

Информационный базис и большие данные



Москва, НАПФ, 29.10.15

Проблема

Любой процесс приводит к формированию информации. Если раньше хранили только самое важное, сегодня хранят все, а интеллектуальные системы круглосуточно ищут в этом связи. В результате, когда нужно узнать несколько основных параметров, тратится масса времени на обработку всего массива. Растут риски ошибок при обработке запросов, связанные с операционными рисками и нарушением принципа Оккама*, в результате информационные риски увеличиваются с расширением информационного пространства.

****Бритва Оккама** (иногда «лезвие Оккама») — методологический принцип, получивший название от имени английского монаха-францисканца, философа-номиналиста Уильяма Оккама (Ockham, Oskam, Ossam; ок. 1285—1349). В кратком виде он гласит: «Не следует множить сущее без необходимости» (либо «Не следует привлекать новые сущности без крайней на то необходимости»). Этот принцип формирует базис методологического редукционизма, также называемый принципом бережливости, или законом экономии (лат. Lex parsimoniae).*

Порой принцип выражается в словах «То, что можно объяснить посредством меньшего, не следует выражать посредством большего» (лат. Frustra fit per plura quod potest fieri per pauciora).

Большие данные

Большие данные (англ. *big data*) в информационных технологиях — серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х годов, альтернативных традиционным системам управления базами данных и решениям класса [Business Intelligence](#). В данную серию включают средства массово-параллельной обработки неопределённо структурированных данных, прежде всего, решениями категории NoSQL, алгоритмами MapReduce, программными каркасами и библиотеками проекта Hadoop. В качестве определяющих характеристик для больших данных отмечают «*mpu V*»: объём (англ. *volume*, в смысле величины физического объёма), скорость (англ. *velocity* в смыслах как скорости прироста, так и необходимости высокоскоростной обработки и получения результатов), многообразие (англ. *variety*, в смысле возможности одновременной обработки различных типов структурированных и полуструктурированных данных)

Информационное общество — общество, в котором большинство работающих занято производством, хранением, переработкой и реализацией [информации](#), особенно высшей её формы — [знаний](#). Для этой стадии развития общества и экономики характерно:

увеличение роли информации, знаний и информационных технологий в жизни [общества](#);

возрастание числа людей, занятых информационными технологиями, коммуникациями и производством информационных продуктов и [услуг](#), рост их доли в [валовом внутреннем продукте](#);

нарастающая [информатизация](#) общества с использованием телефонии, радио, телевидения, сети Интернет, а также традиционных и электронных СМИ;

создание глобального [информационного пространства](#), обеспечивающего:

эффективное информационное взаимодействие людей;

их доступ к мировым информационным ресурсам;

удовлетворение их потребностей в информационных продуктах и услугах.

развитие [электронной демократии](#), [информационной экономики](#), [электронного государства](#), [электронного правительства](#), цифровых рынков, электронных [социальных](#) и хозяйствующих сетей.

Модель развития общества – волны поколений

Война государств – это не война правительств, армий, промышленности - это война территорий. Чем богаче и могущественнее территория, тем сложнее ее завоевать, имея даже заметное военное преимущество. История России многократно это подтверждала. Правители считают, что принимая решения, они оказывают влияние на текущие события, но реакция на любое возмущение в социуме происходит не мгновенно. Это легко промоделировать на примере передачи сигнала по гибкой цепи, также можно сравнить с движением в автомобильной пробке или привести другие примеры. Скорее принимаемые решения совершаются в угоду событиям. Реакция на иные принимаемые сегодня решения, происходит в будущем, и часто в далеком будущем. Поэтому политики, когда сегодня идут против сложившихся тенденций, исправляют ошибки своих предшественников, но в результате создают проблемы своим приемникам. Происходит этакая игра волн. И все хорошо, пока нет ресурсных ограничений. Когда они наступают, сложившаяся модель межпоколенческого переноса проблем нарушается, и правила внутри строя изменяются. Поэтому чем богаче территория государства, тем дольше государство существует.

Монтескье в своем знаменитом сочинении “О духе законов” (1748 год) писал о том, что любая человеческая жизнь и жизнь любого народа - просто отражение и климатических, и географических, и прочих условий обитания.

Эволюция сознания и общества

Важным этапом необратимой эволюции биосферы Вернадский считал её переход в стадию ноосферы. Энергия живого вещества.

Никита Моисеев. Судьба человечества. Путь разума Sustainable development.

Пассиона́рная теория этноге́неза (теория пассионарности и этногенеза) [Льва Гумилёва](#) описывает исторический процесс как взаимодействие развивающихся [этносов](#) с вмещающим [ландшафтом](#) и другими этносами. Теория Гумилева определяет и описывает понятия [этноса](#) (точнее, несколько видов [этнических систем](#)), вводит понятие **пассионарности**, описывает типичные процессы [этногенеза](#) и взаимодействия этносов.

Пассионарность (от [фран.](#) passioner - увлекаться, разжигать страсть) — избыток некой «биохимической энергии» живого вещества, порождающий жертвенность, часто ради высоких целей. Пассионарность — это непреодолимое внутреннее стремление к деятельности, направленной на изменение своей жизни, окружающей обстановки, [статуса-кво](#). Деятельность эта представляется пассионарной особи ценнее даже собственной жизни, а тем более жизни, счастья современников и соплеменников. Она не имеет отношения к этике, одинаково легко порождает подвиги и преступления, творчество и разрушение, благо и зло, исключая только равнодушие; она не делает человека героем, ведущим толпу, ибо большинство пассионариев находятся в составе толпы, определяя её потенциальность в ту или иную эпоху развития этноса.

Понимание пассионарности у Л. Н. Гумилёва несколько различается в разных работах. По сути, это социально-историческое явление, характеризующееся появлением в ограниченном ареале большого числа людей со специфической активностью (пассионариев). Мера пассионарности — удельный вес этих пассионариев в социуме.

Часто под пассиона́рностью понимают [наследуемую](#) характеристику, определяющую способность индивида (и группы индивидов) к сверхусилиям, сверхнапряжению. Сам основоположник пассионарной теории этногенеза приводил примеры, но о природе и сущности пассионарности не говорил. Гумилёв лишь указал, что пассионарность индивида по сути является психологической переменной и зависит, вероятно, от мутаций, вызываемых космическим излучением «Каков характер этого излучения? Здесь мы можем строить только гипотезы. Их две. Первая — о возможной связи пассионарных толчков с многолетней вариацией солнечной активности, обнаруженной Д. Эдди. Вторая гипотеза — о возможной связи со вспышками сверхновых»(Л. Н. Гумилёв «Конец и вновь начало»). В современных исследованиях личностные проявления пассионарного поведения получили название [Вигоросность](#).



Знаменитый физик Илья Пригожин сказал про Интернет примерно следующее: возможно, это окажется самым выдающимся достижением человечества, а возможно, мы все превратимся в колонию слепых муравьев. Все зависит от того, как человечество будет эту штуку использовать. Пока человечество явно тупеет. Растет число жертв интернета (к счастью, растет и число исследований влияния этой штуки на человечество). Совсем недавно было доказано, что более трех сэлфи в день – это психическое заболевание. Что люди, сидящие в соцсетях, способны сосредотачиваться на период, меньший, чем золотая рыбка. Это ужасно. Но остается надежда.

Информационная революция

Мир стоит на пороге новой технологической революции, которую несет с собой новая наука — интеллектуальный анализ данных.

Если взять общий объем данных, находящихся в мировом совокупном storage (совокупное компьютерное пространство для хранения документов. — «Эксперт»), **то 90 процентов из них появились за последние два года. Фантастическая цифра, но это правда.** Оказалось, что из этих данных можно много что извлечь, и эта технологическая возможность породила новые математические задачи.

