

Главный редактор
Дмитрий Волков, с.н.с., ИПМ РАН

Научный редактор
Наталья Дубова

Редакционный совет:

Валерий Аджиев, к.т.н., с.н.с.,
Национальный центр компьютерной анимации,
Университет Борнмута (Великобритания);

Фуад Алескерев, д.т.н., профессор, НИУ ВШЭ;

Михаил Горбунов-Посадов, д.физ.-мат.н.,
зав. отделом ИПМ РАН, доцент, МГУ;

Юрий Зеленков, д.т.н., зав. кафедрой прикладной
информатики, Финансовый университет
при Правительстве РФ;

Сергей Кузнецов, д.физ.-мат.н., профессор, МГУ;

Михаил Кузьминский, к.хим.н., с.н.с., ИОХ РАН;

Александр Легалов, д.т.н., профессор, СФУ;

Владимир Сухомлин, д.т.н., профессор, МГУ;

Павел Храмцов, к.т.н., доцент, МИФИ;

Игорь Федоров, к.т.н., профессор, МЭСИ;

Виктор Шнитман, д.т.н., профессор, МФТИ;

Леонид Эйсымонт, к.физ.-мат.н., научный
консультант, НИИ «Квант»

Корректор Ирина Карпушина

Верстка и графика Мария Рыжкова

Дизайн обложки Денис Киров

Адрес для корреспонденции:

127254, г. Москва, а/я 42

Телефоны:

+7 495 725-4780/84, +7 499 703-1854
+7 495 725-4785 (распространение, подписка)

Факс: +7 495 725-4783

E-mail: osmag@osp.ru

Подписной индекс:

99482 — «Каталог российской прессы» (МАП)

72773 — Объединенный каталог «Пресса России» АПР

59869 — «Каталог. Издания органов научно-технической
информации»

Реклама

000 «Рекламное агентство «Чемпион» +7 499 750-0467



**ОТКРЫТЫЕ
СИСТЕМЫ**
Open Systems Publications

© 2016 Издательство «Открытые системы»

Журнал зарегистрирован
в Министерстве РФ по делам печати,
телерадиовещания и средств массовых коммуникаций
03.07.2015
Свидетельство ПИ № ФС 77-62328

Журнал выходит 4 раза в год

Цена свободная

Выпуск издания осуществлен
при финансовой поддержке
Федерального агентства по печати
и массовым коммуникациям

Учредитель и издатель:

000 «Издательство «Открытые системы»
Россия, 127254, Москва,
проезд Добролюбова, дом 3, строение 3

Президент Михаил Борисов

Генеральный директор Галина Герасина

Директор ИТ-направления Павел Христов

Коммерческий директор Татьяна Филина

Все права защищены.
При использовании материалов
необходимо разрешение редакции и авторов.

В номере использованы иллюстрации
и фотографии: 000 «Издательство «Открытые
системы» и IEEE Computer Society.

Отпечатано в 000 «Богородский
полиграфический комбинат»
142400, Московская область,
г. Ногинск,

ул. Индустриальная, д. 40Б
(495) 783-9366, (49651) 73179

Тираж 4 000 экз.

Содержание №1 (211) 2016

НОВОСТИ. ФАКТЫ. ТЕНДЕНЦИИ.

Начало конца шифрования

Как поссорились разработчики баз данных in-memory

Исполнилось 25 лет первой в мире веб-странице

Операторы связи поддержат Open Compute Project

В DataCamp появится групповое обучение

основам обработки данных

За четыре года на неиспользуемое ПО

растрачено 30 миллиардов долларов

«Умные» счетчики

Данные «холодного» хранения

Для заблокированных по решению властей

веб-страниц ввели особый код

Большая группа по Большим Данным

Eclipse Che выходит в бета-версии

Алгоритм упростит общение с роботами

Watson в роли доктора и продавца

Oracle прекратит выпуск плагина Java

ПЛАТФОРМЫ

10 Рейтинг Top500 и прогресс

высокопроизводительных

вычислений

Эрих Штраймер, Ханс Мойер,

Джек Донгарра, Хорст Саймон

На протяжении двух десятилетий рейтинг
Top500 пользовался успехом в качестве мерила
производительности суперкомпьютеров и был
первоисточником, позволяющим проследить
технологические тенденции. Но насколько этот
рейтинг применим в эпоху экзамастических
суперкомпьютеров?

В ФОКУСЕ: МАШИННОЕ ОБУЧЕНИЕ

14 Практические аспекты машинного обучения

Виктор Китов

Сегодня с машинным обучением связывают
много надежд, однако успешность его приме-
нения определяется не только выбором адек-
ватного задачи алгоритма, но и правильными
шагами на этапах планирования, разработки и
внедрения модели.

