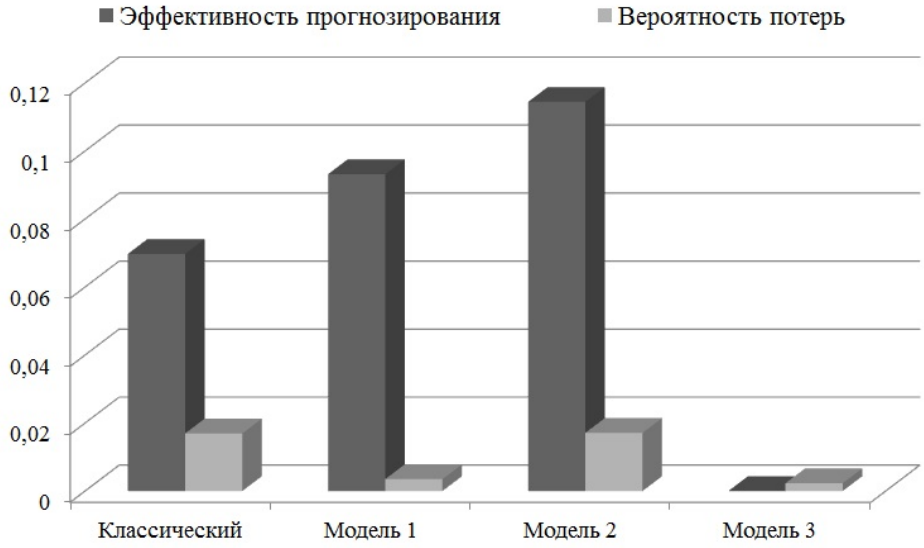


Рисунок 1 – Сравнение оценок полученных с помощью модифицированных моделей

Исходя из величин требуемых параметров качества обслуживания для поступающей последовательности, при формировании оценки пропускной способности, можно сделать вывод о том, что данные модификации также вносят ошибку прогнозирования. Это связано с «медленной» скоростью адаптации представленных моделей оценки пропускной способности к быстроизменяющемуся характеру поступающего трафика. Следовательно, чтобы математическая модель имела такую же скорость адаптации, как и поступающий трафик необходимо, чтобы первая имела такие же параметры, что и последняя.Это возможно, если использовать в качестве прогностической, модель, точно описывающую поступающий трафик – ARFIMA-модель.

В этом случае, соответствующая оценка пропускной способности определяется исходя из:

$$b_j(m\Delta\tau) = \max_{i=1..k}(ARFIMA(c_j((m(k-1)+i)\Delta t))), \quad (15)$$

где $ARFIMA(c_j((m(k-1)+i)\Delta t))$ − оценка отсчетов в рамках интервалов статистики, спрогнозированная ARFIMA-моделью.

Условия работы математической модели соответствуют (4) − (7). Сравнение прогнозных значений требуемых оценок качества согласно классической и ARFIMA-модели, представлено на рис. 2.

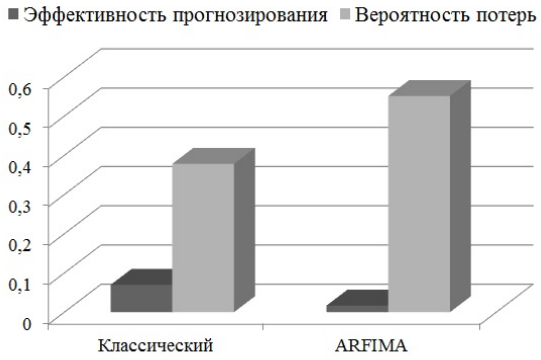


Рисунок 2 – Сравнение оценок классической и ARFIMA-модели

Результаты моделирования показали, что повышение точности прогнозной модели приводит к ухудшению параметров качества обслуживания. Это говорит о низкой точности оценок, используемых для расчета. Для уменьшения влияния этого недостатка на результат работы математической модели рассмотрим следующие модификации оценок.

В представленных математических моделях принятие решения о необходимой оценке пропускной способности сводится к интуитивно простому решению: нахождению максимального или среднего значения. Это обусловлено технологическими возможностями операторов: на практике обычно отслеживаются пиковые значения скоростей поступающего трафика данных за определенный цикл наблюдения.

Математические модели, оперирующие максимальными и средними значениями в этом случае, позволяют «обеспечить» заданные параметры качества обслуживания (величину доли потерь). Несмотря на это, увеличивается объем используемого канального ресурса при удовлетворительном значении коэффициента утилизации.

Данное явление объяснимо свойствами поступающего трафика (пиковые значения обычно больше постоянного уровня трафика в несколько десятков раз). Т.о., значение соответствующей оценки пропускной способности в десятки раз больше необходимого, с удовлетворительными значениями параметров качества обслуживания и значениями коэффициента утилизации.

Для уменьшения значения этой оценки необходимо, чтобы значение расчетной оценки пропускной способности канала была как можно ближе к значению интенсивности поступающего трафика, с выполнением условий (7) – (9). Т.е., чтобы выполнялось следующее выражение:

$$b_j(m\Delta\tau) = \int_{(m-1)\Delta\tau}^{m\Delta\tau} c_j^*(m)\Delta\tau \Big/ \Delta\tau, \quad (16)$$

где $c_j^*(m\Delta\tau)$ − прогнозируемые значения интенсивности трафика j-го сервиса для m-го интервала перераспределения, рассчитанного по следующему выражению:

$$c_j^*(m\Delta\tau) = ARFIMA(c_j(m(k-1)+i)\Delta\tau))). \quad (17)$$

Условия расчета оценок пропускной способности соответствуют (6) – (7). При анализе сравнительных зависимостей эффективности работы математических моделей для оценок пропускной способности (математическая модель Autobandwidth и интегральной математической модели) видно, что интегральная математическая модель позволяет добиться значения коэффициента утилизации большего, чем у процедуры Autobandwidth. Однако, значение величины потерь при этом достигает недопустимого значения. И наоборот, уровень потерь с процедурой Autobandwidth соответствует требованиям по качеству обслуживания, но значение утилизации канала достаточно низкое. Таким образом, необходимо использовать комбинации этих двух алгоритмов, в зависимости от скорости изменения поступающего трафика. Для этого, нужно адаптировать начало и конец интервала регулирования таком образом, чтобы они совпадали с началом и концом зон резкого изменения трафика. Следовательно, необходимо рассмотреть задачу о неравномерных интервалах регулирования.

Пусть для некоторой реализации $tr_j(i)$, где j – сервис, а i – последовательность временных отсчетов, в которые производилась оценка трафика, необходимо зарезервировать пропускную способность для ТЕ-туннеля $(tun_j(\Delta\tau))$. При этом реализации трафика прогнозируются, согласно выражению:

$$tr_j^*((in+l)\Delta t) = ARFIMA(tr_j((i(n-1)+l)\Delta t)), \quad n = 1,2,\ldots t \quad (18)$$

Оценка пропускной способности резервируемого ТЕ-туннеля рассчитывается, согласно формуле (17).

Начало и конец интервала регулирования находится исходя из значений $Q(n\Delta t)$ – функция приращения скорости изменения трафика, которая рассчитывается по реализации $tr_j^*((in+l)\Delta t)$:

$$Q(n\Delta t)=\frac{tr_j^*((i(n-1)+l)\Delta t)-tr_j^*((i(n-2)+l)\Delta t)}{tr_j^*((i(n-1)+l)\Delta t)},$$
$$n=1,2,\ldots t_s \qquad (19)$$

Если $Q(n\Delta t)>1$, то этот отсчет трафика фиксируется как начало интервала регулирования, если значение $Q(n\Delta t)$ меняет знак с отрицательного на положительный, то такой отсчет трафика фиксируется как конец интервала регулирования.

Аналогично происходит расчет, если прогнозируемый трафик резко убывает, только при значениях $Q(n\Delta t)<-1$ для начала интервала регулирования и изменение знака с положительного на отрицательный соответственно. В том случае, если в течении интервала прогнозирования трафик существенно не изменялся, т.е. $-1<Q(n\Delta t)<1$, то началом и концом интервала регулирования считается начало и конец интервала наблюдения.

Итак, математическая модель определения начала и конца интервала регулирования будет выглядеть следующим образом:

$$m\Delta\tau=\begin{cases} n\Delta t, & Q(n\Delta t)>1 \\ n\Delta t, & Q(n\Delta t)<-1, \\ & n\Delta t_p \end{cases} \qquad (20)$$

$$(m+1)\Delta\tau=\begin{cases} n\Delta t, & Q(n\Delta t)=0 \\ n\Delta t, & Q(n\Delta t)=0, \\ & (n+1)\Delta t_p \end{cases} \qquad (21)$$

где Δt_p – интервал прогноза.

Общим ограничением для математической модели остаются выражения (6).

Данная математическая модель позволяет «реагировать» на трафик с высокой пачечностью, так как понятие «интервала регулирования» появляется только в случае начала резкого изменения скорости поступающего трафика. Результаты работы математической модели по оценке величины пропускной способности ТЕ-туннеля представлены на рис. 3.

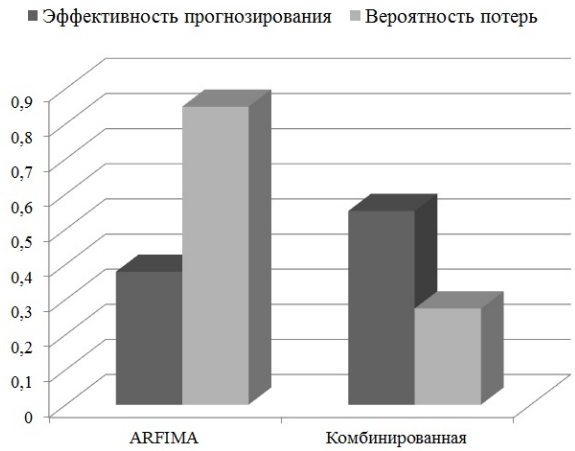


Рисунок 3 – Сравнение оценок ARFIMA-модели и комбинированной

Использование математической модели с адаптивным интервалом регулирования позволяет повысить коэффициент использования канальных ресурсов внутри ТЕ-туннеля на 25%, однако при этом параметр качества ухудшился на 10 %.

Это указывает, что скорость адаптации представленных моделей также недостаточна для обеспечения заданных параметров качества обслуживания при расчете оценки пропускной способности ТЕ-туннеля.

Выводы

На основе свойства самоподобности мультисервисного трафика выдвинуто предположение о зависимости следующих параметров модели: интервал статистики и интервал регулирования, от необходимого уровня параметров качества обслуживания. На основе анализа этой зависимости выработаны рекомендации по использованию интервалов следующих размеров: интервала статистики в 1 или 2 отсчета, интервала регулирования − 9 или 10 отсчетов.

Прогнозирование, согласно классической модели показало низкий уровень эффективности прогнозирования, что объясняется несовершенством используемых математических моделей для определения как оценок пропускной способности. Для устранения этих недостатков предложены модификации классической модели. Для этих моделей величина оценки эффективности прогнозирования пропускной способности до 10%. при величине уровня потерь до 3%. Для предложенных моделей прогнозирования: ARFIMA-модели, с интегральной оценки эффективности прогнозирования и с адаптивным интервалом регулирования, можно сделать следующие выводы.

Оценка эффективности прогнозирования, с использованием ARFIMA-

модели, достигает значения в 53%. Модифицирование механизма расчета оценки эффективности прогнозирования позволяет улучшить его на величину до 10%. Так, при расчете оценки пропускной способности алгоритмом адаптивного интервала регулирования, происходит «отслеживание» резких фрактальных выбросов трафика. Это дает возможность регулировать пропускную способность только по «необходимости», что улучшает оценку эффективности прогнозирования до 63%. Несмотря на получение достаточно высоких показателей эффективности для 3-х предложенных математических моделей, оценки пропускных способностей не удовлетворяют требованиям по качеству обслуживания. Следовательно, для этого необходимы дополнительные меры, например механизмы «сглаживания» трафика.