Теперь, когда мы устроили в своих сетях без пяти минут информационный хаос, давайте попробуем переработать его во что-нибудь полезное, — предлагают компьютерные лингвисты. В экспоненциально растущей свалке слов, которую представляет собой интернет, закопаны информационные сокровища. Они достаются тому, кто умеет искать. (<http://www.nkj.ru/archive/articles/25308/> Наука и жизнь, Читать 2.0: компьютеры учатся добывать из текста смысл)

Риски:

- **информационная перегрузка**
- **прокрастинация**
- **неправильная интерпретация**
- **сингулярность**

Информационная перегрузка

На протяжении всей истории человечества информация была дефицитом и ценностью. Однако последние десятилетия заставляют менять отношение к ней. Данные накапливаются эксабайтами (10^{18} байт), охватывая все: от частоты пульса пользователей «умных» браслетов до снимков удаленных галактик, от расхода воды в каждой квартире до параметров ядерных реакторов. Согласно оценкам компании Cisco, к концу 2014 года только ежемесячный мобильный трафик данных в мире составил 2,5 эксабайт, к 2019-му он достигнет 24,3 эксабайт в месяц, а число подключенных мобильных устройств превысит 10 млрд единиц. На Facebook ежечасно загружается более 10 млн фотографий. На YouTube каждую секунду загружается более часа видео. По прогнозам IDC, к 2020 году общий объем цифровых данных достигнет 40 зеттабайт. Для понимания масштабов: если записать 40 зеттабайт ($40 \cdot 10^{21}$ байт) данных на самые емкие современные диски Blue-ray, суммарный вес дисков без упаковки будет равен весу 424 авианосцев. При этом используется лишь менее 3% из 23% потенциально полезных данных.

В 2011 году американцы потребляли в пять раз больше информации в день, чем в 1986 — по объёму она была сопоставима с 174 газетами. В течение своей жизни каждый человек обрабатывает 34 гигабайта данных или порядка 100 тысяч слов ежедневно.

21 274 телевизионных станции производит 85 тысяч часов оригинальных программ в день, из которых люди в среднем смотрят по 5 часов в сутки, что составляет порядка 20 гигабайт фото- и видеоданных. При этом на YouTube каждый час заливается более 6 тысяч часов роликов. Всего люди создали уже порядка 300 эксабайтов данных (три в двадцатой степени, шесть триллионов романов «Война и мир»).

*Наименование и обозначение единицы количества информации «байт» (1 байт = 8 **бит**) применяются с двоичными приставками «Кило», «Мега», «Гига», которые соответствуют множителям « 2^{10} », « 2^{20} » и « 2^{30} » (1 Кбайт = 1024 байт, 1 Мбайт = 1024 Кбайт, 1 Гбайт = 1024 Мбайт). Данные приставки пишутся с большой буквы. Допускается применение международного обозначения единицы информации с приставками «К» «М» «Г», рекомендованного Международным стандартом Международной электротехнической комиссии МЭК 60027-2 (KB, MB, GB, Kbyte, Mbyte, Gbyte).*

Прокрастинация

Как не сбиться с пути и закончить большой проект? Новый пост известного блоггера и писателя Криса Бейли, который уже несколько лет изучает приемы производительной жизни.

Если бы я не отключался от сети во время написания книги, я практически уверен, что писал бы ее до сих пор. Тим Пичил, который исследует мотивацию и прокрастинацию уже 20 лет, провел интереснейшее исследование производительности. Он изучил, сколько времени средний человек тратит на прокрастинацию, когда подключен к интернету, и открылось нечто удивительное: средний человек 47% времени, проведенного в онлайн, прокрастинирует. И это было еще до взлета социальных сетей!

Прокрастинáция (от [англ.](#) *procrastination* — задержка, откладывание; от [лат.](#) *crastinus* — завтра и [лат.](#) *pro* — на) — в [психологии](#) — склонность к постоянному откладыванию важных и срочных дел, приводящая к жизненным проблемам и болезненным психологическим эффектам. Прокрастинация проявляется в том, что человек, осознавая необходимость выполнения вполне конкретных важных дел (например, своих должностных обязанностей), игнорирует эту необходимость и отвлекает своё внимание на бытовые мелочи или развлечения.

Прокрастинация отличается от [лени](#) тем, что в случае лени субъект не хочет ничего делать и не беспокоится по этому поводу, а в состоянии прокрастинации он осознаёт важность и срочность работы, но не делает её, возможно, находя те или иные самооправдания. От отдыха прокрастинацию отличает то, что при отдыхе человек восполняет запасы энергии, а при прокрастинации — теряет¹.

Неправильная интерпретация

По данным Gartner, к 2018 году половина всех случаев нарушения деловой этики будет связана с неправильным применением анализа данных, в первую очередь, неспособностью извлекать из них выгоду для бизнеса. Риски усиливаются за счет большого объема данных, их разнообразия и изощренности современных аналитических методов.

Gartner опубликовал рекомендации, которые помогут предприятиям не наступать на грабли при работе с большими данными. По мнению аналитиков, примерно 50% всех случаев нарушения деловой этики к 2018 году будут вызваны неспособностью правильно обращаться с огромными объемами информации и инструментами для их обработки. Речь идет о неэффективном использовании ресурсов, возможной потере репутации, об ограничении деятельности бизнеса и даже о правовых санкциях.

Нужны ли вам средства обработки больших данных?

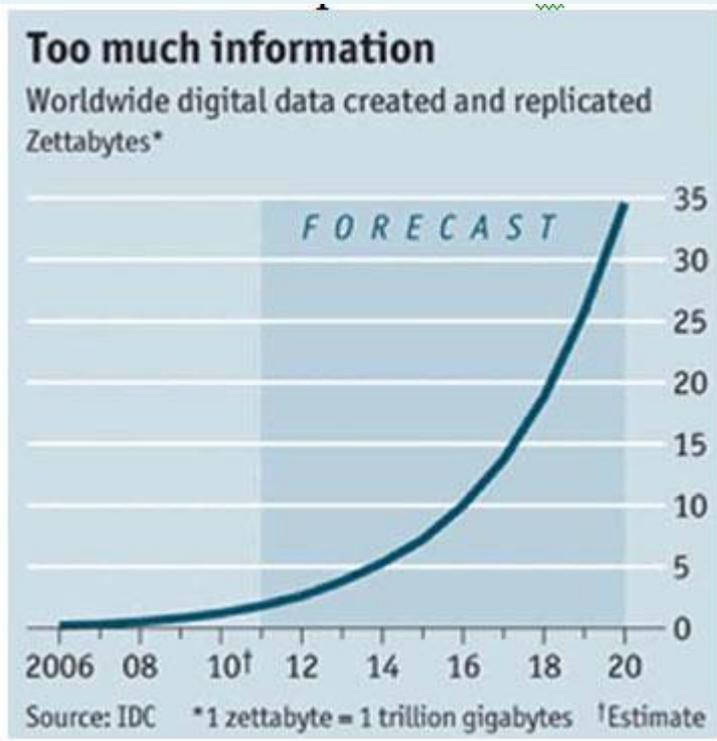
В первую очередь, специалисты Gartner советуют пошагово соотносить аналитические выводы с тем эффектом, который получает предприятие по результатам принятых на их основе бизнес-решений. Некоторые организации по ряду причин не способны применить в своей деятельности результаты исследований. Кроме того, принятие решения не всегда требует инвестиций в инженерные разработки и сложные аналитические проекты.

Сингулярность

Новая научно технологическая революция — гипотетический момент (Сингулярность), по прошествии которого машинное мышление перегонит биологическое, и технический прогресс станет настолько быстрым и сложным, что окажется недоступным человеческому пониманию.

