

УДК: 339.13.024

1.3. Концептуальный подход к построению платформы для рынка данных

Неволин И.В., к.э.н., в.н.с. ЦЭМИ РАН, Москва

Ценность данных связана с программами для их обработки. Это интуитивно понятное соображение может получить формальное основание в виде алгоритмического подхода к определению информации. Предложенный Колмогоровым этот подход оперирует информацией в терминах программы и способа программирования, которые преобразуют объекты. Такой формализм подсказывает естественный способ повышения ценности данных через построение платформы, призванной соединить массивы данных с программами для их обработки. Описанная в статье концептуальная схема такой платформы предусматривает программные сервисы навигации в массивах данных, сервисы ценообразования и поддержки процедуры согласования цен.

Введение

Анонсированный Президентом России национальный проект «Экономика данных» повышает актуальность вопросов, затрагивающих оборот данных. Какие существуют подходы к построению рынка данных? Как устроено ценообразование на рынке данных? Как следует обеспечить безопасность и конфиденциальность данных? Как следует развивать инфраструктуру рынка данных? Эти и подобные вопросы не стали чем-то новым в текущем году, но построение рынка данных переходит в другую категорию. Его масштаб растёт с уровня отдельных организаций и консорциумов до уровня вице-преьера. «Рынок данных» как таковой не звучит в задачах и направлениях нацпроекта¹. Однако направления «Создание и развитие цифровых платформ» и «Искусственный интеллект» непосредственно построены на данных, оборот и ценность которых неотъемлемо связаны с развитием технологий.

Разработчики систем искусственного интеллекта связывают ценность своих разработок с качеством наборов данных для обучения (обучающей выборки). Но и угрозы для таких систем также могут исходить от данных. Например, «отравленные данные» и «дрейф данных» являются теми аспектами, которые характеризуют атаку на системы искусственного интеллекта и некорректность их работы из-за отрыва обучающего набора от реальных процессов с течением времени. Соответственно, ценность программных разработок, информации и управляющих сигналов, получаемых с их помощью, неразрывно связана с ценностью данных. В этой связи целью работы является установление формальной связи между программой и данными. В задачи исследования входит выбор формального описания такой связи и формулирование концептуальных положений для платформы, которая могла бы повысить ценность обоих компонентов продукта – программы и обрабатываемых ею данных – при их «наилучшем» соединении.

Количественное измерение информации

Исторически первым был комбинаторный подход к определению количества информации. Содержательный смысл этого подхода можно выразить через количество двоичных знаков, которые устраняют неопределённость относительно наблюдаемого объекта – системы, описываемой последовательностью двоичных знаков (элементов системы). Развитие комбинаторных идей в терминах теории вероятностей дало дополнительные инструменты анализа и оказалось перспективным для применения систем с большим количеством несвязанных или слабо связанных сообщений, для которых, однако, возможно установить вероятностные закономерности. Проблема такого подхода заключается в интерпретации отрицательных значений количества информации, а также зависимости от вероятностных распределений выборочных наблюдений [Колмогоров, 1965]. Следующим шагом исследований в этой области стал алгоритмический подход к информации, предложенный А.Н. Колмогоровым.

Пусть множество объектов $X = \{x\}$ пронумеровано, и $D = \{n(x)\}$ – множество двоичных последовательностей – номеров элементов x , представленных двоичной записью. При этом требуется взаимно однозначное соответствие между X и D . Функция $l(x)$ обозначает длину последовательности $n(x)$. Предполагая $l(n(x)) \leq l(x) + \text{const}$ на $D \subset X$ и $l(x, y) \leq C_x + l(y)$ при $x \in X, y \in X$, вводят понятие программы p и метода программирования $\varphi(p, x) = y$, ставящего в соответствие объект y программе p и объекту x .

Информацией при этом называется величина $I_A(x; y) = K_A(y) - K_A(y|x)$, где

$$K_A(y|x) = \begin{cases} \min_{A(p,x)=y} l(p) \\ \infty, \neg p|A(p, x) = y \end{cases}$$

Метод программирования A является «наилучшим» в том смысле, что любой другой метод связан с программой большей длины, т.е. $K_A(y|x) \leq K_\varphi(y|x) + C_\varphi$.