Таким образом, разработан метод оценки пропускной способности, учитывающий фрактальные свойства поступающего трафика при обеспечении заданных параметров качества обслуживания.

Литература

1. Костромицкий, А. И. Подходы к моделированию самоподобного трафика / А. И. Костромицкий, В. С. Волотка // Восточно-Европейский журнал передовых технологий, 2010. – №46. – с. 46-49

2. Rose, O. Estimation of the Hurst Parameter of Long Range Dependent Time Series [Электронный ресурс] / O. Rose. – Режим доступа: http://phobos.informatik.uni-wuerzburg.de/TR/tr137.pdf. – Заглавие с экрана.

3. Ложковский, А. Г. Модель трафика в мультисервисных сетях с коммутацией пакетов [Электронный ресурс] / А .Г. Ложковский. – Режим доступа: http://archive.nbuv.gov.ua/portal/natural/nponaz/2010_1/09.pdf. – Заглавие с экрана.

4. Some New Findings on the Self-Similarity Property in Communications Networks and on Statistical End-to-End Delay Guarantee: Technical Report [JNG05-01] / Department of Computer Science, Hong Kong Baptist University; manager Joseph Kee-Yin Ng; contractors: Shibin Song, Bi Hai Tang. – Hong Kong, 2001. – 14 p.

5. Ute Zeigler. Mathematical modeling, simulation and optimization of dynamic transportation networks with applications in production and traffic [Электронный ресурс] / Ute Ziegler. – Режим доступа: http://darwin.bth.rwthaachen.de/opus3/volltexte/2013/4452/pdf/4452.pdf. – Заглавие с экрана.

6. Кричевский, А. М. Прогнозирование временных рядов с долговременной корреляционной зависимостью: дис. канд. техн. наук : 05.13.01 / Кричевский Андрей Михайлович. – СПб., 2008 – 179 с.

7. The Internet TrafficArchive: NASA-HTTP. - Режим доступа: http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html. – Заглавие с экрана.

Климов В.В. Метод прогнозирования оценки пропускной способности. *Необходимость прогнозирования определенных оценок эффективности работы сети, таких как пропускная способность, обусловлена сложностью сбора и обработки соответствующих данных из-за их динамически изменяющегося характера. Целью данной статьи является разработка метода оценки пропускной способности, учитывающего фрактальные свойства поступающего трафика при обеспечении заданных параметров качества обслуживания. В качестве базовой исследуется классическая модель экспоненциального сглаживания, основными временными параметрами которой являются интервалы статистики и регулирования. Решение об их величинах принимались с учетом оценки вероятности потерь и эффективности прогнозирования. Под последним понимается процентное соотношение между спрогнозированными оценками трафика и пропускной способности. Данное условие представлено в аналитическом описании классической модели. Сделан результат о низком уровне оценок эффективности прогнозирования пропускной способности для классической модели и ее модификации. Предложено использовать в качестве модели предсказания трафика ARFIMA-модель. Это позволило повысить эффективность предсказания. Для улучшения это тенденции предложена модель, позволяющая «отслеживать» резкие фрактальные выбросы трафика. Это дает возможность регулировать пропускную способность только по «необходимости». Разработанный метод оценки пропускной способности, учитывающий фрактальные свойства поступающего трафика при обеспечении заданных параметров качества обслуживания.*

Ключевые слова: *оценка эффективности, трафик, фрактальные свойства, экспоненциальное сглаживание, интервалы статистики и регулирования, пропускная способность, вероятность потерь, ARFIMA-модель, эффективность прогнозировани*

V.V. Klimov. Estimating Throughput Predictive Method. *The need to predict certain estimates of network performance, such as throughput, is due to the complexity of collecting and processing the relevant data due to their dynamically changing nature. The purpose of this article is to develop a method for assessing the throughput that takes into account the fractal properties of incoming traffic while ensuring the specified quality of service parameters. The classical exponential smoothing model is investigated as the basic one, the main time parameters of which are the intervals of statistics and regulation. The decision on their values was made taking into account the assessment of the probability of losses and the effectiveness of forecasting. The latter refers to the percentage between the predicted traffic and throughput estimates. This condition is presented in the analytical description of the classical model. A result is made about a low level of estimates of the efficiency of forecasting throughput for the classical model and its modifications It is proposed to use the ARFIMA model as a traffic prediction model. This improved the prediction efficiency. To improve this trend, a model has been proposed that allows "tracking" sharp fractal traffic surges. This makes it possible to adjust the bandwidth only when "needed". The developed method for assessing the throughput, taking into account the fractal properties of the incoming traffic while ensuring the specified quality of service parameters.*

Key words: *efficiency assessment, traffic, fractal properties, exponential smoothing, statistics and regulation intervals, bandwidth, loss probability, ARFIMA model, predicting efficiency*

*Статья поступила в редакцию 19.*09.*2020*
Рекомендована к публикации профессором Павлышом В. Н.

Инженерное образование

УДК 378

ТЕХНОЛОГИЯ ВОРКШОП КАК ДИНАМИЧЕСКИЯ СИСТЕМА ПОДГОТОВКИ БУДУЩИХ СПЕЦИАЛИСТОВ

Е. И. Приходченко
Донецкий национальный технический университет, г. Донецк
88rapoport88@mail.ru

Аннотация
Требования к будущему специалисту на рабочем месте в сегодняшней действительности предъявляются очень высокие. Не исключение составляет и подготовка магистров, умение ими в нестандартной обстановке применить полученные знания. В данной статье рассматривается технология воркшоп, которая направлена на развитие личностных качеств будущего специалиста. Также рассмотрена трёхсферная модель динамических знаний и проанализированы основные подходы к организации технологии воркшопа.

Постановка проблемы

По Европейским образовательным стандартам сегодняшний выпускник высшей школы должен обучаться на протяжении всей жизни, постоянно обновлять свои знания, повышать, развивать и углублять свой уровень компетенций. Технология воркшоп как раз и нацелена на их решение. Различными подходами к организации технологии воркшопа занимались такие ученые, как: В.И. Загвязинский. Его интересовало влияние развивающего влияния на обучение [1]. Педагогическими инновациями в образовании занимались М.Е. Киригина, Н.Р. Юсуфенкова, А.В. Хуторской, В.С. Лазарев, Б.П. Мартиросян, В.М. Полонский, А.И. Вагизова, С. Д. Пивкин, Н.Ш. Валеева, А.И. Мингалеева, К.Ф. Фопель [2-7, 10]. Разработку занятий с применением технологии воркшопа предлагают Е.И. Канивец, Л.Н. Онипченко и др. [8,9].

Цель и задачи исследования

Исходя из поставленной проблемы, изучить теоретический и практический концепты технологии воркшоп. Для этого нами будут решаться следующие задачи:

– провести сравнительный анализ взглядов различных ученых на обозначенную проблему;

– описать методики, применяемые в технологии воркшоп с целью развития более высокого уровня динамической системы знаний

Основное содержание статьи

Требование сегодняшнего времени – обучение на протяжении жизни ставит перед будущими специалистами необходимость повышения компетентности, уровня мотивации к обучению, выработки умений самоорганизовываться, чтоб быть конкурентоспособным на рынке труда. Будущие государственные служащие, как никто другой, должны быть готовы к принятию самостоятельных решений, действовать в нужных жизненных ситуациях ответственно, с максимально положительным вектором их направленности.

Именно технология воркшоп (термин К. Фопеля) направлена на развитие указанных личностных качеств: Термин «воркшоп» обозначает «мастерская», «технология» - от гр. texnos – исскуство, logos – учение; искусство обучать, выбирая из своих «запасов» в «мастерской» те методы (гр. metodos – путь достижения истины), которые наиболее полно решают поставленные проблемы. Ещё Филипп Дормер Стенхон Честерфилд в своём произведении «Письма к сыну» писал, что если один способ оказывается неудачным, попробуй другой и выбирай всякий раз наиболее подходящий. «Воркшоп» - это интенсивное учебное мероприятие, на котором учатся прежде всего благодаря собственной активной работе» [8, С. 143]. Вызвать желание студентов участвовать в занятии можно различными интерактивными методами. Здесь ограничений нет. Главное, чтоб педагог сам владел применяемой методикой на высоком уровне и умел заинтересовать ею студентов, как пример приведен метод «ПОПС-формула»:

– Позиция;
– Обоснование;
– Примеры (не менее трёх);
– Следствие (итог).

Автор данной методики Дэвид Раймон Маккальд Мейтон (КАР).

В центре внимания технологии воркшоп находится «интенсивное групповое взаимодействие в корпоративной культуре, самостоятельность в обучении, умение получать динамические знания, которые и становятся частью профессиональной личности» [8, С. 145].

Динамические знания отличаются от поверхностных и технических, ориентированных на профессиональные умения, на внутреннюю взаимосвязь в той или иной дисциплине, тем, что самостоятельная работа студентов здесь оказывается доминирующей. Они стремятся познавать новое. Они просто хотят учиться. Модель динамического обучения – трёхсферная. В начале обучения необходимо создать ситуацию «расслабленного внимания». Это предпосылки для оптимального функционирования мозга. Второй шаг создания ситуации, в которой студенты могут погрузиться в «комплексный опыт». Третий шаг состоит в активной оценке участниками своего опыта – т.е. обучаемые анализируют свой опыт и проясняют для себя, чему они научились в прошедшей учебной ситуации. Активное оценивание – путь к пониманию, которое выходит за пределы простого воспоминания» [8, с.145].

Принципы динамического обучения базируются на результатах нейропсихологических, неврологических, психологических исследований паттернов (от англ. patron – образец), т.е. принятых в определенной культуре образцов и стереотипов поведения в созданной содержательной, интересной учебной среде. При получении динамических знаний решающее значение имеет возможность получить от обучения удовольствие и почувствовать каждому участнику учебного процесса продуктивным членом группы.

Технология воркшопа – это прямой путь к созданию ситуации успеха для студентов, значительному повышению их желания учиться.

Пробуждая интерес к учебе, повышаем и уровень самооценки, самостоятельности в коллективном деле, формируем аналитические, творческие, коммуникативные социальные навыки, расширяем границы практических компетенций, чтобы уметь действовать в новых, необычных ситуациях, сначала учебных, а в будущем – и в жизненных.

Рассмотрим некоторые подходы к организации технологии воркшоп. Так, Эдвард де Боно предлагает развивать во время проведения занятий с применением выше указанной технологии так называемое латеральное мышление (термин введен вышеуказанным ученым), что в переводе с английского языка обозначает «нестандартное», «боковое», «нелинейное», «движение в сторону», используя полученную информацию как стимул, как призыв к действию, к созданию новых, оригинальных идей, которые должны быть эффективными, неповторимыми.

По мнению автора методики Э. де Боно мышление не заменяет логическое, а дополняет его, позволяет сделать скачок вперёд, освобождает от шаблона.

Интересен метод «Plus Minus Interesting» (PMD). Целью его является поочередное нахождение положительных (Plus), отрицательных (Minus), интересных (Interesting) сторон в анализе и решении проблем.

Генерирующее значение в технологии воркшоп принадлежит динамическим знаниям, знаниям, полученных в деятельности, в активной самостоятельности и коллективной работе. Это могут быть дидактические игры, которые активизируют познавательную деятельность, вырабатывают постоянное стремление к новым, более полным и глубоким знаниям, способствуют созданию благоприятных условий для удовлетворения широкого круга интересов, желаний, запросов, творческих устремлений студенческой молодежи.