Machine Intelligence Research Institute (MIRI; ранее — **Singularity Institute** и **Singularity Institute for Artificial Intelligence**) — некоммерческая организация, основной целью которой является создание безопасного искусственного интеллекта, а также изучение потенциальных опасностей и возможностей, которые могут появиться при создании ИИ. Организация поддерживает идеи, первоначально выдвинутые Ирвингом Гудом и Вернором Винджем, касательно «интеллектуального взрыва» или сингулярности, и идеи Елиезера Юдковски о создании дружественного ИИ. Юдковски исследует в Институте сингулярности в США проблемы глобального риска, которые может создать будущий сверхчеловеческий ИИ, если его не запрограммировать на дружественность к человеку

Прогноз объема данных, накопленных человечеством*



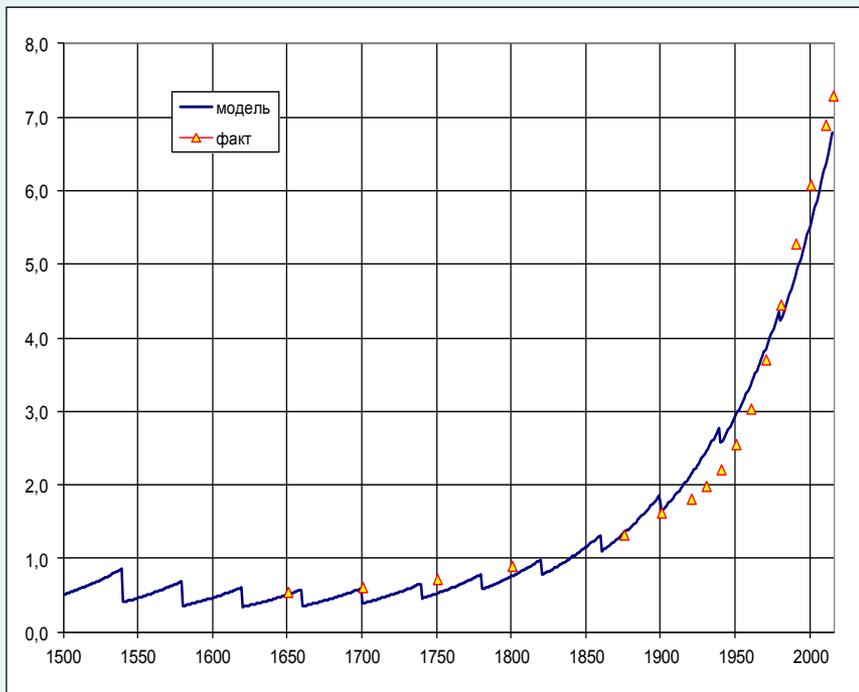
В гуманитарных науках нет и быть не может законов типа законов Ньютона и почему для изучения мира человеческих страстей и духовного мира человека нельзя опереться на четкие законы, а надо принять ЛОГИКУ развития событий. И учиться следовать этой ЛОГИКЕ.

Гуманитарные науки изучают прежде всего мышление человека, особенности его духовного мира, и, познавая его логику, люди тем самым совершенствуют свое мышление, т.е. меняют самого человека, идеалы, к которым он стремится. Следовательно, и его действия**.

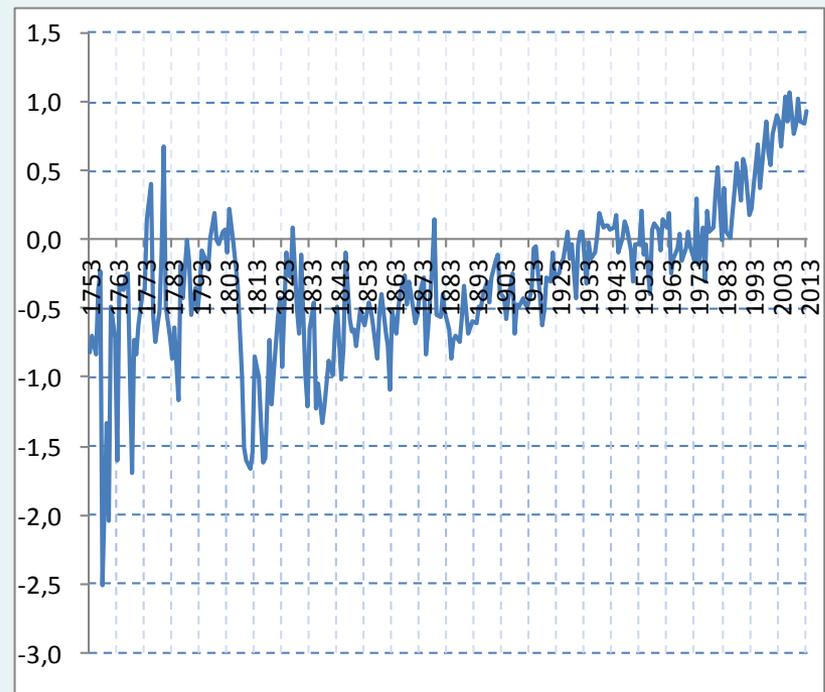
*<http://spkurdyumov.ru/economy/informacionnye-texnologii-i-innovacii/>

**Н. Мусеев Судьба цивилизации

Численность населения Земли



Изменение температуры земной поверхности

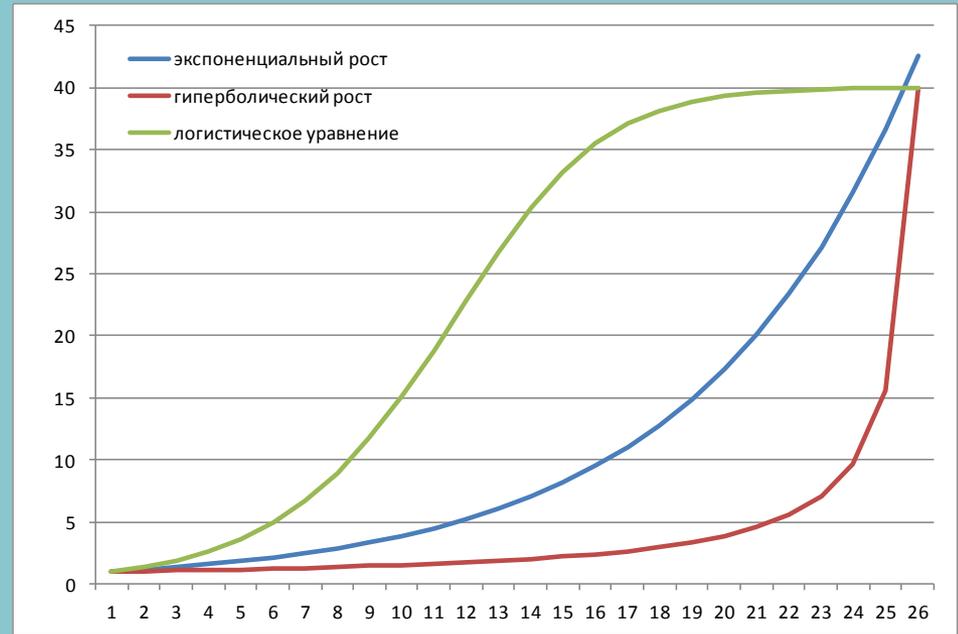


Пределы роста

$$\frac{dP}{dt} = \alpha * P \quad \text{модель экспоненциального роста)}$$

$$\frac{dP}{dt} = \alpha * P^2 \quad \text{(модель гиперболического роста)}$$

$$\frac{dP}{dt} = (\alpha - \beta * P) * P \quad \text{(логистическое уравнение)}$$



$$\frac{dX_1}{dt} = (a - b * X_2) * X_1; \quad \text{(модель Вольтерра-Лотки)}$$

$$\frac{dX_2}{dt} = (-c + d * X_1) * X_2.$$

$$\frac{dX(t)}{dt} = -a * X(t) - b * Y(t) + F_1(t);$$

модель Ланчестера

$$\frac{dY(t)}{dt} = -c * X(t) - d * Y(t) + F_2(t).$$

Естественные ограничения роста: эпидемии

год	число умерших	болезнь	страна
1625	35 000	чума	Великобритания
1656	60 000	чума	Италия
1665	100 000	чума	Лондонская чума
1672	400 000	чума	Италия
1771		чума	Москва
1792	800 000	чума	Египет
1793	5 000	желтая лихорадка	США
1812		сыпной тиф*	Россия
1855	12 000 000	чума	Китай, Индия
1896	3 000 000	чума	Индия
1910	1 000 000	чума	Индия
1916		полиомелит	США
1918	75 000 000	Испанка**	Европа
1921	1 000 000	чума	Индия
1957	2 000 000	азиатский грипп	
1968	33 800	гонконгский грипп	США
1980-2006	25 000 000	СПИД	

**Во время похода Наполеона в Россию в 1812 г. французская армия потеряла от сыпного тифа 1/3 своих солдат, а армия Кутузова половину войска*

***Испанка была вероятней всего, самой массовой за всю историю человечества. В 1918—1919 годах (18 месяцев) во всем мире от испанки умерло приблизительно 50-100 млн человек или 2,7-5,3 % населения Земли. Было заражено около 550 млн человек, или 29,5% населения планеты. Эпидемия началась в последние месяцы Первой мировой войны и быстро затмила это крупнейшее кровопролитие по масштабу жертв. В мае 1918 года в Испании было заражено 8млн людей или 39% её населения*

Естественные ограничения роста: катастрофы

Сверхизвержение вулкана Тамбора на острове Сумбава в Индонезии в 1815 году имело разрушительные последствия. Тогда в результате этого стихийного бедствия погибло свыше 70 тысяч человек (самое большое число жертв вулкана за всю историю цивилизованного человечества, хотя непосредственно взрыв уничтожил только 12 тысяч человек, остальные погибли от цунами, ядовитых газов и от падающих камней), несколько островов архипелага стали полностью негодными для проживания.

