

ПРОБЛЕМЫ ХРАНЕНИЯ И УТИЛИЗАЦИИ ИНФОРМАЦИИ

Елисов Максим Вячеславович¹
Самарский университет, г. Самара

Аннотация. Статья посвящена исследованию особенностей хранения информации. Предложена модель каталогизации данных с учетом предложенной авторами классификации информации. Проведен экономический анализ современных способов хранения и обработки информации. Предложен способ оптимизации бизнес-процессов.

Ключевые слова: информация, данные, хранение данных, оптимизация, утилизация данных, цифровизация.

PROBLEMS OF STORAGE AND DISPOSAL OF INFORMATION

Yelisov M.V.
Samara University, Samara

Abstract. The article is devoted to the study of the features of information storage. A data cataloging model is proposed taking into account the classification of information proposed by the authors. An economic analysis of modern ways of storing and processing information is carried out. A method for optimizing business processes is proposed.

Keywords: information, data, data storage, optimization, data utilization, digitalization.

Проблема хранения и обработки больших массивов данных требует регулярного пересмотра, с учётом новейших научных методов и достижений. Для каждого этапа цифровизации имеется возможность выделить положительные и отрицательные аспекты, которые должны быть учтены при дальнейшем развитии последующих этапов.

Хранение информации было большой проблемой для человечества в течении тысяч лет. Всегда было проблематично содержать большое количество информации в одном месте в течении длительного периода времени, и с увеличением общего количества информации также встали ребром вопросы о её систематизации, анализе, а также об утилизации уже устаревших или признанными ложными данных. Многие компании сейчас

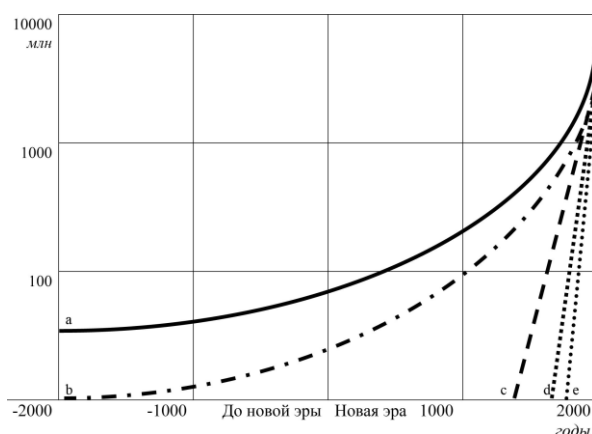
¹Студент 2 курса бакалавриата Естественнонаучного Института Самарского университета. Научный руководитель: Трусова А. Ю. кандидат физико-математических наук, доцент кафедры математики и бизнес-информатики Самарского университета.

ищут способ оптимизировать системы хранения данных: сократить расходы и использовать ресурсы носителей максимально, однако используемые методы недостаточны и поверхностны.

Хранение и накопление являются одними из основных действий, осуществляемых над информацией и главным средством обеспечения ее доступности в течение некоторого промежутка времени. В настоящее время определяющим направлением реализации этой операции является концепция базы данных, склада (хранилища) данных.

В прошлом столетии широко стала известна проблема “Информационного взрыва”, то есть резкое увеличение скорости и объёмов публикаций (объёма информации) в масштабах планеты, лавинообразное нарастание массы разнообразной информации в современном обществе – рисунок 1.

В 2002 г. человечество произвело $18 \cdot 10^{18}$ байт (18 эксабайт). За пять лет до того человечеством было произведено информации больше, чем за всю предшествующую историю. Объём информации в мире ежегодно возрастает на 30%. 92% новой информации хранится на магнитных



носителях.

Рисунок 1 - Распространение среди человечества новых функционально-поведенческих отличительных признаков в виде полезных навыков информационного взаимодействия. а — население Земли; б — грамотные люди; с — книги; d — получение радио-, телевизионной информации (кол-во приёмников); e — информационная связь через телефоны, компьютеры, интернет (кол-во телефонов, компьютеров, пользователей Интернет).

Такое стремительное увеличение массива данных влечёт за собой увеличение числа носителей, что требует гигантских расходов ресурсов. В частности, стремительно возрастает число крупных (имеющих в своём составе более 5000 серверов) дата-центров - рис. 2 - что требует создания дорогостоящей инфраструктуры. К 2025 году объём всех данных во всем мире составит 163 зеттабайт ($1.63 \cdot 10^{23}$ байт), при этом около 80% данных неструктурированы, что затрудняет их использование. Общий анализ данных не проводится, и не исключена ситуация, в которой целый дата-центр будет хранить информацию, признаваемую ложной.

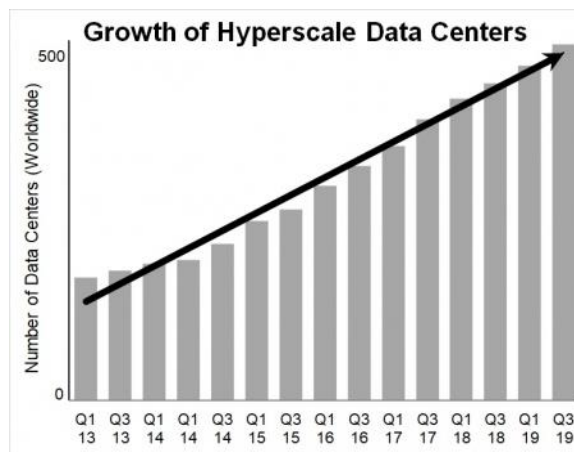


Рисунок 2 - Рост числа крупных дата-центров с 2013 по 2019 гг.