Работа в малых группах дает возможность приобрести навыки общения и сотрудничества. Одной из дидактических игр может быть методика «аквариум» – развивает качества иммерсивности, т.е. погружения в предложенную для рассмотрения проблему, ассеритивности – умения строить доказательную базу, аргументированно подводить к результативному умозаключению. Важно мнение каждого участника, никто не должен быть пассивным, ведь инновационное обучение подразумевает развитие креативных способностей к совместным действиям в новых ситуациях, формирование уверенности, настойчивости в достижении целей.

Заслуживает внимания метод проектов (лат. projects – брошенный наперёд), базирующийся на интегративной основе, формируя навыки поиска и обработки необходимой, наиболее важной, целесообразной информации, привлекая к конкретным профессионально важным проблемам, к пониманию роли знаний в жизни и обучении. Проектный метод представляет собой мотивированную практико-ориентированную учебную деятельность студентов, направленную на самореализацию исследовательских способностей, формирование компетенций социального проектирования и моделирования, прекращение их интеллектуального потенциала.

Инновационные подходы в обучении способствуют становлению персонального, личностного пути развития каждого студента, т.е. построению индивидуальной образовательной траектории, способствующей реализации каждого обучаемого его совокупности организационно-деятельности, креативных и познавательных способностей.

По определению Клауса Фопеля, воркшоп – это возможность открыть для себя, что знаешь и умеешь больше, чем думаешь до проведения занятия. Т. Спурнова рассматривает воркшоп как метод обучения, который основан на изучении практических аспектов какого-либо вопроса,

создавая новые, уникальные, зачастую, интеллектуальные продукты, обмениваясь опытом в интерактивном, коммуникативном формате.

Выводы

Таким образом, технология воркшоп имеет стимулирующее воздействие на внутренние мотивы самообразования студентов: формирует их созидательные установки; определяет проблемно-рефлексивный деятельностный подходы; включается педагогическое сопровождение, создание ситуации заинтересованности обучаемых к занятиям; обеспечивает ценностно-ориентированное единство коллектива учебной группы, развивает эффективность творческой деятельности; ориентирует на получение наиболее оптимального результата. Она характеризуется высоким уровнем самостоятельности, возможностью регулировать темп работы и подбирать дифференцированные задачи; совершенствовать память всех участников учебного процесса; развивать навыки исследовательской работы, потребность насыщения личностными знаниями с опытом; определяет индивидуальную позицию в отношении ключевых проблем.

Литература

1. Загвязинский, В. И. Теория обучения: совершенная интерпритация / В.И. Загвязинский: учебное пособие для студ.высш.пед. учеб.заведений – М.:Изд-во «Академия», 2001. – 192 с.

2. Юсуфбекова, Н.Р. Общие основы педагогической инноватики: опыт разработки теории инновационных процессов образовании / Н.Р. Юсуфбекова. – М.: Педагогическое общество, 1991. – 91 с.

3. Хуторской, А.В. Педагогическая инноватика: учебное пособие для студ.высш.учеб.завед / А.В. Хуторской. – М.:Академия, 2008. – 256 с.

4. Лазарев, В.С. Педагогическая инноватика: объект, предмет и основные понятия / В.С. Лазарев, Б.П. Мартиросян // Педагогика, 2004. – №4 – С. 59-67.

5. Полонский, В. М. Инновации в образовании (методологический анализ) / В.М. Полонский // Инновации в образовании, 2007. – №3 – С.108-113.

6. Вагизов, А.И. Подходы к формированию системы инновационного образования РФ / А.И. Вагизов // Вестник Казан.тенол.ун-та, 2011. Т. 14, №23 – С. 293-298.

7. Пивкин, С.Д. Самоопределение обучающихсяв инновационном курсе / С.Д. Пивкин, Н.Ш. Валеева, А.И. Мингалеева // Вестник Казан.тенол.ун-та, 2012.-Т.15, №6. – С. 258-262.

8. Канивец, Е.Н. Неоторые теоритические и практические аспекты организации обучения на основе воркшопа / Е.Н. Канивец, Л.П. Онипченко // Наукові праці ДонНТУ: Серія: педагогіка і психологія і соціологія, 2012. – №12. – С.143-147.

9. Приходченко, Е. И. Педагогика: инновационные подходы / Е. И. Приходченко. – [Саарбрюккен] : LAP Lambert Academic Publishing, 2018. – 689 с.

10. Фопель, К. Ф. Эффективный воркшоп. Динамическое обучение. Перевод с нем. / К. Ф. Фопель – М.: Генезис, 2011. – 368 с.

Приходченко Е. И. Технология воркшоп как динамическая система подготовки будущих специалистов. *Требования к будущему специалисту на рабочем месте в сегодняшней действительности предъявляются очень высокие. Не исключение составляет и подготовка магистров государственной службы, умение ими в нестандартной обстановке применить полученные знания. В данной статье рассматривается технология воркшоп, которая направлена на развитие личностных качеств будущего специалиста. Также рассмотрена трёхсферная модель динамических знаний и проанализированы основные подходы к организации технологии воркшопа.*

Ключевые слова: *магистр государственной службы, технология воркшоп, динамические знания.*

Prikhodchenko K. I. The workshop technology as a dynamic system for future specialists preparing. *The requirements for a future specialist in the workplace in today's reality are very high. Not an exception is the preparation of masters of public service, their ability to apply the acquired knowledge in an unusual environment. This article discusses the technology of the workshop, which is aimed at developing the personal qualities of a future specialist. The three-sphere model of dynamic knowledge is also considered and the main approaches to organizing the technology of the workshop are analyzed.*

Key words: *master of public service, technology workshop, dynamic knowledge.*

Статья поступила в редакцию 21.11.2020
Рекомендована к публикации профессором Мальчевой Р. В.

УДК 004.92

Разработка специализированного устройства на базе ПЛИС для реализации операций сложения и вычитания чисел с плавающей запятой

О. А. Авксентьева, Д. И. Выростков, Р. В. Мальчева
Донецкий национальный технический университет, г. Донецк
avksentievaolga@gmail.com; findarn@gmail.com

Аннотация
В статье рассматриваются особенности проектирования специализированных вычислительных устройств при использовании представления чисел с плавающей запятой. Рассмотрены стандартный и модифицированный алгоритмы сложения и вычитания чисел с плавающей запятой, мантиссы которых представлены в прямом коде. Разработана структура операционного модуля специализированного устройства. Проанализированы базовые элементы программируемых логических интегральных схем, а также особенности проектирования цифровых устройств на их базе. Определены факторы, влияющие на выбор модели реализации алгоритма в базисе ПЛИС. Разработаны варианты реализации устройства сложения и вычитания чисел с плавающей запятой в базисе ПЛИС с использованием только LUT-элементов, а также на базе LUT- и EMB-элементов. Произведено их сравнение и сформулированы выводы и рекомендации по использованию разных схем памяти ПЛИС. Намечены направления дальнейшей работы.

Введение

В настоящее время перед вычислительной техникой стоит не столько задача реализации того или иного алгоритма, сколько соответствие одному или нескольким критериям при его реализации, таким как: быстродействие, надёжность, стоимость и другие. Дополнительными требованиями при разработке являются определённые ограничения, накладываемые на процессы проектирования. Эти ограничения могут задаваться с позиции разных точек зрения: производственной эффективности, рыночной окупаемости, либо с точки зрения учебной или научно-исследовательской деятельности. Таким образом, один и тот же алгоритм можно реализовать с помощью разных технологий, и даже в пределах одной технологии существует бесконечное множество вариаций его представления [1].

Цель исследования и постановка задачи

Значительная доля современных вычислений приходится на операции с целочисленным и дробным представлением чисел. Представление и тех и других описано в стандарте IEEE754-2008 в качестве чисел с плавающей запятой. В некоторых вычислителях математические операции над числами с плавающей запятой выполняются программно, что занимает значительное процессорное время. Поэтому задача разработки специализированного устройства для их осуществления является актуальной задачей, т.к. позволит значительно повысить быстродействие компьютерной системы, выполняющей математические операции над числами, представленными в формате с плавающей запятой [1, 2].

Целью данной работы является анализ алгоритмов и способов реализации в базисе программируемых логических интегральных схем (ПЛИС) арифметических операций над числами, представленными в формате с плавающей запятой.

Для достижения цели в работе решаются следующие задачи:

– анализ особенностей выполнения арифметических операций над числами, представленными в формате с плавающей запятой и различных вариантах кодирования;

– анализ технологии проектирования цифровых элементов в базисе ПЛИС;

– разработка структурной схемы вычислительного модуля;

– реализация вычислительного модуля в базисе ПЛИС;

– определение задач для дальнейших исследований.

Анализ алгоритма и разработка обобщенной структуры операционного модуля специализированного устройства

Сложение и вычитание чисел с плавающей запятой – это элементарные действия, реализация которых необходима в любом вычислителе. В настоящее время для реализации различных вычислительных алгоритмов применяется базис ПЛИС.

Среди достоинств ПЛИС необходимо отметить быстродействие, низкую энергопотребление и, наиболее важное, возможность многократного перепрограммирования внутренней логики схемы. Одна из важных задач при реализации схемы алгоритма в базисе ПЛИС является задача сокращения используемых логических элементов. Решение этой проблемы, с одной стороны, позволяет добиться большего быстродействия и уменьшению потребления энергии. С другой стороны, это позволит запрограммировать одновременно несколько одинаковых алгоритмов, что значительно ускорит обработку некоторого массива данных за счёт параллелизма.

В данной работе предлагается реализовать на ПЛИС модифицированный алгоритм сложения и вычитания чисел с плавающей запятой, у которых порядки чисел представлены в положительном нуле (ПН), а мантиссы – в прямом коде (ПК).

Введём следующие понятия:

A, B – исходные нормализованные числа в заданном формате с плавающей запятой (первый и второй операнды);

C – нормализованное число, результат вычисления;

D – заданное действие (сложение/вычитание).

$$C = A \pm B \rightarrow Nc'MC * 2^{PC} = Na'MA * 2^{PA} \pm Nb'MB * 2^{PB},$$

где Na, Nb, Nc – знаки чисел A, B, C;
MA, MB, MC – мантиссы чисел A, B, C;
PA, PB, PC – порядки чисел A, B, C.

Известно [1, 2], что стандартный алгоритм выполнения сложения/вычитания чисел сводится к следующим шагам:

- проверка мантисс на равенство нулю;
- выравнивание порядков;
- выполнение действия над мантиссами;
- проверка результата на соответствие нормализованному представлению;
- проверка на ряд особых случаев, таких как переполнение, результат равен нулю, потеря точности и т.д.).

При выполнении операции в прямом коде вводится такое понятие как действие над модулями чисел (Dm), которое определяет, которое из действий (сложение или вычитание) будет происходить аппаратно:

$$Dm = \overline{Na}\,\overline{D}\,Nb + Na\,\overline{D}\,\overline{Nb} + \overline{Na}\,D\,\overline{Nb} + Na\,D\,Nb =$$
$$\overline{Na}\left(\overline{D}\,Nb + D\,\overline{Nb}\right) + Na\left(\overline{D}\,\overline{Nb} + D\,Nb\right) =$$
$$\overline{Na}(D\,Nb \oplus D\,Nb) + Na\left(\overline{D\,Nb \oplus D\,Nb}\right) =$$
$$Na \oplus Nb \oplus D.$$

Знак результата *Nc* и модуль мантиссы результата mC вычисляются в соответствии с реализацией конкретного алгоритма. В стандартном алгоритме порядки и мантиссы обрабатываются как независимые друг от друга числа. Стоит отметить, что стандартный алгоритм сложения/вычитания чисел с плавающей запятой является последовательным, что сказывается на временных затратах [3].