Кроме того, это извержение породило глобальные климатические аномалии, включая такой феномен, как "вулканическая зима". Так, 1816 год стал известен как "год без лета" из-за чрезвычайно низких температур, которые установились в Европе и Северной Америке. Этот необычный холод привел к катастрофическому неурожаю, из-за чего весной 1817 цены на зерно выросли в десять раз, а среди населения разразился голод. В результате множество европейцев эмигрировали в Америку

Однако даже эта катастрофа меркнет перед тем, что устроил людям **вулкан Тоба** с индонезийского острова Суматра 74 тысячи лет тому назад. Тогда при сверхизвержении из земных недр было выброшено больше тысячи кубических километров магмы (для сравнения — этого количества вполне достаточно для того, чтобы покрыть территорию всей России восьмисантиметровым слоем лавы). После взрыва из-под земли со сверхзвуковой скоростью вырвался столб раскаленного газа и пепла, который почти мгновенно достиг края стратосферы — отметки в 50 километров. За трое суток на поверхность излилось более 2800 кубических километров магмы: кое-где толща застывшей лавы достигала десятков метров.

От Мальтуса

В 1798 году Мальтус опубликовал свою книгу *Essay on the Principle of Population* («Очерк о законе народонаселения»).

Три основных тезиса «Очерка»:

Из-за биологической способности человека к продолжению рода, его физические способности используются для увеличения своих продовольственных ресурсов.

Народонаселение строго ограничено средствами существования.

Рост народонаселения может быть остановлен лишь встречными причинами, которые сводятся к нравственному воздержанию, деструктивному действию пороков или несчастьям (войны, эпидемии, голод).

Также Мальтус приходит к выводу, что народонаселение растёт в геометрической прогрессии, а средства существования — в арифметической.

В 1850-м году Французская Академия составила прогноз развития города Парижа на 100 лет - к 1950-му году. Этот прогноз сбился во многом за одним исключением: французские академики считали, что главными проблемами для Парижа будут транспортные, потому что увеличится население, и, следовательно, главной транспортной проблемой будет уборка конского навоза - они не представляли, что будут другие виды транспорта.

до Римского клуба

Во всей истории человечества не было периода, подобного XX веку: численность населения выросла в четыре раза, глобальный ВВП — в 24 раза, а потребление всех материалов увеличилось в восемь раз. При этом потребление биомассы выросло в 3,6 раза, металлических руд — в 27, а строительных материалов — в 37 раз. Рост потребления материалов составлял от 1 до 4% в год.

В начале 70-х годов по предложению Клуба [Джей Форрестер](#) применил разработанную им методику моделирования на ЭВМ к мировой проблематике. Результаты исследования были опубликованы в книге «Мировая динамика» (1971), в ней говорилось, что дальнейшее развитие человечества на физически ограниченной планете Земля приведет к экологической катастрофе в 20-х годах следующего столетия.

Сегодня появились возможности резко повысить эффективность использования материалов благодаря их армированию SWCNT (*одностенных углеродных нанотрубок*) . Добавление 0,2% SWCNT на единицу массы алюминия позволяет повысить предел прочности при растяжении на 150% или сократить потребность в алюминии в два с половиной раза. При легировании в пропорции 0,1% предел прочности при растяжении повышается на 100%. Для меди добавление 0,1% SWCNT на единицу массы меди позволяет повысить предел прочности при растяжении на 100%, или сократить потребность в меди вдвое. Для бетона добавление всего лишь 0,001% SWCNT позволяет повысить упругость на сжатие на 70%, что должно привести к снижению его потребления в 1,7 раза на единицу полезной функции. Аналогичные результаты достигаются и при легировании других материалов SWCNT. То есть добавки SWCNT к базовому материалу ведут к сокращению потребности в нем за счет повышения его потребительских свойств (прочности, упругости и др.) и, следовательно, к снижению энергопотребления и, соответственно, выбросов парниковых газов.

В результате кумулятивное снижение выбросов за 2015–2100 годы составит 331 млрд тонн CO₂, что кратно шести объемам всех антропогенных выбросов за 2014 год и восьми годовым объемам выбросов ПГ от сжигания топлива и от промышленных процессов. Для этого необходимо увеличить производство и использование SWCNT до 167 тыс. тонн в 2035 году, до 289 тыс. тонн в 2050-м и до 429 тыс. тонн в 2100-м.

Эволюция

Первая промышленная революция, обеспечившая переход от ручного труда к машинному, растянулась на многие десятилетия. Принято связывать ее с изобретением парового двигателя Ньюкмана в XVII в., но процесс перехода от мануфактур к фабрикам продолжался в развитых странах в течение XVIII–XIX вв (двигатель Уатта). Промышленная революция сильно повлияла не только на развитие науки и техники, но и на изменение структуры общества, урбанизацию, появление новых специальностей.

Вторая промышленная революция была связана с электрификацией и организацией конвейерного производства в XX в. сначала автомобилей, а потом и большинства других товаров. Резко повысилась производительность труда, изменились подходы к управлению предприятиями.

В начале XXI в. стало появляться все больше публикаций на тему **третьей промышленной революции**. Она базировалась на отказе от использования полезных ископаемых, переходе к возобновляемым источникам энергии в сочетании с внедрением компьютеров в производство, автоматизацией и переходом к цифровому аддитивному производству, который журнал

Но, несмотря на то что реалии третьей революции еще далеко не распространились по миру, назрела новая «революционная ситуация» – немецкая концепция «Индустрия 4.0». В зависимость от ее реализации был поставлен сам факт существования в будущем промышленности Германии на фоне глобального переноса производств в Азию и другие развивающиеся страны. Ключевой двигатель «Индустрии 4.0» – усиленная интеграция «киберфизических систем», или CPS, в заводские процессы. Производственные мощности будут взаимодействовать с производимыми товарами и адаптироваться при необходимости под новые потребности потребителей. При этом целые этапы производства будут происходить без участия человека. **Это производственная часть интернета вещей, который стремительно проникает в нашу жизнь**

Способы защиты от информационной перегрузки

- Фильтрация информации
- Информационный базис



Фильтрация информации

Человеческий мозг способен справляться с большим объёмом информации, но его работа имеет определённую цену. У людей возникают проблемы, когда появляется необходимость отделить важные данные от второстепенных, этот процесс утомляет. Нейроны, требующие для жизни кислород и глюкозу, быстро истощаются.

Каждое обновление статуса, которое вы читаете на Фейсбуке, каждый твит, каждое текстовое сообщение, которое вы получаете от друга, борется за ресурсы вашего мозга с более важными вопросами: надо ли обратить свои сбережения в акции, где вы оставили свой паспорт или как быстро прийти к взаимопониманию с другом, с которым вы поссорились.

Дэниел Левитин, нейробиолог

Человеческий мозг проигрывает войну против растущего объема информации, который ему приходится обрабатывать. Нейробиологи [утверждают](#), что мозг человека сильно устарел, а внимание постепенно становится самым нашим ценным ресурсом. И редким тоже — чтение книг превращается в задачу, [требующую](#) практически непосильной в век Твиттера и BuzzFeed концентрации.

Но природа и учёные подарили людям разнообразные «допинги для мозга». И в информационный век, предполагающий обработку невероятного количества данных, многим перестаёт хватать кофеина.