Согласно оценке IDC, расходы на аналитику и хранение больших данных составили \$150 млрд в 2017 году, тогда как в 2020 году компании потратили уже \$210 млрд, чтобы справиться с наплывом данных. Используемые сейчас способы снижения затрат направлены на максимальное использование всего объёма электронных носителей, например, путём гиперконвергенции. Этот способ объединяет все доступные носители в единую базу, и рационально распределяет данные для их дальнейшего хранения. Этот метод сокращает время простоя и позволяют существенно снизить расходы на поддержание функционирования инфраструктуры (ТСО), однако это лишь уменьшает “симптомы”, сама же “болезнь” несистематизации данных становится всё острее, и наносит серьёзный ущерб экономике.

Увеличение количества информации ведёт к тому, что сортировка - единственный способ контроля над ней. Сортировка позволит систематизировать данные, сопоставить их друг с другом, и утилизировать те, что не отвечают заданным критериям. Информация, содержащая в себе заведомо ложные данные не позволяет использовать её, то есть является вредительской для человечества и должна быть направлена на утилизацию.

Так как контроль над информацией в рамках общества может производиться только в рамках государственных структур, то при большом объёме работ рационально будет создать расширенное архивное ведомство, чьей главной задачей будет не только хранение, но и анализ, сортировка и утилизация информации.

Прежде всего необходимо чётко установить критерии оценивания:

Достоверность — лживость; Оригинальность — вторичность; Полнота — куцость; Ценность — бесполезность.

Достоверность - сопоставление информации с реальностью, оригинальность - изложение этой информации в уже известных работах, полнота - насколько широко была рассмотрена тема, ценность - сколько пользы обществу может принести эта информация.

Главным из них является достоверность информации, ведь насколько бы данные не были ценными и оригинальными, но если они не соответствуют действительности, то и потенциал реализовать не получится. Определение достоверности информации должно проходить путём перекрёстного анализа данных, используя уже подтверждённые исследования. Вторичность информации также определяется путём сравнения с уже признанными данными. Хранение вторичной и ложной информации тратит ресурсы, при этом не принося никакой пользы государству или компании, а порой приносит и прямой вред, вводя в заблуждение. В качестве примера можно привести современное положение с вирусами: в сети Интернет распространяется ложная информация о болезнях и вакцинах, что может ввести людей в заблуждение, что приводит к недооценке опасности и дальнейшим неправильным действиям. Однако данные, что некогда считались истинными, но сейчас признанными не соответствующими действительности могут быть сохранены в исторических или социокультурных целях.

Полная структуризация информации способствует научной деятельности - ведь все статьи и исследования будут собраны и каталогизированы; учебной деятельности - так как любые учебники и обучающие материалы так же будут собраны вместе; в любой отрасли производства. При этом даже без утилизации информации, одна лишь их структуризация позволит уменьшить количество места, занимаемого на электронных носителях, то есть сэкономить ресурсы.

Сама утилизация - это серьёзная проблема. Информация - это стратегический ресурс, и даже будучи признанной негодной к использованию в производстве, она может быть использована во вредительских целях.

Цифровую информацию можно уничтожить путём форматирования носителя, а данные, хранящиеся на физических носителях можно уничтожить только уничтожением носителя. Например, путём физической деструктуризации, либо химическим разложением.

Для полноценной структуризации информации необходимо создание единого соответствующего гос. аппарата в масштабах всей Земли, что на данный момент невозможно, а без соответствующей структуризации невозможна и полная утилизация данных в сети Интернет. Однако создание подобной системы в рамках одного государства обеспечит определённый контроль над информацией, позволит частично утилизировать устаревшую информацию, что в конечном итоге обеспечит экономическую, научную и информационную выгоды. Также контроль над данными обеспечит стратегическую выгоду, и не даст попасть стратегически важной информации в руки геополитических конкурентов. Более того, введение подобной системы каталогизации и анализа в рамках отдельных компаний позволит существенно уменьшить объёмы используемых носителей, снизить затраты и увеличить скорость обработки данных.

Таким образом, изучив вопросы хранения и обработки информации авторы рекомендуют обратить внимание на проблему недостаточного анализа данных при всё возрастающем темпе увеличения их количества.

Список использованных источников

- 1 Советов Б. Я. Информационные технологии: учебник для прикладного бакалавриата — 7-е изд., перераб. и доп. — Москва: Издательство Юрайт, 2019. — 327 с.
- 2 Еремин А. Л. Ноогенез и теория интеллекта. Краснодар: СовКуб, 2005. — 356 с.
- 3 Lyman P., Varian H.R. How much information. Release of the University of California. Oct.27, 2003.
- 4 David Reinsel, John Gantz, John Rydning. The Digitization of the World. From Edge to Core. An IDC White Paper – #US44413318, November 2018.
- 5 Нежданов И.Ю., Аналитическая разведка для бизнеса. Москва: Ось-89, 2008 — 336 с.