Существует модифицированный алгоритм сложения/вычитания чисел с плавающей запятой с представлением мантисс в прямом коде. Он сводится к представлению порядков и мантисс единым модулем чисел, над частями которых выполняются разные, но взаимосвязанные операции.

Известно, что для перехода из дополнительного кода (ДК) в ПН при выполнении операций на двоичном сумматоре необходимо инвертировать знак числа. Также известно, что коды для представления положительных чисел в ПК и ДК совпадают [1].

Таким образом, при действии $|A| + \overline{|B|} + 1$ (вычитании в ДК) на двоичном сумматоре получен результат в формате PC'MC. Из-за особенностей арифметики в ПН, PC без последующего инвертирования знака получим в ДК, а, поскольку действие над мантиссами производится над модулями, то MC тоже будет представлена в ДК. Кроме этого, необходимо также сравнить порядки, то есть произвести вычитание PA – PB с учетом арифметики положительного нуля.

Для разработки устройства введём следующие обозначения:

E_{outp} – выходной перенос сумматора, на котором выполняется вычитание порядков;

$ZERO_m$ – флаг, указывающий на то, что результат вычисления мантисс равен нулю;

$ZERO_p$ – флаг, указывающий на то, что результат вычисления порядков равен нулю.

Введём промежуточную переменную $ZERO_s = ZERO_m * ZERO_p$, тогда:

если $E_{outp} * ZERO_S$, то $|A| = |B|$;

если $E_{outp} * \overline{ZERO_S}$, то $|A| > |B|$;

если $\overline{E_{outp}} * \overline{ZERO_S}$, то $|A| < |B|$.

Введём переменную - флаг w, отображающую равенство/неравенство порядков.

Пусть w = 1, если $|A| > |B|$.

Таким образом, знак результата NC определяется следующим образом:

$$Nc = w * Na \vee \overline{w}(Nb \oplus D).$$

Обобщенная структура операционного модуля представлена на рис. 1. Несмотря на представление в модифицированном алгоритме порядка и мантиссы единым двоичным модулем,

для их обработки всё равно используются разные комбинационно-логические схемы и память.

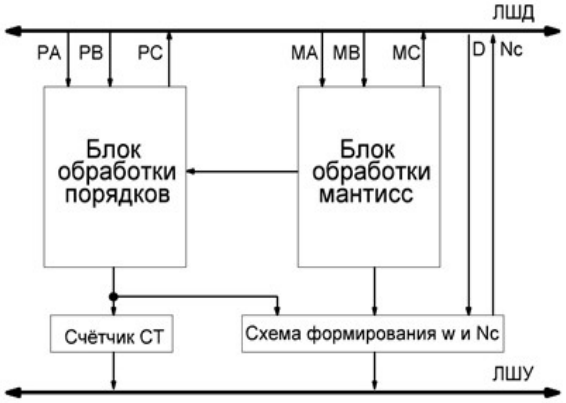


Рисунок 1 - Структура операционного модуля специализированного устройства

Таким образом, выделяются блок обработки порядков и блок обработки мантисс. Через локальную шину данных (ЛШД) в блоки поступают исходные значения – порядки (PA, PB) и мантиссы (MA, MB).

Эти же блоки определяют значения результата – также порядок (PC) и мантиссу (MC). Разница порядков PA, PB из блока обработки порядков поступает в счётчик, что определяет количество необходимых сдвигов мантиссы числа с наращиваемым порядком.

Эта информация поступает на локальную шину управления (ЛШУ), которая связана с ЛШД через шинный интерфейс.

В операционном модуле представлены несколько комбинационно-логических схем (КЛС), которые формируют признак равенства-неравенства чисел (w) и знак результата (Nc) на основе данных из обоих блоков, а также заданного действия над числами (D).

От рассмотренной структуры перейдем к блочной диаграмме операционного модуля специализированного устройства (рис. 2) .

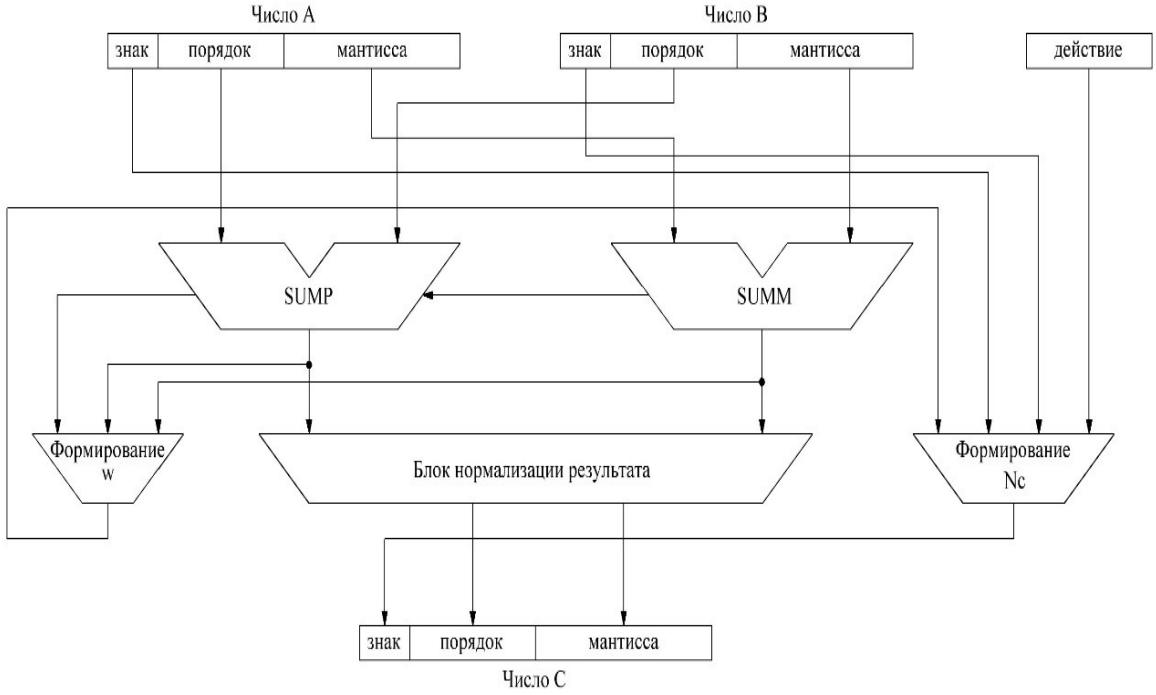


Рисунок 2 – Блочная диаграмма операционного модуля специализированного устройства

Операции над порядками осуществляются в комбинационно-логической схеме (КЛС) SUMP, операции над мантиссами – в комбинационно-логической схеме SUMM. Результаты вычислений блоков помещаются в блок нормализации результата, а также в КЛС формирования признака равенства/неравенства порядков w. После этот признак, знаки чисел A и B, а также действие над числами поступает в КЛС формирования знака результата Nc.

SUMP и SUMM представляют собой не только сумматоры, но и необходимую временную память для хранения операндов, либо счётчик (в случае SUMP).

Реализация операционного модуля в базисе ПЛИС

Реализация различных алгоритмов на базе ПЛИС можно свести к двум этапам: проектирование отдельных функционально целостных блоков; логическое соединение готовых блоков для получения конечной функциональности устройства.

Следуя этим шагам, для реализации устройства необходимо выполнить проектирование следующих функциональных блоков:

- обработки порядков;
- обработки мантисс;
- нормализации результата;
- формирования признака равенства/неравенства порядков;
- формирования знака результата.

Основной ресурс ПЛИС - конфигурируемый логический блок (Configurable Logic Block - CLB) - состоит из таблицы соответствия (Look Up Table - LUT), мультиплексора и D-триггера (рис.3).

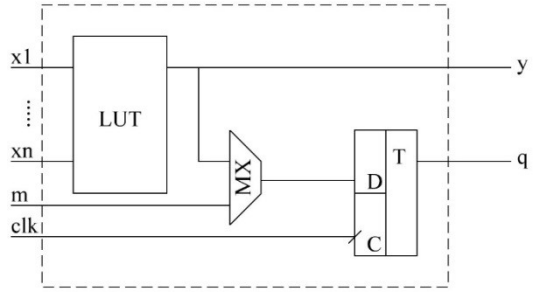


Рисунок 3 - Конфигурируемый логический блок: общая структура

Входными данными CLB является набор входов x1…xn для LUT-элемента, сигнал тактовой частоты CLK и флаг синхронной/асинхронной записи m. Выходные данные представлены результатом срабатывания логической функции LUT-элемента и значением, которое хранит D-триггер. Таким образом, логический элемент ПЛИС, в зависимости от реализации, может выступать и как элемент комбинационно-логических схем, и как элемент памяти [4, 5]. На рис.4. показана структура LUT-элемента (элемент табличного типа, программируемая таблица соответствий, таблица перекодировки).

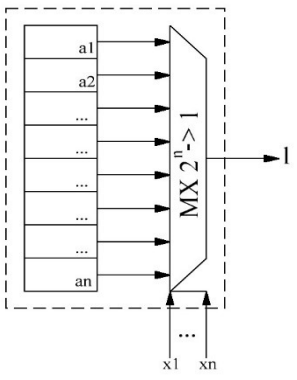


Рисунок 4 - Структура типового LUT-элемента

Он представляет из себя набор 2^n ячеек статической памяти, в которые при компиляции логики ПЛИС загружаются все возможные состояния заданной логической функции. Это обеспечивает высокое быстродействие логического элемента ПЛИС, т.к. нужные значения уже записаны и хранятся в специальных ячейках конфигурируемой памяти. Несмотря на наличие у этих ячеек адресов (a1:an), архитектура ПЛИС не предоставляет доступ к ним. Все они подключены к мультиплексору $2^n \rightarrow 1$ с n адресных входов. В качестве адресных входов используются входные переменные логической функции (x1:xn), которые и определяют выбираемый из памяти результат l. Количество входов LUT-элемента может быть различно даже в пределах архитектуры одной ПЛИС, т.к. многие ПЛИС поддерживают несколько режимов работы основного логического элемента. Преимуществом LUT-элементов является равное время задержки (срабатывания) логической функции для функций с одинаковым количеством переменных. Оценка времени задержки таким образом становится значительно проще [5].

На рис.5 приведена структурная схема операционного модуля в ПЛИС-базисе. На схеме LUT-блоки, используемые в качестве схем памяти, названы LUT-mem:

LUT-mem1, LUT-mem2, LUT-mem5 – регистры устройства, хранящие порядок чисел A, B, C соответственно;

LUT-mem3, LUT-mem4, LUT-mem6 – регистры устройства, хранящие мантиссы чисел A, B, C соответственно;

LUT-mem7 – счётчик, хранящий разность порядков и определяющий количество сдвигов.

LUT-блоки, реализующие КЛС, обозначены как LUT-GL:

LUT-GL1 – КЛС блока обработки порядков, архитектурно организованная как сумматор положительного нуля;

LUT-GL2 – КЛС блока обработки мантисс, архитектурно организованная как сумматор прямого кода;

LUT-GL3, LUT-GL4, LUT-GL7 – коммутаторы регистров порядка;

LUT-GL5, LUT-GL6, LUT-GL9 – коммутаторы регистров мантисс;

LUT-GL8 – коммутатор счётчика;

LUT-GL10 – КЛС, определяющая равенство/неравенство порядков w;

LUT-GL11 – КЛС, определяющая знак результата Nc;

LUT-GL12 – КЛС, анализирующая текущее значение в счётчике.