Другая интересная тактика по борьбе с перегрузкой в информационном поле состоит не в адаптации организма к нему, а наоборот — в тщательной фильтрации поступающих данных. Жёстко регулировать темы и вещи, достойные внимания, откинув «мусор» и негативные факторы: может быть, как раз это секрет адаптации к новому информационному пространству — позволить ему в большей степени проходить мимо себя.

Некоторые законы

Первый закон Паркинсона гласит, что общий объём работы всегда будет увеличиваться, чтобы заполнить собой всё отпущенное на работу время.

Закон телефона. То, насколько эффективен телефонный разговор, обратно пропорционально тому, сколько на него было затрачено времени.

Закон информации (касаемо компьютеров). Количество данных увеличивается, чтобы в итоге на носителе не осталось свободного места, а увеличение памяти и самих носителей — это предпосылка к появлению новых технологий, для которых нужно ещё больше памяти и носителей

Некоторые примеры

Дорожка к Кармана была рассчитана в начале 50-х годов на арифмометре за год. Сейчас на это требуется меньше секунды, но качественно результат не меняется.

Информационный базис

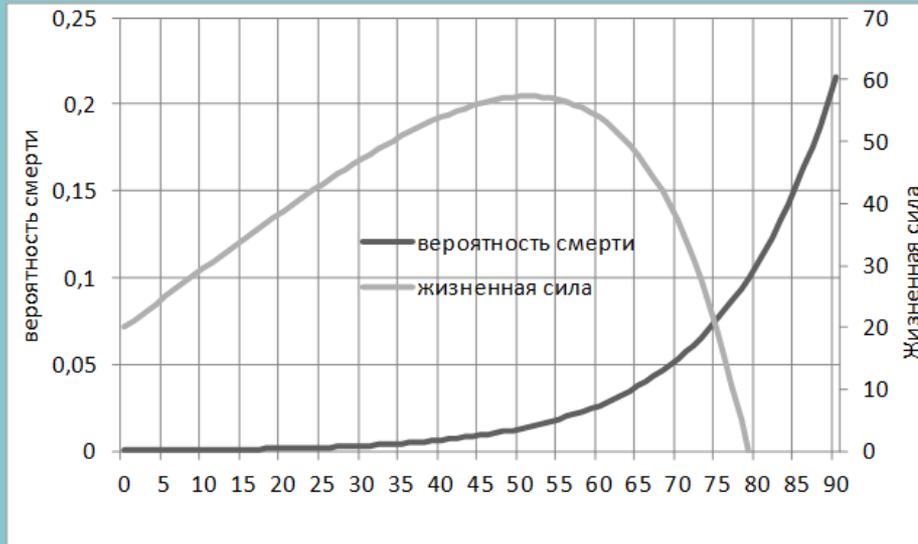
Разделение информации по степени значимости является актуальной задачей. Для ее решения целесообразно **создать систему ключевых (базисных) показателей.**

Например, показатели отчетности сопоставляются квартал к кварталу, год к году. **Но для целей анализа тенденций необходимо видеть не только текущие значения, но и отслеживать тенденции.** Сегодня у информационных систем существуют трудности, во-первых, увеличивается время обработки запросов, во-вторых при большом наборе параметров анализ тенденций оборачивается серьезной проблемой.

Для решения описанных проблем предлагается усовершенствовать информационную систему и сегментировать информацию, дополнив ее надстройкой в которой сопровождаются (рассматриваются) только базисные показатели, определяемый на основании опыта или специального исследования. Затем формируется регламент сопровождения базисных показателей. Какие преимущества предоставляет такое решение? Так как размерность массива невелика и не ограничивает скорость обработки, результаты при соблюдении прав доступа можно публиковать в сети Интернет, что открывает доступ к данным в любом месте в любое время, независимо ушел домой штатный программист или руководитель. Возможность сопоставления исторических данных позволяет менее болезненно реагировать на текущие события. Например, легко получить ответ на вопрос встречалось ли ранее изменение подобное наблюдаемому сегодня.

Отличие между информационным базисом и большими данными аналогично разнице между стратегическим и операционным управлением.

Информационный базис не ограничивается только архитектурой баз данных – возможно это стиль жизни.



Многие процессы цикличны. Даже пассионарность Гумилева – не что иное, как дифференциальное уравнение. У одних процессов длина волны большая, у других малая, но рост описывается д.у. Рост прекращается, когда встречается сопротивление среды, потом происходит смена модели и новый дифур. Наверное, число растущих процессов инвариантно. Поэтому нужны новые идеи и приложения в областях, которые развиваются. Периоды роста и спада в ряде отраслей соизмеримы с периодами активности человеческой жизни. В зависимости от местоположения человека на кривой смертей, нужно решать в каких нишах следует работать. Работа в развивающихся нишах приводит к росту жизненной силы, поэтому переход в активные ниши в старшем возрасте выглядит правильным, поскольку увеличивает продолжительность активной жизни.

«До пятидесятилетнего возраста я, как и большинство людей, не придавал особого значения физической культуре, — пишет А. А. Микулин. — Однако, когда я тяжело заболел и оказался в больнице, у меня появилось желание и время разобраться в сложном устройстве человеческого организма. Изучив много книг на эту тему, я и разработал свою систему, исправленную и дополненную критикой и опытной проверкой моих друзей. Эта система позволила мне преодолеть болезни, перестать стареть и сохранить работоспособность на протяжении многих лет, включая и настоящее время». Микулин А.А Долголетие должно быть активным.

Примеры построения информационного базиса

- **База данных документов**
- **Статистические базы данных**
- **DataPoint**

База данных документов

База Данных Документов (БДД) создается с целью сбора, централизации и систематизации основополагающих документов, упрощения к ним доступа, автоматического резервного копирования, анализа и построения структурированного архива.

БДД представляет, информационную оболочку в формате HTML, позволяющую дальнейшее развитие и дополнение, упрощающую доступ к документам, облегчающую их классификацию и поиск.

БДД позволяет осуществлять контроль деятельности как отдельных сотрудников, так и организации в целом.

Меню в формате HTML, позволяющее легко вносить изменения и дополнения в структуру документов

Например, меню из 3-х уровней, содержащее 10 пунктов на каждом уровне, позволяет выбрать нужный документ из 1000 возможных только за 3 клика мыши. Дополнительный клик мыши позволит расширить поиск до 10 000 документов

Наличие поисковой системы позволит работать с системой пользователю, не знакомому со структурой документов и расширить возможности системы.

Примеры построения базы данных документов

The image displays a complex web application interface with multiple overlapping windows. The primary window shows a table of HIV/AIDS statistics in Russia from 1987 to 2002. A secondary window displays a document titled 'Положение детей в Российской Федерации' (Situation of Children in the Russian Federation), which includes text about social protection and child welfare. A third window shows a list of documents for download.

Year (date as of December 31)	New Cases Registered During the Year	Cumulative Since
1987	23	
1988	47	
1989	268	
1990	130	
1991	82	
1992	86	
1993	108	
1994	158	
1995	189	
1996	1,433	
1997	3,858	
1998	3,709	
1999	19,851	
2000	59,281	
2001	88,494	
2002	49,917	

Положение детей в Российской Федерации

Международный Банк Реконструкции и Развития (Всемирный Банк), членом которого Российская Федерация является с 1992 года, расширяет сотрудничество...

Уровни благосостояния детей, их социальное и физическое благополучие зависят от устойчивости семьи как основной ячейки общества, от результативности государственной политики в этой области и от объема инвестиций в детей. Инвестиции в детей представляют собой вложение в будущий человеческий капитал, который является ключевым фактором личного благосостояния, экономического роста и социально-экономического развития в целом. Такие инвестиции обеспечивают более высокую производительность труда и за счет этого обеспечивают возможности роста эффективности функционирования экономики. Совершено осознанные инвестиции в детей позволяют в будущем получить образование на государственном социальном расходовании, так как они уменьшают риск разрушающего автоблагосостояния поведения, сокращают потребность в будущих социальных расходах. Инвестиции в детей способствуют большей социальной справедливости, позволяют детям из малообеспеченных семей преодолеть исключение из общества и создать возможности для вертикальной социальной мобильности.