18 Машинное обучение для понимания естественного языка

Павел Велихов

Технологии обработки естественного языка
сегодня шагнули далеко вперед, и немалая
заслуга в этом принадлежит машинному обу-
чению, применяемому, в частности, для пони-
мания текстов.

22 Машинное обучение для планирования запросов

Олег Иванов

Рост объемов данных требует от СУБД уве-
личения производительности выполнения
запросов. Оптимизация плана выполнения
запроса с использованием средств машинного
обучения позволяет в разы уменьшить время
его обработки.

26 Машинное обучение для Больших Данных

Андрей Орлов

Теоретические основы машинного обучения по-
явились практически одновременно с первыми
компьютерами, однако при его практическом
применении всегда приходится учитывать спе-
цифику конкретных систем. Работа с Большими
Данными средствами Nadoop требует инстру-
ментов адаптации алгоритмов машинного
обучения к этой платформе.

28 Оценка компетентностей студентов на основе анализа социальных сетей

Михаил Захаров, Анатолий Карпенко,
Елена Смирнова

В третьем поколении Федеральных государ-
ственных образовательных стандартов выс-
шего профессионального образования задачи
обучения формулируются в терминах компе-
тентностей учащихся, что делает актуальной
задачу их оценки. На помощь приходят средст-
ва машинного обучения, позволяющие оценить
компетентности студентов на основе анализа
их поведения в социальных сетях.

ИНТЕРНЕТ ВЕЩЕЙ

31 Референтные архитектуры для Интернета вещей

Михаэль Вейрих, Кристофер Эберт

Интернет вещей призван предоставить новые
возможности развития и повысить произво-
дительность труда благодаря свободе комму-
никации различных устройств друг с другом.
Однако воплощению этих идей мешает не-
хватка архитектурных шаблонов, описываю-
щих протоколы связи с учетом особенностей
конкретных отраслей.

34 Интернет вещей для энтузиастов: метеостанция с визуализацией

Эдуард Лебедев

Интернет вещей развивается сегодня бла-
годаря не только компаниям, предлагающим
решения, специально предназначенные для
работы в Интернете вещей, но и энтузиастам,
которым остро необходимы новые инстру-
менты работы с взаимодействующими «ум-
ными» вещами.

СУБД

36 Стоимостные оптимизаторы для СУБД: вчера и сегодня

Леонид Борчук

Оптимизаторы запросов для реляционных
СУБД прошли долгий путь развития, превра-
тившись сегодня в сложную программу оценки
стоимости вариантов, однако в СУБД нового
поколения стоимость выполнения запросов
во внимание почти не принимается. Что это —
шаг назад или два вперед? И нужна ли вообще
преемственность для стоимостных оптимиза-
торов мира Больших Данных?

ОПЫТ

40 Частное облако для масс

Владимир Сигунов

На базе OpenStack сегодня можно построить
частное облако. Однако опыт реальных проектов
показывает, что малым и средним предприя-
тиям требуются типовые масштабируемые аппа-
ратно-программные решения, поставляемые
вместе с технической поддержкой.

МИР

42 ДИСПАК — операционная система атомного проекта

Ирина Крайнева

Атомный проект сыграл определяющую роль в
политическом и социально-экономическом раз-
витии СССР в послевоенный период, в частно-
сти, став стимулом к развитию вычислительной
техники, вычислительной математики и про-
граммирования. Одной из наиболее известных
программистских разработок, зародившихся в
недрах атомного проекта, стала операционная
система ДИСПАК для ЭВМ БЭСМ-6.

ИТ-УНИВЕРСИТЕТЫ

44 На пути к Открытой Науке

Сергей Парин

Новые информационные системы для органи-
зации научных коммуникаций позволяют по-
высить ответственность ученого за качество
выпускаемых работ и обеспечить более широ-
кое и открытое использование их результатов.
Одна из таких систем — «Соционет».

БИБЛИОТЕКА

46 ИТ на рубеже эпох

Александр Тьренко

Темы ноябрьского, декабрьского, январского
и февральского номеров журнала Computer
(IEEE Computer Society, Vol. 48, No. 11, 12 2015
и Vol. 49, No. 1, 2 2016) — суперкомпьютеры
экзафлопсной производительности, перспе-
ктивы развития компьютеров после окончания
действия закона Мура, слияние человека с
кибернетическими системами и новые тен-
денции в мире облаков.