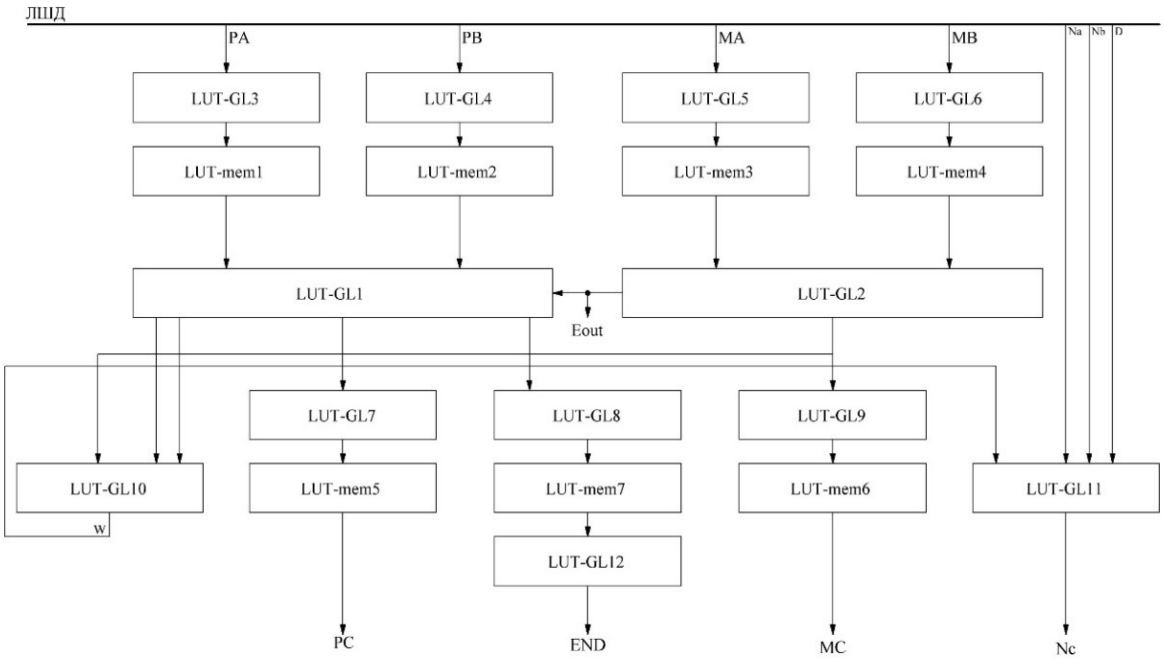


Рисунок 5 – Структурная схема устройства сложения и вычитания чисел с плавающей запятой в базисе ПЛИС (LUT- элементы)

Помимо LUT-элементов, память в архитектуре ПЛИС зачастую может быть представлена различными встроенными блоками памяти (Embedded Memory Block – EMB) . Учитывая, что на представление одного разряда требуется один логический элемент ПЛИС, память можно реализовать на блоках встроенной памяти. Как показали проведенные ранее исследования [6, 7], это позволит значительно сократить количество LUT-элементов в схеме. Структурная схема устройства сложения и вычитания чисел с плавающей запятой в базисе ПЛИС показана на рис.6.

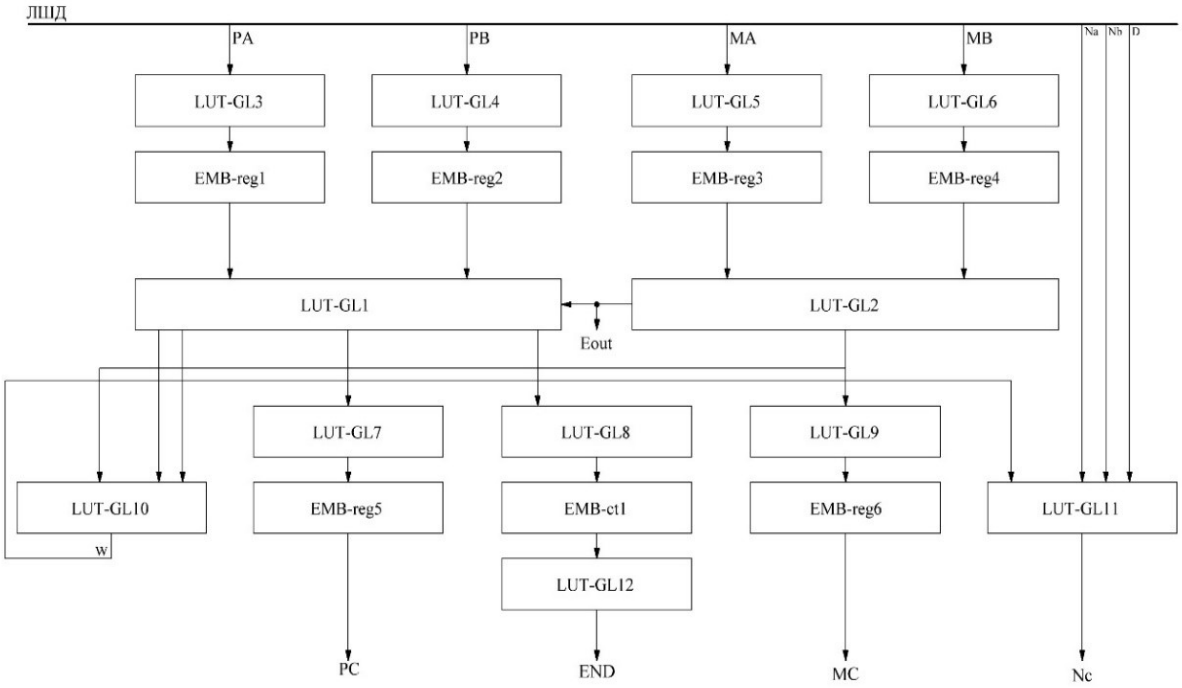


Рисунок 6 – Структурная схема устройства сложения и вычитания чисел с плавающей запятой в базисе ПЛИС (LUT- и EMB-элементы)

На схеме регистры, реализуемые на блоках выделенной памяти, обозначены как EMB-reg, а счётчик – EMB-ct:

EMB-reg1, EMB-reg2, EMB-reg5 - регистры устройства, хранящие порядок чисел A, B, C, соответственно;

EMB-reg3, EMB-reg4, EMB-reg6 - регистры устройства, хранящие мантиссу чисел A, B, C, соответственно;

EMB-ct1 - счётчик, хранящий разность порядков и определяющий количество сдвигов.

Выводы

На примере алгоритма сложения/вычитания чисел с плавающей запятой рассмотрены возможные модели синтеза операционных модулей в базисе ПЛИС. Конечная реализация зависит от поставленных целей. В общем случае, необходимо выбирать между количеством занятых конфигурируемых логических блоков и быстродействием устройства.

Приоритетной задачей при программировании ПЛИС является сокращение количества используемых конфигурируемых логических блоков (и, следовательно, элементов табличного типа). Это достигается с помощью отказа от использования CLB в качестве элементов памяти, поскольку для одного разряда требуется один CLB, который может включать в себя несколько LUT-элементов. При этом элементы памяти реализуются в встроенной памяти ПЛИС – EMB-блоках. С другой стороны, обращение к этим блокам осуществляется медленнее, чем к блокам CLB.

Значительно выгоднее реализовывать схемы памяти на EMB-блоках, несмотря на некоторое снижение быстродействия устройства, поскольку CLB, несмотря на возможность использования в качестве простого триггера, имеют сложную внутреннюю логику, позволяющую реализовывать многие арифметические операции на аппаратном уровне. Также, поскольку количество этих блоков в кристалле ПЛИС ограничено, использовать CLB-блоки в качестве схем памяти зачастую нецелесообразно.

Направления дальнейших исследований

В качестве направлений дальнейших исследований намечаются следующие этапы:

- более детальное проектирование вычислительного модуля с учетом рекомендаций [5 – 7];
- моделирование и тестирование модуля с использованием UML и HDL [8, 9];
- включение полученной модели в обеспечение виртуальной лаборатории [10, 11], разработанной и используемой на кафедре компьютерной инженерии;
- использование для изучения и моделирования архитектур процессорных элементов при подготовке бакалавров и магистров по направлению «Информатика и вычислительная техника».

Литература

1. Зыков А.Г. Арифметические основы ЭВМ [Текст] / А.Г. Зыков, В.И. Поляков. – Санкт-Петербург: Университет ИТМО, 2016. – 140 с.

2. Баркалов, А. А. Классические принципы организации и проектирования центрального процессора [Текст]/ А. А. Баркалов, Л. А. Титаренко, В. Н. Опанасенко. – Донецк: УНИТЕХ, 2011. – 338 с.

3. Сергиенко А.М. Архитектура компьютеров: Конспект лекций [Текст] / А. М. Сергиенко – Киев: НТУУ «КПИ», 2015. – 198 с.

4. Потехин В.А. Схемотехника цифровых устройств: учебное пособие для вузов [Текст] / В.А. Потехин – Томск: В-Спектр, 2012. – 250 с.

5. Barkalov A. Logic Synthesis for FPGA-Based Control Units—Structural Decomposition in Logic Design [Текст] / A. Barkalov, L. Titarenko, K. Mielcarek, S. Chmielewski // Lecture Notes in Electrical Engineering. Germany, Berlin: Springer, 2020. Volume 636. – 240 p.

6. Баркалов, А. А. Уменьшение числа LUT-элементов в схеме автомата Мили / А. А. Баркалов, Р. В. Мальчева // Научные труды Донецкого национального технического университета. Серия: Вычислительная техника и автоматизация. Донецк: ДОННТУ, 2013.- № 2(25). - С. 168-174.

7. Баркалов, А. А. Оптимизация схемы автомата Мура, реализуемой в базисе ПЛИС / А. А. Баркалов, Р. В. Мальчева, К. А. Солдатов // Радиотехника, информатика, управление, 2012.- № 1(26). - С. 44-47.

8. Jasinski R. P. Effective Coding with VHDL. Principles and best practice [Текст] / R. P. Jasinsky - The MIT Press, 2016. – 524 p.

9. Malcheva, R. Application of Multilevel Design on the base of UML for Digital System Developing / R. Malcheva // Lecture Notes in Electrical Engineering. 2011. Т. 79. С. 93-117.

10. Мальчева, Р. В. Разработка виртуальной лаборатории для изучения и моделирования архитектур процессорных элементов / Р. В. Мальчева, О.А. Авксентьева // Программная инженерия: методы и технологии разработки информационно- вычислительных систем (ПИИВС-2016): сборник научных трудов I научно-практической конференции. 16-17 ноября 2016 г. – Донецк: ГОУВПО «ДОННТУ», 2016. - С. 102-108.

11. Мостовая, В. А. Разработка виртуальной лаборатории для изучения архитектуры процессора / В. А. Мостовая, О. А. Авксентьева, Р. В. Мальчева // В сборнике: Современные информационные технологии в образовании и научных исследованиях (СИТОНИ-2019). Материалы VI Международной научно-технической конференции. Под общей редакцией В. Н. Павлыша. – Донецк: ГОУВПО «ДОННТУ», 2019. - С. 407-411.