В России правильно организованные и адекватные инвестиции, направленные на обеспечение благополучия детей имеют особое значение в связи с двумя обстоятельствами. Во-первых, реорганизация экономического и социального устройства общества на рыночных принципах в 1990-х годах резко увеличила риск бедности и создала дополнительные риски для детей. Для снижения этих рисков должна быть укреплена соответствующая система социальной защиты семьи и детей. Во-вторых, актуальность этой работы значительно возрастает в связи со сложной демографической ситуацией и резко сократившейся за 1990-е годы численностью детского населения.

В течение 1990-х годов Правительством РФ предприняты важные шаги по развитию системы социальной защиты детей. В основном создана современная законодательная и нормативно-правовая база, система органов управления, действует сеть учреждений и организаций, осуществляющих помощь семьям и детям, оказавшимся в трудной жизненной ситуации. Вместе с тем озабоченность вызывает рост числа детей, регистрируемых как лишенных родительского попечения и рост численности детей, помещенных в интернатные учреждения: по состоянию на 1 января 2002 года в интернатных учреждениях в РФ находилось более 683 тыс. детей.

В обществе осознается необходимость перехода к более эффективной системе социальной защиты детей, что предполагает: (i) укрепление системы социальных услуг, обеспечивающую своевременную диагностику и профилактику семейного неблагополучия; (ii) переход на преимущественно семейное устройство детей, оставшихся без попечения родителей и (iii) перепрофилирование и повальное свертывание системы интернатных учреждений. Переход к новой системе позволит обеспечить сокращение числа детей, оставшихся без попечения родителей, снижение притока детей в интернатные учреждения.

На данном CD – диске представлены материалы по проблеме социальной защиты детей, разработанные специалистами и консультантами Всемирного Банка в сотрудничестве с Министерством и ведомствами РФ, при активной участии администрации ряда регионов Российской Федерации, сотрудников учреждений социальной защиты, образования и здравоохранения. Эти материалы включают краткое резюме доклада «Положение детей в Российской Федерации. Итоги социальной политики 1990-х годов и перспективы», статистические данные и материалы, интентивную новеллу, бюджетных и экономических затрат содержания и реформирования системы социальной защиты детей. Представлены материалы региональных исследований систем социальной защиты детей, проведенные в 2002-2004 гг. в Костромской, Ростовской областях, Алтайском крае и Республике Бурятия. Приведены ссылки на информационные источники.

ГОТОВО

Eng

ВСЕМИРНЫЙ БАНК

Статистические базы данных

В сети Интернет опубликован новый демографический ресурс — <http://www.oaoospos.ru/demography/>, содержащий данные о половозрастном составе населения и показатели таблиц дожития для 37 стран, включая Россию в течение длительного исторического периода. Самый большой период у Швеции (данные с 1751 года). Данные по России представлены в региональном разрезе за период 1990 – 2014 г.

Ресурс разработан с целью систематизации массива данных, широко использующихся в различных областях, включая социальную политику, пенсионную индустрию и т. д., и содержит две базы данных: «Россия и мир» и «Регионы России». Раздел с данными по России доступен по ссылке: <http://oaoospos.ru/demography/index.php?type=russia&action=general&lang=ru>

Собранный массив данных позволяет получить информацию, связанную с основными демографическими показателями (продолжительность жизни, доли населения и т. д.) и их тенденциями, а также может использоваться при составлении аналитических материалов, отчетов, презентаций и т. д.

Интерфейс ресурса разработчики старались сделать максимально простым, понятным и удобным.

С помощью данного ресурса можно проводить сравнительный анализ показателей по различным регионам России по таким характеристикам:

- ожидаемая продолжительность жизни с любого возраста;
- нагрузка пожилыми;
- численность населения в абсолютном и долевым выражении в заданной возрастной группе;
- численность умерших в заданной возрастной группе.

Статистические базы данных. Версия 1998 года – рынок РТС. Версия 2015 года – демографический ресурс.

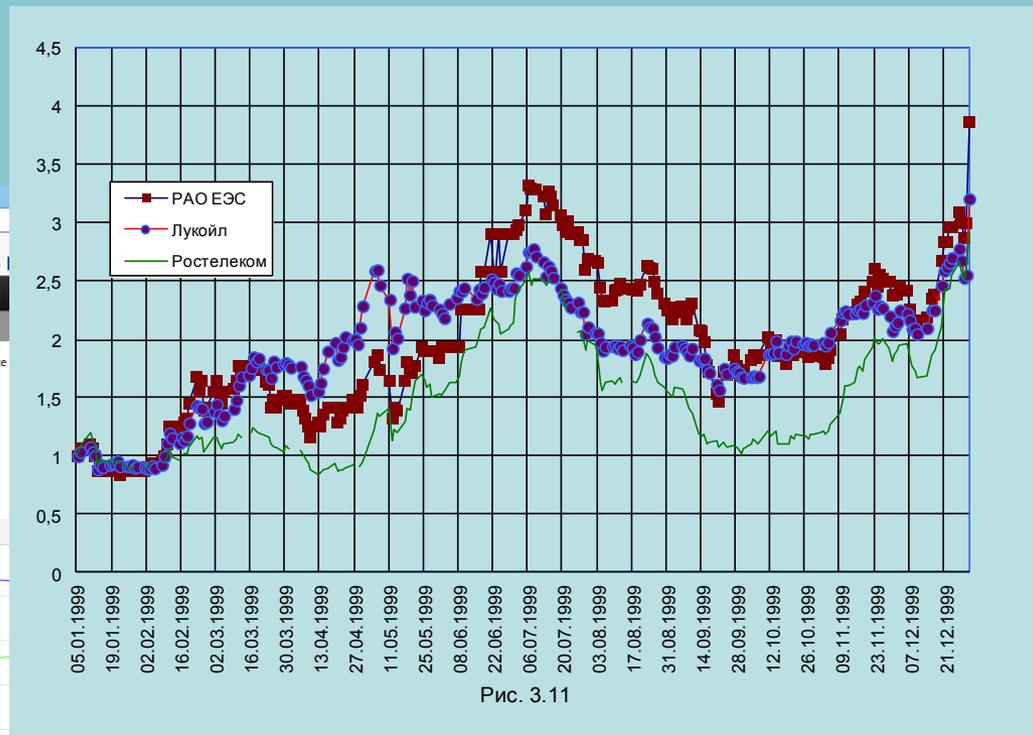
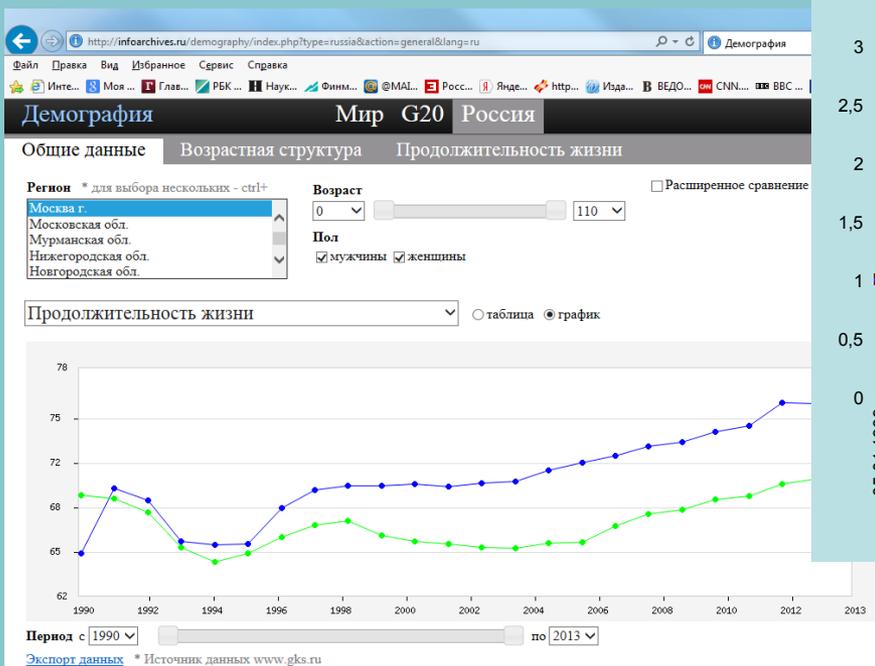


Рис. 3.11

DataPoint

DataPoint – это удобное средство хранения, накопления (обновления) и отображения временных рядов. <http://www.infoarchives.ru/datapoint>

Назначение

Создание возможности для пользователя получать актуальное состояние, историю изменения, экспресс-прогноз в отношении выбранных им показателей.

