

В.Г. Абашин



Экология данных

Первая редакция

Монография



Москва 2025

В.Г. Абашин

МОНОГРАФИЯ

«Экология данных»

Москва 2025

УДК 004.6
ББК 32.973
А 136

Авторы:

Абашин В.Г., канд. тех. наук, доцент.

Рецензенты:

Коротеев М.В., канд. эк. наук, доцент, заведующий кафедрой искусственного интеллекта факультета информационных технологий и анализа больших данных Финансового университета при Правительстве РФ;

Лукиянов П.В., к.т.н, доцент кафедры информационных систем и цифровых технологий, Орловского государственного университета имени И.С. Тургенева.

Абашин, Валерий Геннадьевич

А 136 Экология данных. Монография – М.: Мир науки, 2025. – Сетевое издание. Режим доступа: <https://izd-mn.com/PDF/46MNNPM25.pdf> – Загл. с экрана.

ISBN 978-5-908044-08-0

DOI: 10.15862/46MNNPM25

Монография посвящена новому подходу оценки и сравнения информационных систем – экологии данных. Он базируется на формализуемом понятии токсичности данных и занимает промежуточное место между вероятностным и объёмным подходом к измерению информации. Экология данных позволяет оценивать потенциальные негативные факторы, к которым может привести использование конкретной информационной системы. Также в монографии приведён обзор современных подходов и технических решений к хранению, передаче и обработке информации.

Монография будет полезна студентам ИТ специальностей, а также направлений, связанных с менеджментом и управлением в сфере ИТ.

ISBN 978-5-908044-08-0

© Абашин Валерий Геннадьевич

© ООО Издательство «Мир науки», 2025

Оглавление

Введение	6
ГЛАВА 1. Информация и данные в современном мире. Экологический аспект	8
1.1. Информация и данные	8
1.2. Глобальные технологии хранения и обработки данных	10
1.3. Жизненный цикл аппаратного обеспечения вычислительных устройств, комплексов и систем	16
1.4. Влияние аварий и катастроф на экологичность данных	18
1.5. Общие сведения об информационной экологии	22
ГЛАВА 2. Введение в экологию данных	24
2.1. Токсичность данных, токсичность информационных систем	24
2.2. Токсичность алгоритмов обработки данных	27
2.3. Экология данных и вычисления	28
2.4. Экология данных сервисов на базе вычислительных сетей	32
ГЛАВА 3. Общие технологии управления данными с точки зрения экологии данных	37
3.1. Экологичность на примере ряда базовых информационных технологий	37
3.2. Интерфейсы обмена данными	44
3.3. Вклад тестирования и конфигурирования оборудования в экологичность	47
3.4. Влияние резервирования данных на их токсичность	49
ГЛАВА 4. Специализированные подходы к хранению данных с точки зрения экологии данных	53
4.1. Файловые системы как структурированное хранилище информации	53
4.2. Базовые данные	59
4.3. Специализированные базы данных	62
4.4. Описание методологии определения токсичности хранения информации	64
ГЛАВА 5. Распределенные технологии хранения и управления данными с точки зрения экологии данных	67
5.1. Файловая система как распределённая система хранения данных	67
5.2. Распределённые СХД и СУБД	70
5.3. Облачные распределённые СХД и СУБД	73
5.4. Сравнение токсичности методов распределённого хранения с точки зрения экологии данных	75
ГЛАВА 6. Эффективное управление данными	77
6.1. Выбор критерия для определения эффективности управления данными	77

6.2. Экология данных и многокритериальная система управления данными	80
6.3. Основные концепции эффективного управления данными	82
6.4. Проектирование персональной системы управления данными	86
ГЛАВА 7. Безопасность хранения данных	91
7.1. Общие принципы защиты данных	91
7.2. Организация безопасного хранения данных	93
7.3. Криптографические методы защиты	94
7.4. Управление рисками и реагирование на инциденты	95
ГЛАВА 8. Перспективные направления в хранении данных	98
8.1. Перспективные аппаратные технологии в хранении данных	98
8.2. Перспективные концепции хранения данных	100
Заключение	104
Литература	106

Введение

Информация является объективной неопределимой сущностью нашего мира. Она многогранна и изучается с различных точек зрения. Исторически первым подходом к измерению информации стал энтропийный подход, который отталкивается от цели применения информации. Объёмный подход связывает цель обработки и хранения информации с имеющейся материальной средой, пригодной для работы с информацией. В данной монографии предлагается рассмотреть информационные технологии с точки зрения их влияния на экологию планеты, человека, социальные отношения.

В предыдущий период развития информационных технологий доминировали концепции призывающие сохранять всю имеющуюся информацию, создавая цифровую копию реальной жизни. Данное стремление было обусловлено крайней нехваткой ресурсов для получения, хранения и обработки информации. На сегодняшний день мы живем в условиях информационного шторма, который начинает неуправляемо деформировать современную цивилизацию, разрушать социальные связи, формировать персональные информационные пузыри, приводить к функциональной инвалидности человека за счёт формирования клипового мышления.

Новые подходы к обработке информации сфокусированы на решении технических проблем, связанных с повышением скорости и обработки информации. Большие данные — отталкиваются от концепции учёта особенностей аппаратной реализации и их ограничений. Причём эта концепция не позволяет повысить уровень автоматизации для всех задач. Например, при создании распределённых таблиц снова требуется вычислять или предугадывать необходимый объём хранения. Предметная область снова оказывает влияние на выбор решения для хранения данных. Приходится выбирать чем жертвовать: скоростью, надёжностью, целостностью данных. Использование чисто технических подходов приводит к существенному дисбалансу. Открытая библиотека Википедия (wikipedia.org) находится на 7 месте по посещаемости на планете и генерирует трафик сопоставимый с мировыми лидерами интернета, однако её бюджет примерно в 1665 раз меньше чем у Альфабет.