Авксентьева О. А., Выростков Д. И., Мальчева Р. В. Разработка специализированного устройства на базе ПЛИС для реализации операции сложения и вычитания чисел с плавающей запятой. *В статье рассматриваются особенности проектирования специализированных вычислительных устройств при использовании представления чисел с плавающей запятой. Рассмотрены стандартный и модифицированный алгоритмы сложения и вычитания чисел с плавающей запятой, мантиссы которых представлены в прямом коде. Разработана структура операционного модуля специализированного устройства. Проанализированы базовые элементы программируемых логических интегральных схем, а также особенности проектирования цифровых устройств на их базе. Определены факторы, влияющие на выбор модели реализации алгоритма в базисе ПЛИС. Разработаны варианты реализации устройства сложения и вычитания чисел с плавающей запятой в базисе ПЛИС с использованием только LUT-элементов, а также на базе LUT- и EMB-элементов. Произведено их сравнение и сформулированы выводы и рекомендации по использованию разных схем памяти ПЛИС. Намечены направления дальнейшей работы.*

Ключевые слова: *устройство, плавающая запятая, ПЛИС, LUT-элемент, EMB-элемент.*

Avksentieva O. A., Vyrostkov D. I., Malcheva R. V. Developing of FPGA-based specialized device to realize the addition and subtraction of numbers with floating point notation. *The article discusses the design features of specialized computing devices when using the floating-point representation of numbers. The standard and modified algorithms for adding and subtracting floating-point numbers, whose mantissas are represented in the direct code, are considered. The structure of the operating module of a specialized device is developed. The basic elements of programmable logic integrated circuits, as well as the design features of digital devices based on them, are analyzed. The factors influencing the choice of the algorithm implementation model in the FPGA basis are determined. Variants of the implementation of the device for adding and subtracting floating-point numbers in the FPGA basis using only LUT elements, as well as on the basis of LUT and EMB elements, have been developed. Their comparison is made and conclusions and recommendations on the use of different FPGA memory schemes are formulated. The directions of further work are outlined.*

Keywords: *device, floating-point, FPGA, LUT elements, EMB elements.*

Статья поступила в редакцию 21.11.2020
Рекомендована к публикации профессором Зори С. А.

УДК 004.942

Тестирование аналоговых и аналогово-цифровых схем методами цифровой обработки сигналов

Д.О. Нестеренко, Ю.Е. Зинченко, В.Н. Соленов
Донецкий национальный технический университет
des_jawa@mail.ru, zinchenko.yuri@gmail.com

Рассмотрены проблемы тестирования различных подклассов схем. Определен наиболее перспективный подход для реализации автоматизированной системы диагностики на базе аппаратных средств цифровой обработки сигналов. Проведен сравнительный анализ рабочих частот, объемов встроенной памяти и стоимостных параметров микросхем, производимых компаниями Altera и Xilinx. Определен оптимальный вариант элементной базы для построения автоматизированных систем диагностики.

Введение

Современный мир немыслим без техники. Она настолько прочно вошла в нашу жизнь, что мы порой даже не замечаем, как часто имеем дело с ней. С развитием цифровых технологий связан новый этап эволюции техники: она становиться все компактнее, быстрее, умнее. Цифровые схемы прочно вошли в нашу жизнь и заняли позицию верного помощника. Однако, наряду с цифровыми, существовали и будут существовать аналоговые и аналогово-цифровые схемы. Они представляют собой механизмы сопряжения цифровых схем с окружающей средой. В чистом виде аналоговые схемы могут встречаться в источниках питания, различных генераторах и так далее. Как и любые схемы имеют тенденцию выходить из строя, так и исследуемый класс аналоговых и аналогово-цифровых схем, разработанных уже более десятка лет тому назад, тоже часто содержит неисправности. Неисправности, как правило, проявляются в отклонении эксплуатационных параметров от нормы или в полной потере функциональности электронных схем.

Класс электронных схем очень широк, однако его можно условно разделить на три основных подкласса: аналоговые, цифровые и комбинированные аналогово-цифровые схемы. Конкретные электронные устройства чаще всего состоят из целого комплекса модулей и элементов с различной схемотехникой и природой сигналов. То есть любая схема может содержать как цифровые, так и аналоговые узлы.

Проблема тестирования схем этих подклассов является неоднозначной задачей. Так, для диагностики цифровых узлов применяются методы синтеза цифровых тестов и исследования тестовых реакций. При этом проверяется наличие или отсутствие логических уровней в контрольных точках схемы. Метод анализа аналоговых схем сложнее, поскольку в процессе диагностики необходимо также исследовать амплитудные, частотные характеристики сигналов, их спектральные составляющие.

Наиболее эффективными средствами борьбы с неисправностями схем являются автоматизированные системы диагностики (АСД). Наибольшей популярностью пользуются те их них, которые базируются на принципах цифровой обработки сигналов (ЦОС). Они гарантируют необходимую точность измерений параметров, а также высокую скорость и возможность автоматизации процесса диагностики неисправных схем.

Исследования

Принципы ЦОС значительно изменили традиционное устройство систем диагностики. Одним из важнейших изменений является тот факт, что современные АСД на базе ЭВМ с ЦОС полностью имитируют аппаратное обеспечение диагностического комплекса. То есть такие системы не имеют традиционных аналоговых инструментов, как вольтметры, амперметры, осциллографы и др. Физически в этих системах присутствует ЭВМ, её периферийные устройства и интерфейсные схемы связи с объектом диагностики. Однако, отсутствие традиционных аналоговых приборов совсем не означает снижения функциональности или направленности АСД на тестирование цифровых схем. Аналоговая часть таких систем реализована в виде программных моделей. Измерительные средства выполнены в виде компьютерных программ и вследствие этого лишены многих традиционных проблем, таких как: перекрестные помехи, нелинейность, шум, плавание, неправильная калибровка, температурный эффект и даже старение [1].

С целью повышения быстродействия и

мобильности, а также для упрощения сложности программного обеспечения АСД структура диагностического комплекса аналогово-цифровых схем может быть реализована на основе современных вычислительных устройств на базе FPGA со встроенными ядрами ЦОС. Среди производителей данного класса микросхем наиболее известны Altera и Xilinx [2].

Ведущим мировым производителем FPGA Xilinx была разработана среда Embedded Development Kit (сокращенно, EDK), которая предоставляет возможность использования процессорных элементов и целого спектра периферийных устройств. Количество стандартных компонентов непрерывно увеличивается и меняется, в настоящее время среди процессорных элементов наиболее распространены семейства MicroBlaze и PowerPC [3]. Структура MicroBlaze построена по принципу гарвардской архитектуры, в основе которой положена концепция разделенных шин адреса и команд. Организация магистралей микропроцессора, согласно этой концепции, обеспечивает достижение высокой скорости выполнения операций [4].

Основными элементами архитектуры микропроцессорных ядер семейства MicroBlaze являются:

- 32-разрядное АЛУ;
- блок регистров общего назначения;
- блок выполнения операций с плавающей запятой;
- регистр состояния (статуса);
- программный счетчик (счетчик команд);
- буфер команд (конвейерный регистр);
- блок дешифрации команд;
- контроллер интерфейса шины команд;
- контроллер шины данных;
- регистры специального назначения Exception Address Register (EAR) и Exception Status Register (ESR);
- настраиваемая логика;
- кэш-память команд;
- кэш-память данных;
- схема управления прерываниями.

Настраиваемая логика и кэш-память команд и данных не является обязательным элементом архитектуры. Необходимость их наличия определяется разработчиком согласно конфигурации проектируемой системы [5].

При проектировании систем на кристалле на основе микропроцессорных ядер семейства MicroBlaze, разработчик может использовать набор компонентов периферийных модулей, представленных в виде IP-ядер (Intellectual Property) [6].

В состав проекта могут входить:

- таймер счетчик (Timer/Counter);
- сторожевой таймер;
- универсальный асинхронный приёмопередатчик UART (Universal Asynchronous Receiver/Transmitter);
- контроллер прерываний Interrupt Controller (INTC);
- контроллер стандартного интерфейса ввода/вывода (GPIO Controller);
- контроллер интерфейса памяти SDRAM (SDRAM Controller);
- контроллер интерфейса памяти DDR SDRAM (DDR SDRAM Controller);
- контроллер интерфейса памяти Flash Memory (Flash Memory Controller);
- контроллер интерфейса блочной памяти BlockRAM (BRAM Controller;
- контроллер интерфейса Serial Peripheral Interface (SPI Master and Slave Bus Controller), отвечающий спецификациям фирмы Motorola;
- другие интерфейсные модули проектировщика.

Одним их наиболее распространенных отладочных комплексов фирмы Xilinx является Spartan 3E (S3E), который полностью удовлетворяет элементной базе благодаря наличию всех необходимых компонентов м возможности удобного использования преимуществ шинной архитектуры при программировании FPGA-устройств.

Исходя из условий проектирования АСД могут быть использованы среды разработки ISE и EDK, входящие в состав комплексной среды отладки проектов для плат фирмы Xilinx – Xilinx Platform Studio (XPS), которая поставляется с соответствующими устройствами. ISE ориентирована на разработку и отладку HDL-модулей, используемых в качестве периферии для МП Microblaze, программируемых с помощью EDK.

Среди микросхем со встроенной возможностью ЦОС фирмы Altera можно выделить следующие: Cyclone III, Arria GX, Stratix III. Микросхема Arria GX оборудована каналами последовательной связи и поэтому чаще всего используется в качестве моста, соединяющего интерфейс PCI Express с Gigabit Ethernet или Serial Rapid IO. Также она может использоваться в качестве сопроцессора с ЦП или процессором ЦОС. Микросхема Cyclone III объединяет в себе такие качества, как низкое энергопотребление, высокая функциональность и низкая цена, что делает ее незаменимой для использования в качестве контроллера в следующих отраслях: машиностроение, военная промышленность и др., где возникают задачи контроля дисплеев любых размеров, обработки видео и аудио информации, беспроводных коммуникаций. Stratix III является наиболее функциональным и дорогим решением, содержит аппаратную реализацию ЦОС. Эта микросхема с успехом используется для построения аппаратуры беспроводных сетей, устройств

вещания аудио и видео информации, аппаратуры тестирования и диагностики, измерительных средств. Она также может быть использована для реализации военных и медицинских задач.

Сравнительная характеристика микросхем приведена на рис. 1 – 4.

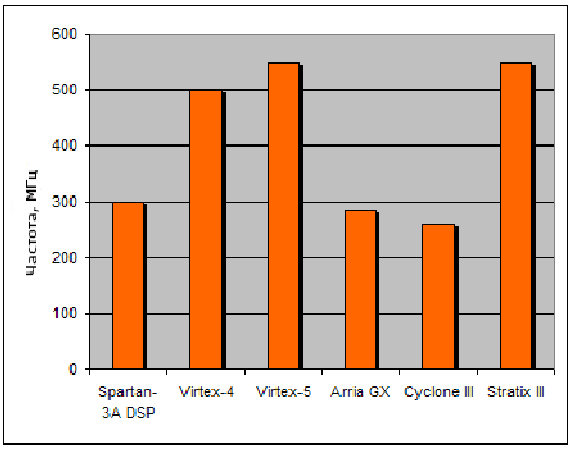


Рисунок 1 – Сравнительный анализ рабочих частот микросхем компаний Altera и Xilinx