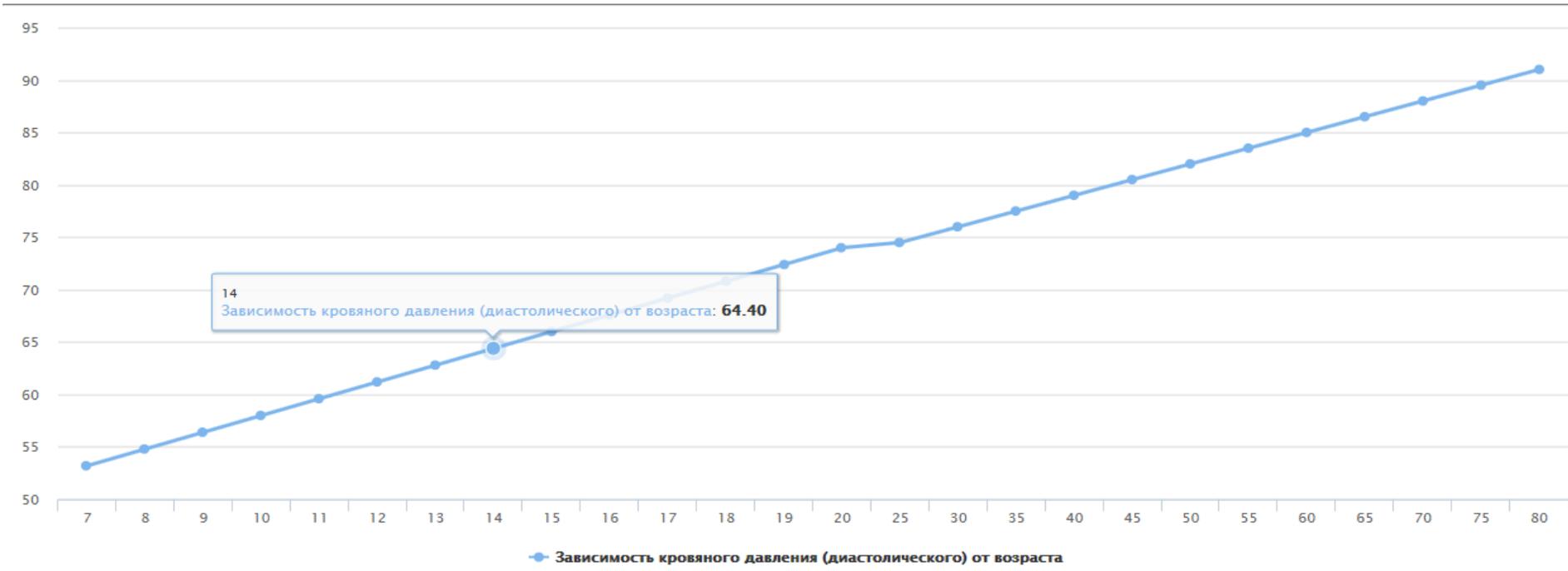
Перечень показателей:

- Базовый набор – «общие» показатели: показатели определяются администратором сервиса (курсы валют, прогнозы погоды, демографические показатели, макроэкономические показатели и др....)
- Персональный набор: показатели определяются и загружаются пользователем (почтовые отправления, показатели деятельности предприятия, цены продуктов, акций, загруженность телефонных каналов, среднее время ответа на запрос и т.д.).

Категория	Ряд	Пользователь	Публичный
TOP-25	Зависимость тормозного пути (м) от скорости (км/ч) при коэффициенте сцепления с дорогой равным 0.5	okrt@yandex.ru	1
TOP-25	Зависимость продолжительности жизни (лет) от потребления морепродуктов (кг/год)	okrt@yandex.ru	1
TOP-25	Зависимость кровяного давления (диастолического) от возраста	okrt@yandex.ru	1
TOP-25	Зависимость кровяного давления (систолического) от возраста	okrt@yandex.ru	1
TOP-25	Численность студентов в России (тыс. человек)	okrt@yandex.ru	1
TOP-25	Численность аспирантов в России (чел.)	okrt@yandex.ru	1

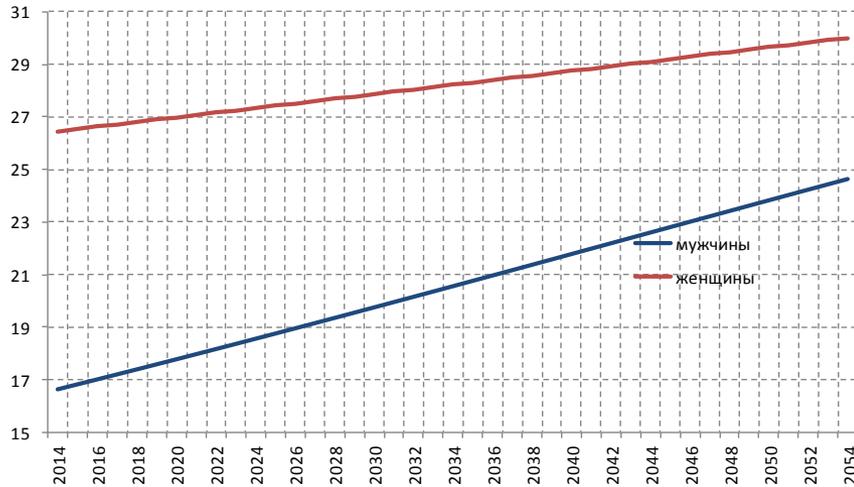
Показано 1 из 11 (всего 11 рядов) - отфильтровано из 41 рядов

[График](#) [Данные](#)

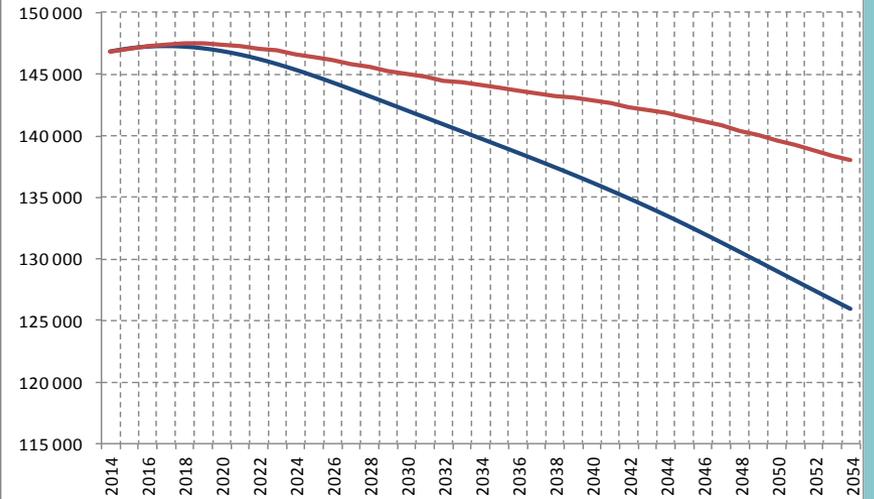


Пример 1. Демографические последствия роста продолжительности жизни

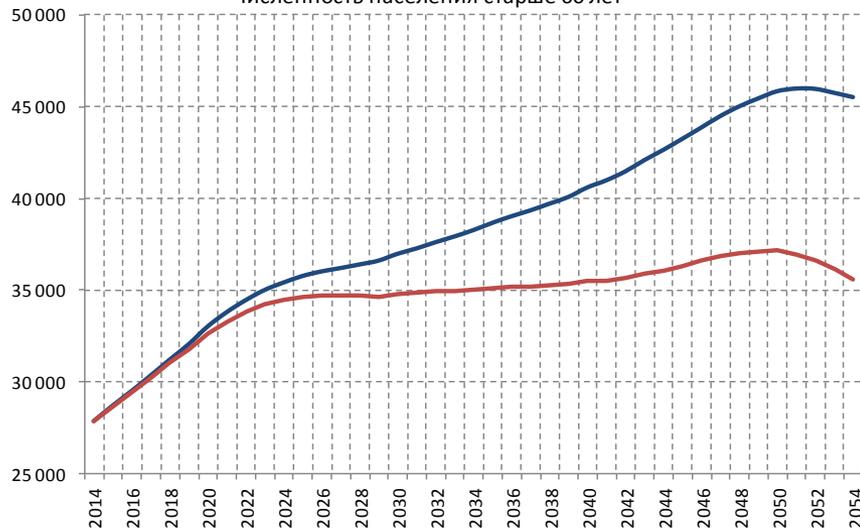
Ожидаемая продолжительность жизни с пенсионного возраста