Основной принцип экологии данных это умеренность. В основе монографии положен датацентрический подход при работе с данными. Считается что данные несут ценность, а если они несут ценность, то они должны быть управляемы. Это повышает общую управляемость организацией, делает её более эффективной. Управление данными порождает новые профессии, такие как инженер данных, шеф по данным и другие.

Цель монографии — обобщение наработок в области работы с данными в виде нового подхода оценки качества данных и информационных систем с целью повышения их эффективности.

Задачами монографии являются: определение и формализация понятия экологии, токсичности данных и информационных систем; систематизация подходов к управлению данными; оценка токсичности различных подходов к работе с данными и их свойств.

Для полноценного освоения материала необходимо иметь начальные знания об операционных системах на базе GNU/Linux, базовые навыки работы с эмулятором терминала, иметь базовое представление о программировании на языке sh и Python. Однако в случае отсутствия указанных навыков монография все равно позволит ознакомиться с основными современными концепциями хранения и обработки информации.

Монография будет интересна научным сотрудникам интересующимся вопросами развития средств хранения и обработки информации, информационной экологией, технической экологией, для студентов технических направлений желающих ознакомиться с современными подходами к хранению информации, а также для управленцев и экономистов, желающих разобраться в современных высокоуровневых концепциях работы с данными.

Данные это стихия. С ними мы можем справиться только вместе.

Заключение

Токсичные информационные системы наследие взрывного роста популярности вычислительных машин. Следующий этап развития ЭВМ должен обеспечить повышение экологичности вычислительной техники. В данной монографии предпринимается попытка очертить круг проблем, границ, в рамках которых будут производиться дальнейшее исследование токсичности алгоритмов обработки данных и хранения данных. Данные вопросы требуют отдельной детальной проработки.

Предлагается считать экологию данных подмножеством энтропийного подхода к измерению информации с целью, обладающей некоторой априорной определенностью, связанной с требованием к минимизации негативного воздействия информации и технологий её обработки на окружающую среду, человека, социальные отношения. В этом смысле экология данных с одной стороны не является энтропийным подходом, т. к. учитывает влияние на экологию информационной системы, т. е. является только одним свойством энтропийного подхода, с другой стороны не может быть описана объёмным подходом, т. к. учитывает влияние энтропийной составляющей на человека и окружающий мир в целом.

Чтобы избежать токсичности информационная система должна быть максимально динамичной, но достаточно статичной чтобы быть полезной. Для поддержания информационных систем в таком состоянии требуются специалисты занимающиеся управлением данными. Без этих людей нельзя собрать и подготовить данные к обработке, сложно защитить данные.

Токсичность алгоритмов можно рассматривать как вероятностный подход конкретизированный для целей определения токсичности. Сложность применения токсичности алгоритмов связана с отсутствием методологии её определения. Для создания такой методологии требуется отдельное исследование. Также отсутствует статистика токсичности элементов информационных систем и используемых в них типов данных (персональных, критических, оперативных для разных ИС, прочих). Сбор статистических данных, которые позволят формализовать расчет токсичности, вычислять токсичность данных и информационных систем является одной из основных задач дальнейших исследований.

На основании вышеизложенного был сделан вывод, тестирование и конфигурирование как элементы жизненного цикла влияют на токсичность информационной системы, а значит они также требуют оценки с точки зрения экологии данных. Это направление также является целью дальнейших исследований. Влияние этапов жизненного цикла на экологичность данных слабоформализуемая проблематика, требующая большого количества статистических данных.

В результате исследования выявлена взаимосвязь отсутствия стратегии резервирования и однозначного повышения токсичности информационной системы, однако количественное измерение этой характеристики возможно только в условиях определенности используемого аппаратного и программного обеспечения, а также при наличии описания информационной системы и данных, используемых в ней. Исследование количественных характеристик токсичности подсистем резервирования должно быть статистически обосновано и является частью дальнейших исследований.

Регламентные меры с точки зрения экологии данных также могут становиться источником токсичности и требуют контроля за их соблюдением и постоянного обновления в соответствии с изменяемыми условиями.

Оптимизация информационных систем, выражаемая в том числе через уменьшение токсичности, трансформируется в социальную ответственность бизнеса, которая невозможна без комплексного повышения эффективности, что приводит к снижению расходов и повышению привлекательности корпоративного бренда.

Абашин Валерий Геннадьевич

Экология данных

Монография издана в авторской редакции

Сетевое издание

Компьютерный набор, верстка – Абашин В.Г.

Ответственный за выпуск – Алимова Н.К.

Научное издание

Системные требования:

операционная система Windows XP или новее, macOS 10.12 или новее, Linux.

Программное обеспечение для чтения файлов PDF.

Объем данных 3,19 Мб

Принято к публикации «22» июля 2025 года

Режим доступа: <https://izd-mn.com/PDF/46MNNPM25.pdf> свободный. – Загл. с экрана. –
Яз. рус., англ.

ООО «Издательство «Мир науки»

«Publishing company «World of science», LLC

Адрес:

Юридический адрес – 127055, г. Москва, пер. Порядковый, д. 21, офис 401.

Почтовый адрес – 127055, г. Москва, пер. Порядковый, д. 21, офис 401.

<https://izd-mn.com/>

**ДАННОЕ ИЗДАНИЕ ПРЕДНАЗНАЧЕНО ИСКЛЮЧИТЕЛЬНО ДЛЯ ПУБЛИКАЦИИ НА
ЭЛЕКТРОННЫХ НОСИТЕЛЯХ**