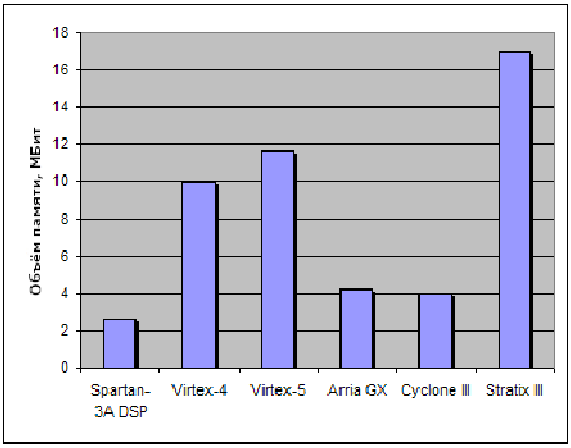


Рисунок 2 – Сравнение объёмов встроенной памяти

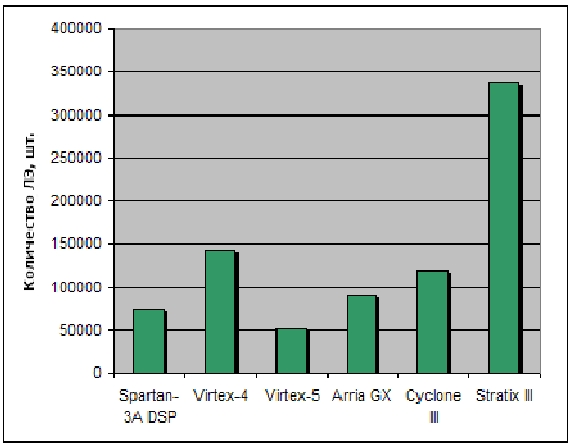


Рисунок 3 – Количество логических элементов

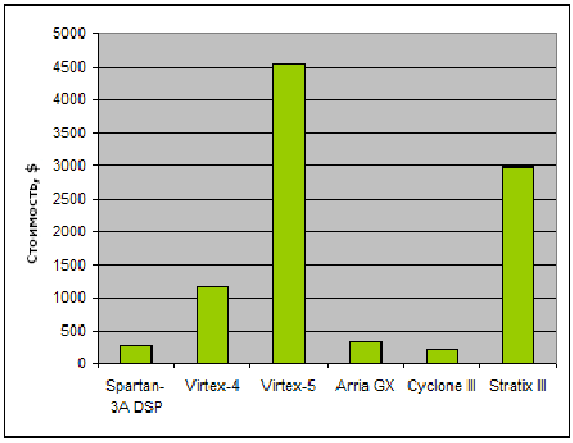


Рисунок 4 – Сравнение средних стоимостей микросхем

Аналогичными продуктами фирмы Xilinx является микросхемы: Virtex-4, Virtex-5 и Spartan-3 и др. Семейство Virtex используются при проектировании оборудования различных проводных и беспроводных сетей, систем цифровой обработки сигналов (видео и аудио информации) и др. Серия микросхем Spartan-3 используется в оборудовании сетей, устройств обработки видео и аудио, в дисплейных панелях, а также в различных периферийных устройствах.

Выводы

Анализ этих данных позволяет сделать следующий вывод: наиболее функциональными решениями среди микросхем с аппаратной реализацией методов ЦОС, являются микросхемы серий Virtex и Stratix, но они очень дороги. Поскольку одной из приоритетных задач является небольшая стоимость системы диагностики, то для её реализации следует выбрать продукт Cyclone III компании Altera.

Продукция компании Altera с архитектурой FPGA семейства Cyclone III широко применяется для интеграции в портативных устройствах, объединяя в себе такие качества, как невысокая себестоимость, низкое энергопотребление, и в то же время способна обрабатывать большие объемы данных.

Успешное применение устройств FPGA в области диагностики аналоговых и аналогово-цифровых подтверждено исследованиями и разработками кафедры компьютерной инженерии ДонНТУ [7-11].

Литература

1. Matthew Mahoney. DSP-Based Testing of Analog and Mixed-Signal Circuits // IEEE Computer Society Press Los Alamitos, California.

2. Максфилд, К. Проектирование на ПЛИС. Курс молодого бойца. – Москва, 2007.

3. Kabisatpathy, P. Fault Diagnosis of Analog Integrated Circuits / Prithviraj Kabisatpathy, Alok Barua, Satyabroto Sinha // Springer, 2005.

4. Зотов, В. Embedded Development Kit — система проектирования встраиваемых микропроцессорных систем на основе ПЛИС серий FPGA фирмы Xilinx / В. Зотов // Компоненты и технологии, 2003. – № 3.

5. Зотов, В. MicroBlaze — семейство тридцатидвухразрядных микропроцессорных ядер, реализуемых на основе ПЛИС фирмы Xilinx / В. Зотов // Компоненты и технологии, 2003. – № 4.

6. MicroBlaze Processor Reference Guide: [Online resource]. – Access mode: http://www.xilinx.com/support/documentation/sw_ manuals/xilinx11/mb_ref_guide.pdf.

7. Масякин, Е.А. Разработка и исследование методов и структур аппаратного генерирования аналоговых тестов на базе FPGA / Е.А. Масякин, Т.А. Зинченко, Ю.Е. Зинченко // Материалы VI международной научно-технической конференции студентов, аспирантов, молодых учёных " Информатика и компьютерные технологии (ИКТ-2010)". 23-25 ноября 2013 г. - Донецк: ДонНТУ, 2010. – С. 69-75.

8. Шигимагин, А.В. Разработка и исследование методов и структур аппаратного анализа аналоговых тестовых реакций на базе FPGA / А.В. Шигимагин, Ю.Е. Зинченко // Материалы VI международной научно-технической конференции студентов, аспирантов, молодых учёных " Информатика и компьютерные технологии (ИКТ-2010)". 23-25 ноября 2013 г. - Донецк: ДонНТУ, 2010. – С. 76-82.

9. Коваленко, И.А. Идентификация и моделирование параметров аналоговых устройств на базе специально обученных нейронных сетей / И.А. Коваленко, А.М. Ковалев, Е.В. Лобанов, Ю.Е. Зинченко // «Информатика и компьютерные технологии», сборник трудов VIII международной научно-технической конференции студентов, аспирантов и молодых учёных – 18-19 сентября 2012 г. – Донецк: ДонНТУ, 2012. - С. 90-95.

10. Коваленко, И.А. Диагностика и обнаружение неисправностей в аналоговых интегральных схемах с использованием искусственных нейронных сетей и метода псевдослучайного тестирования / И.А. Коваленко, А.М. Ковалев, Е.В. Лобанов, Ю.Е. Зинченко, В.В. Ханаев // «Информационные управляющие системы и компьютерный мониторинг», сборник трудов III Всеукраинской научно-технической конференции студентов, аспирантов и молодых учёных - 16-18 апреля 2012 г. -Донецк: ДонНТУ, 2012. - С. 631-637.

11. Бєловолова, М. А. Архітектура системи діагностики аналогових пристроїв на базі нейронної мережі та спектрального аналізу / М. А. Бєловолова, Ю. Є. Зінченко // Наукові праці Донецького національного технічного університету, серія: «Проблеми моделювання та автоматизації проектування», 2013. № 1 (12)-2(13).

Нестеренко Д.О., Зинченко Ю.Е., В.Н. Соленов. Тестирование аналоговых и аналого-цифровых схем методами цифровой обработки сигналов. *Рассмотрены проблемы тестирования различных подклассов схем. Определен наиболее перспективный подход для реализации автоматизированной системы диагностики на базе аппаратных средств цифровой обработки сигналов. Проведен сравнительный анализ рабочих частот, объемов встроенной памяти и стоимостных параметров микросхем, производимых компаниями Altera и Xilinx. Определен оптимальный вариант элементной базы для построения автоматизированных систем диагностики.*

Ключевые слова: *тестирование, цифровая обработка, сигнал, микросхема, сравнение.*

D. Nesterenko, Y. Zinchenko, V. Solenov. Using digital signal processing methods for analog and analog-digital circuits testing. *The problems of testing various subclasses of schemes are considered. The most promising approach for the implementation of an automated diagnostic system based on digital signal processing hardware is determined. A comparative analysis of the operating frequencies, the amount of internal memory and the cost parameters of the chips produced by Altera and Xilinx is carried out. The optimal variant of the element base for building automated diagnostic systems is determined.*

Keywords: testing, digital processing, signal, chips, comparing.

Статья поступила в редакцию 04.12.2020
Рекомендована к публикации профессором Скобцовым Ю.А.

Об авторах

Авксентьева Ольга Александровна (1951 г. р.) – старший преподаватель кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Андриевская Наталия Климовна (1971 г. р.) – старший преподаватель кафедры автоматизированных систем управления факультета компьютерных наук и технологий Донецкого национального технического университета.

Бельков Дмитрий Валерьевич (1965 г. р.) – кандидат технических наук, доцент, доцент кафедры прикладной математики факультета компьютерных наук и технологий Донецкого национального технического университета.

Выростков Дмитрий Иванович (1999 г. р.) – студент 4 курса группы КС-18 кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Едемская Евгения Николаевна (1960 г. р.) – старший преподаватель кафедры искусственного интеллекта и системного анализа факультета компьютерных наук и технологий Донецкого национального технического университета.

Зинченко Юрий Евгеньевич (1959 г. р.) – кандидат технических наук, доцент, доцент кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Исаков Андрей Юрьевич (1997 г. р.) – выпускник магистратуры группы КСм-18 кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Крахмаль Мария Вячеславовна (1996 г. р.) – аспирант кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Климов Владимир Владимирович (1981 г. р.) – аспирант ГО ВПО «Донецкий институт железнодорожного транспорта».

Мальчева Раиса Викторовна (1958 г. р.) – кандидат технических наук, доцент, профессор кафедры компьютерной инженерии, заместитель по науке декана факультета компьютерных наук и технологий Донецкого национального технического университета.

Нестеренко Денис Олегович (1985 г. р.) – студент 2 курса магистратуры группы КСзм-19 кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета

Ногтев Евгений Александрович (1993 г. р.) – ассистент кафедры программной инженерии Донецкого национального технического университета.

Приходченко Екатерина Ильинична (1950 г. р.) – доктор педагогических наук, профессор, профессор кафедры социологии и политологии социально-гуманитарного института Донецкого национального технического университета, заслуженный учитель Украины, академик Международной академии наук педагогического образования.

Соленов Вячеслав Николаевич (1997 г. р.) – студент 1 курса магистратуры группы КСм-20 кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Чередникова Ольга Юрьевна (1972 г. р.) – кандидат технических наук, доцент, доцент кафедры компьютерной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Чернышова Анна Викторовна (19хх г. р.) – старший преподаватель кафедры программной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Щедрин Сергей Валерьевич (1975 г. р.) – ассистент кафедры программной инженерии факультета компьютерных наук и технологий Донецкого национального технического университета.

Щербов Игорь Леонидович – проректор Донецкого национального технического университета.

**Требования к статьям,
направляемым в редакцию научного журнала
«Информатика и кибернетика»**

Редколлегией принимаются к рассмотрению статьи, в которых рассматриваются важные вопросы в области информатики и кибернетики. Научный журнал издаётся с 2015 года, периодичность издания – 4 раза в год.

В журнале предусмотрены следующие рубрики:

- информатика и вычислительная техника;
- компьютерные и информационные науки;
- инженерное образование.

В соответствии с номенклатурой специальностей научных работников МОН ДНР первые две рубрики соответствуют следующим укрупненным группам специальностей научных работников:

05.01 – «Инженерная геометрия и компьютерная графика»,

05.13 – «Информатика, вычислительная техника и управление».

С 01.02.2019 Научный журнал включён в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёной степени кандидата наук, на соискание учёной степени доктора наук (приказ МОН ДНР № 135) по группам специальностей 05.01.00 и 05.13.00.

Рубрика «Инженерное образование» предназначена опубликования сотрудниками научно-методических статей.

Журнал также включён в базу данных РИНЦ (Российский индекс научного цитирования) (лицензионный договор № 425-07/2016 от 14.07.2016).

Статьи, представляемые в данный сборник, должны отвечать следующим требованиям. ***Содержание статьи*** должно быть посвящено актуальным научным проблемам и включать следующие необходимые элементы:

- постановку проблемы в общем виде, её связь с важными научными и практическими задачами;
- анализ последних исследований и публикаций, в которых решается данная задача и на которые опирается автор, выделение нерешенных ранее частей общей проблемы, которым посвящается статья;
- формулировка цели статьи и постановка задач, решаемых в ней;
- изложение основного материала с полным обоснованием полученных научных результатов;
- выводы и перспективы последующих исследований в данном направлении.