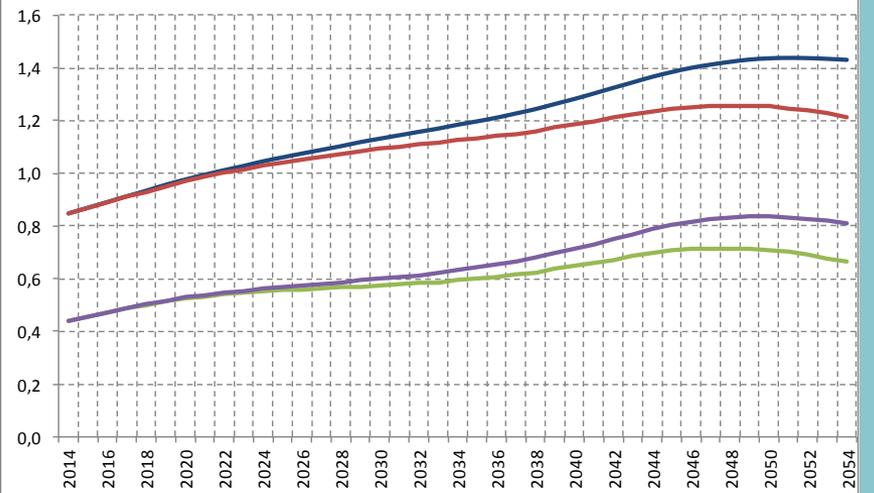
Численность населения



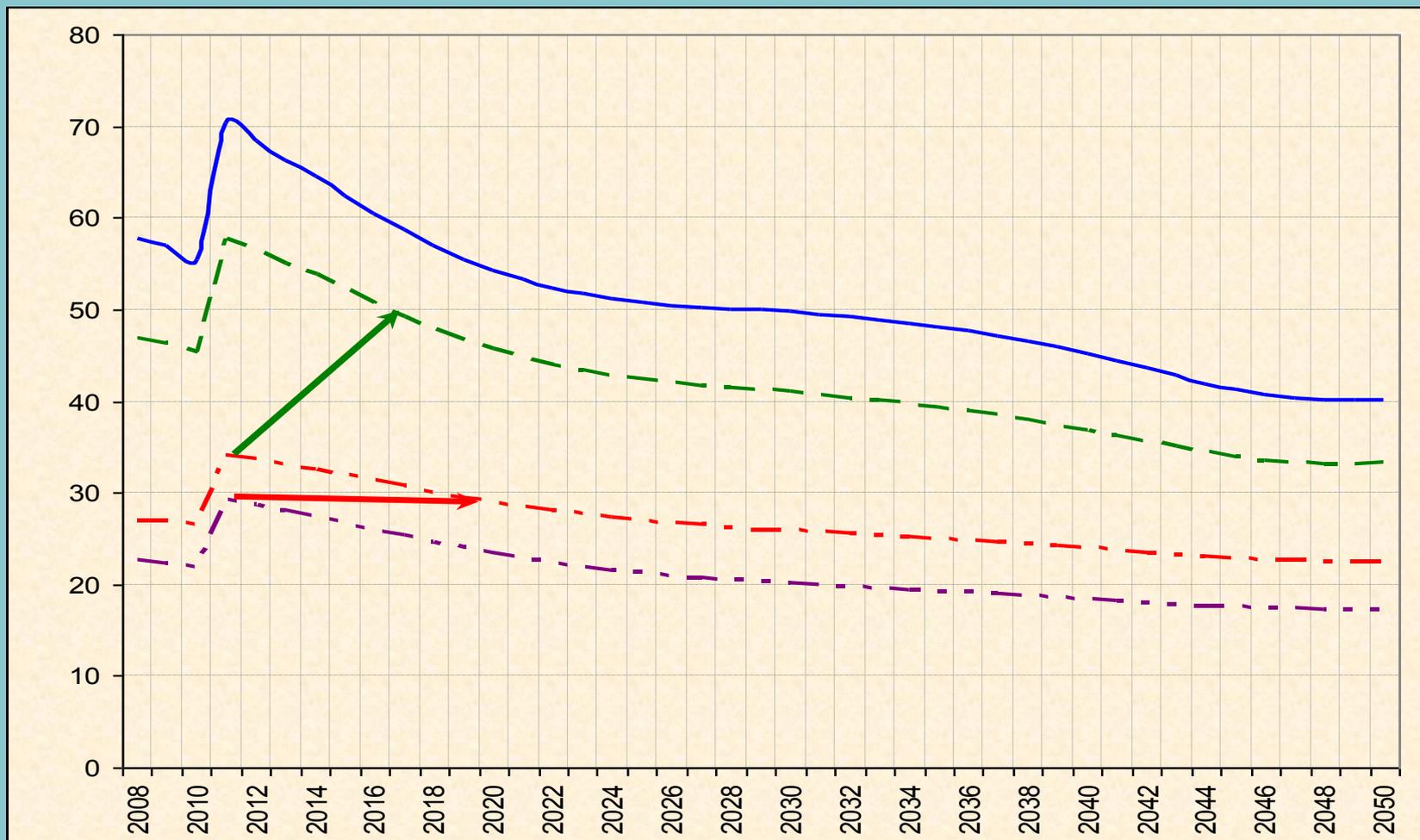
Численность населения старше 60 лет



Демографическая и пенсионная нагрузка



Пример 2. Принцип Оккама:
Оценка возможной ставки замещения для пенсионной системы



Пример 3. Очистка данных

Очистка данных: проблемы и актуальные подходы*

Очистка данных (data cleaning, data cleansing или scrubbing) занимается выявлением и удалением ошибок и несоответствий в данных с целью улучшения качества данных. Проблемы с качеством встречаются в отдельных наборах данных - таких, как файлы и базы данных, - например, как результат ошибок при вводе, утери информации и других загрязнений данных. Когда интеграции подлежат множество источников данных, например - в Хранилищах, интегрированных системах баз данных или глобальных информационных Интернет-системах, - необходимость в очистке данных существенно возрастает. Это происходит оттого, что источники часто содержат разрозненные данные в различном представлении. Для обеспечения доступа к точным и согласованным данным необходима консолидация различных представлений данных и исключение дублирующейся информации.

*http://www.olap.ru/basic/data_clean.asp

Приложение

Единицы измерения количества информации

Измерения в байтах								
ГОСТ 8.417-2002			Приставки СИ		приставки МЭК			
Название	Обозначение	Степень	Название	Степень	Название	Символ		Степень
байт	Б	10^0	-	10^0	байт	В	Б	2^0
килобайт	Кбайт	10^3	кило-	10^3	кибибайт	KiB	КиБ	2^{10}
мегабайт	Мбайт	10^6	мега-	10^6	мебибайт	MiB	МиБ	2^{20}
гигабайт	Гбайт	10^9	гига-	10^9	гибибайт	GiB	ГиБ	2^{30}
терабайт	Тбайт	10^{12}	тера-	10^{12}	тебибайт	TiB	ТиБ	2^{40}
петабайт	Пбайт	10^{15}	пета-	10^{15}	пебибайт	PiB	ПиБ	2^{50}
эксабайт	Эбайт	10^{18}	экса-	10^{18}	эксбибайт	EiB	ЭиБ	2^{60}
зеттабайт	Збайт	10^{21}	зетта-	10^{21}	зебибайт	ZiB	ЗиБ	2^{70}
иоттабайт	Ибайт	10^{24}	иотта-	10^{24}	йобибайт	YiB	ЙиБ	2^{80}