Каждый элемент должен быть выделен соответствующим названием раздела, например, «введение», «постановка задачи», «цель и задачи работы», «цель статьи», «цель исследования», «цель разработки», «анализ … », «сравнительная оценка … », «разработка … », «проектирование … », «программная реализация», «тестирование … », «полученные результаты», «выводы», «литература». Разделы «введение», «выводы», «литература» являются обязательными. Включать в названия разделов нумерацию не разрешается.

В основном тексте статьи формулируются и обосновываются полученные авторами утверждения и результаты. Выводы должны полностью соответствовать содержанию основного текста. Языки публикаций: русский, английский.

Объём статьи, формат страницы

Для оформления статьи следует использовать листы формата А4 (210х297 мм) с полями по 2,5 см со всех сторон. Нумерацию страниц выполнять не нужно.

Рекомендуемый объём статьи – 6-12 страниц. Рукописи меньшего объёма могут быть рекомендованы к публикации в качестве коротких сообщений.

Последняя страница текста статьи должна быть заполнена не менее чем на две трети, но содержать не менее трёх пустых строк в конце.

Форматирование текста

Подготовка статьи осуществляется в текстовом редакторе Microsoft Office Word.

Весь текст статьи оформляется шрифтом Times New Roman 10 пт с одинарным междустрочным интервалом, если ниже в требованиях не сказано иного. Абзацный интервал «перед» – 0 пт, «после» – 0 пт.

На первой строке с выравниванием по левому краю располагается УДК.

Заголовок (название) статьи оформляется шрифтом Times New Roman 14 пт, полужирное начертание, с выравниванием по центру (без абзацных отступов). Заголовок статьи следует печатать с прописной буквы без точки в конце, переносы слов не допускаются. Абзацный интервал «перед» – 12 пт, «после» – 12 пт.

После названия статьи следует информация об авторах, которая выравнивается по центру (без абзацных отступов). На одной строке указываются инициалы и фамилии всех авторов через запятую. Между двумя инициалами ставится пробел. С новой строки указывается название вуза (организации) и город (для каждого автора, если не совпадают). На следующей строке указываются адреса электронной почты (один адрес либо каждого автора – по желанию). Адрес электронной почты оформляется в виде гиперссылки.

К тексту аннотации применяется курсивное начертание, с выравниванием по ширине, отступы слева и справа по 1 см. Заголовок «Аннотация» выделяется полужирным начертанием. Объём аннотации – 450-550 символов (без пробелов). Абзацный интервал «перед» – 12 пт, «после» – 12 пт.

Основной текст статьи разбивается на две колонки шириной по 7,5 см (промежуток между столбцами – 0,99 см), выравнивается по ширине. Абзацный отступ первой строки – 1 см. Автоматический перенос слов не применяется.

Заголовки разделов выполняются шрифтом Arial 10 пт, полужирное курсивное начертание. Абзацный отступ отсутствует, интервал перед абзацем – 12 пт, после абзаца – 6 пт. Для заголовка «Введение» установить интервал «перед» – 0 пт, «после» – 6 пт.

Таблицы в тексте статьи

Название следует помещать над таблицей с абзацного отступа (1 см) в формате: слово «Таблица», пробел, номер таблицы, пробел, тире, пробел, название таблицы. Название таблицы записывают с прописной буквы без точки в конце строки и выравнивают по ширине. В ячейках таблицы устанавливается выравнивание текста по центру по вертикали. По горизонтали текст выравнивается по центру либо по левому краю. Границы ячеек таблицы должны быть только чёрного цвета, толщина линии – 1 пт. На все таблицы должны быть приведены ссылки в тексте статьи, при ссылке следует писать слово «табл.» с указанием её номера, например, « … данные приведены в табл. 5». Таблицы нумеруются в пределах статьи. Таблица располагается сразу после ссылки на неё, если это возможно (например, после окончания абзаца). Если же таблица не помещается на текущей странице, то она должна быть расположена в начале следующей страницы (или колонки). При необходимости допускается включение в статью таблицы, ширина которой превышает ширину колонки. В этом случае таблица и её название размещаются по центру страницы. Таблица не должна выступать за границы полей страницы. Таблица и её название отделяются от основного текста статьи одной пустой строкой до и после.

Рисунки в статье

Ссылки на иллюстрации по тексту статьи обязательны и оформляются в виде « ... на рис. 2» и т. п. Рисунок и его подпись выравниваются по центру колонки (без абзацных отступов), положение рисунка – «в тексте». Размещается рисунок после его первого упоминания в тексте, если это возможно (например, после окончания абзаца). Если же иллюстрация не помещается на текущей странице, то она должна быть расположена в начале следующей страницы (или колонки). При необходимости допускается включение в статью рисунка, ширина которого превышает ширину колонки. В этом случае рисунок и его подпись выравниваются по центру страницы. Иллюстрация не должна выступать за границы полей страницы. Подпись рисунка оформляется в формате: слово «Рисунок», пробел, номер иллюстрации, пробел, тире, пробел, название рисунка. Название рисунка записывают с прописной буквы без точки в конце строки. Для подписи иллюстрации применяют курсивное

начертание. Иллюстрация и её подпись отделяются от основного текста статьи одной пустой строкой до и после. Не допускается выполнять рисунки с помощью встроенного графического редактора Microsoft Office Word. Если на иллюстрации имеется текст, размер шрифта должен быть не менее чем аналогичный текст, набранный шрифтом Times New Roman 10-го размера. Иллюстрация не должна содержать много незаполненного пространства.

Формулы

Формулы и уравнения рекомендуется набирать с использованием MathType (предпочтительно) или MS Equation. Формулы и математические символы не должны существенно отличаться по размеру от основного текста. Обязательной является нумерация формул, на которые имеется ссылка в тексте статьи. Ссылки в тексте на порядковые номера формул дают в скобках, например, « ... согласно формуле (2)». Формулы размещаются по центру колонки, а их номера – по правому краю. Как для строки с формулой, так и для первой строки пояснений (при наличии), абзацный отступ убирается. Первая строка пояснения начинается со слова «где», после которого следует поставить табуляцию на 1 см, затем само пояснение в формате: символ, подлежащий объяснению, пробел, тире, пробел, поясняющий текст, запятая, обозначение единицы измерения физической величины. Пояснения перечисляются через точку с запятой, выравниваются по ширине. Вторая и последующие строки пояснений начинаются с абзацного отступа (1 см). Весь блок текста, связанный с формулой (только формула, несколько формул подряд или формула с пояснениями), отделяется от основного текста одной пустой строкой до и после. Переносить формулы на следующую строку допускается только на знаках выполняемых операций, причем знак в начале следующей строки повторяют. При переносе формулы на знаке умножения применяют знак «×». Формулы и математические уравнений могут быть записаны в тексте документа, если их высота не превышает высоту строки. При этом следует учитывать, что знаки математических операций отделяются от чисел или символов пробелами с обеих сторон. Например, «Если учесть, что $y < 0$ и $2x + y = 1$, то из формулы (3) можно выразить x…». К символам, которые приведены в формуле, при дальнейшем их употреблении (в том числе в пояснениях к формуле) должно применяться курсивное начертание. При этом к любым числам (верхние и нижние индексы, содержащие цифры и т. п.), а также к математическим знакам курсивное начертание не применяется. Не допускается вставлять формулы, выполненные в виде рисунков.

Перечисления: оформление списков

Основной текст статьи может содержать перечисления, оформленные в виде маркированного списка. В качестве маркера элемента списка разрешается использовать только короткое тире «–». Каждый элемент перечисления записывается с новой строки с абзацного отступа, равного 1 см. После символа короткого тире текст располагается с отступом в 1,5 см от левой границы строки, выравнивается по ширине, при переносе на новые строки располагается без отступов. Нумерованные и многоуровневые списки включать в статью не разрешается.

Литература

В тексте статьи обязательны ссылки на все литературные источники, номер источника указывается в квадратных скобках. Ссылки на неопубликованные работы не допускаются. Рекомендуемое количество источников, на которые ссылается автор, не менее 10. Перечень источников приводится в порядке их упоминания в статье. Библиографическое описание каждого литературного источника оформляется в соответствии с ГОСТ Р 7.0.100–2018. Перечень литературных источников оформляется в виде нумерованного списка. В качестве маркеров элементов списка используют порядковые арабские цифры с точкой. Каждый источник представляет собой отдельный элемент перечисления, записывается с новой строки с абзацного отступа, равного 1 см. После порядкового номера с точкой текст располагается с отступом в 1,5 см от левой границы строки, выравнивается по ширине, при переносе на новые строки располагается без отступов.

В конце статьи обязательно приводятся аннотации на русском и английском языках, каждая заканчивается перечнем 5-6 ключевых слов.

К тексту аннотации применяется курсивное начертание, с выравниванием по ширине, отступы слева и справа по 1 см. Слово «Аннотация» опускается. Текст аннотации начинается с ФИО авторов и названия статьи, выделяемых полужирным начертанием. Аннотация на русском языке совпадает с аннотацией, приведенной в начале статьи. В тексте аннотации на английском языке после фамилии автора указывается только первая буква имени с точкой. Абзацный интервал «перед» – 12 пт, «после» – 12 пт. Ключевые слова оформляются с новой строки аналогично тексту аннотации. Заголовок «Ключевые слова:» (англ. «Keywords:») выделяется полужирным начертанием. Ключевые слова перечисляются через запятую.

Порядок представления статьи и сопроводительные документы

В редакцию необходимо представить:

– файл с текстом статьи;

– файл, содержащий фамилию, имя и отчество авторов полностью; ученую степень, ученое звание; место работы с полным указанием должности, подразделения и наименования организации, города (страны); номера телефонов и e-mail для связи;

– экспертное заключение о возможности публикации статьи, подписанное руководителем и заверенное печатью организации, в которой работает автор статьи;

– выписка из заседания кафедры или письмо организации с просьбой об опубликовании и указанием, что изложенные в статье результаты ранее не публиковались.

Статьи и сопроводительные документы следует высылать на электронный адрес infcyb.donntu@yandex.ru.

К сведению авторов

Если статья оформлена с нарушением указанных выше требований и правил, редакция после предварительного рассмотрения может отклонить статью.

На рецензирование статьи направляются членам редакционной коллегии журнала. Все статьи публикуются при наличии положительной рецензии.

В статью могут быть внесены изменения редакционного характера без согласования с автором. Ответственность за содержание статьи и качество перевода аннотаций несут авторы.

Публикация статей в научном журнале «Информатика и кибернетика» осуществляется на некоммерческой основе.

Все номера Научного журнала размещаются на сайте http://infcyb.donntu.org/.

.

CONTENT

Informatics and computer engineering

Engineering education

Электронное периодическое издание

Научный журнал

ИНФОРМАТИКА И КИБЕРНЕТИКА

(на русском, английском языках)

№ 3 (21) - 2020

Ответственный за выпуск Р. В. Мальчева

Технический редактор Р. В. Мальчева

Компьютерная верстка А. И. Воронова

Подписано к выпуску 10.12.2020. Усл. печ. лист. 9,7. Уч.-изд. лист. 6,5.
Адрес редакции: ДНР, 83001, г. Донецк, ул. Артема, 58, ГОУ ВПО «ДонНТУ»,
4-й учебный корпус, к. 36., ул. Кобозева, 17.
Тел.: +38 (062) 301-07-35, +38 (071) 334-89-11
E-mail: infcyb.donntu@yandex.ru, URL: http://infcyb.donntu.org