¹ Подробнее на TAdviser <https://click.ru/3BHTE7>

Алгоритмический подход, предложенный Колмогоровым, оперирует программами и методами программирования – преобразования некоего «кусочка сведений» (данных, сигнала) в изучаемый объект. Согласно алгоритмическому подходу, информация есть разница между размером тривиального описания объекта (применения к объекту тривиального преобразования) и размером программы, которая «наилучшим образом» преобразует сведения в объект. Взятые в кавычки термины отражают вольности пересказа формальных процедур. В частности, упрощение уводит от частично рекурсивных функций, предложенных Колмогоровым, – то есть таких, что при заданном наборе функций и правил их получения нельзя построить полный перечень функций, который обеспечивает отображение всех сведений в объекты. Понятие «рекурсивности» ассоциируется с понятием «вычислимости» или «проблемой остановки» – ответом на вопрос о том, можно ли до запуска программы строго ответить на вопрос о том, завершится ли обработка данных или программа будет выполнять операции бесконечное время. И определение, данное Колмогоровым, предусматривает отсутствие процедуры преобразования данных об объекте в сам объект. Появление термина «данные» уместно в данной части по двум причинам. Во-первых, набор инструкций, как и сведения об объекте, представимы в двоичном коде – последовательности нулей и единиц. Во-вторых, алгоритмический подход, как отмечает его автор, применим тогда, когда кодирование сигнала по длине представления сопоставимо с программой. Говоря более привычным языком, размер данных в смысле требуемой для хранения компьютерной памяти должен быть сопоставим с размером программы, которая эти данные обрабатывает. Думается, что последние десятилетия, когда большие данные стали входить в повседневную жизнь пользователей компьютеров и деятельность предприятий, прежде всего, с развитием поисковых машин и корпоративных информационных систем, это условие выполняется.

Таким образом, алгоритмический подход оказывается полезным в случае прикладных исследований проблем передачи и обработки данных. Ценность данных определяется, прежде всего, программой, способной преобразовать двоичный код в описание объекта по заранее заданному пользователем формату. Интуитивно это понимается на уровне форматов данных: массив неструктурированных бесполезным.

Далее для краткости и удобства чтения «массив данных» сокращается до «массив» - их следует понимать как синонимы. Использование термина «массив данных» соответствует принятой терминологии, утверждённой национальным стандартом [ГОСТ Р ИСО/МЭК 20546-2021, 2021]. Среди специалистов в области искусственного интеллекта также распространён термин «набор данных». Его можно рассматривать как кальку от английского «dataset», но этот англоязычный термин соответствует в стандарте именно «массиву данных». Это утверждённое стандартом сочетание используется в данной статье.

Создание ценности

Современные экономические исследования активно используют информационные технологии: для обработки актуальных данных используются готовые программы и библиотеки, инструменты собственной разработки. На фоне растущего количества компьютерных моделей и социально-экономических данных, собираемых государственными ведомствами и бизнесом, а также актуальной потребности в программных решениях для использования в государственном управлении всё острее встаёт вопрос о навигации в упомянутом многообразии. На это накладываются проблемы разрозненности сведений: открытые модели публикуются во множестве научных журналов, закрытые – остаются на предприятиях, данные исследователям, зачастую, доступны в агрегированном виде с существенным запаздыванием в их актуальности, несмотря на детальную информацию, собираемую, в том числе, экосистемами практически в реальном времени. Алгоритмизация обработки данных программными средствами подсказывает направление для повышения их ценности – пригодности для извлечения из них информации. И это направление состоит в поиске средств для «наилучшей» обработки массива данных с целью создания новых продуктов и повышения качества управленческих решений.

Система как алгоритм и средство повышения ценности

Решение проблемы доступности данных, их повторного использования, распространения программных модулей для анализа социально-экономических процессов, поднятой, в частности, в работах [Козырев, 2019; Неволин, 2023], может состоять в создании информационной системы, работоспособность которой поддерживают организационные механизмы и программные сервисы. Именно последние являются концептуальной основой системы – это, прежде всего, программные сервисы. Это они позволяют получить больше информации о данных. Именно они влияют на ценность данных. Тем не менее, следует предложить общую схему информационной системы, которая помогает сформировать базовое представление о взаимодействии отдельных сервисов.

Работу информационной системы обеспечивают три крупных блока (Рисунок 1): пользовательский интерфейс, программные сервисы и каталог данных. Стоит отметить, что в системе не предусматривается хранение пользовательских массивов – каталог содержит лишь основные сведения о них, метаданные. Интерфейс системы позволяет пользователю формировать запросы – на загрузку, редактирование и чтение метаданных; на поиск массивов; запрос и улучшение цен за доступ к данным. Эти запросы обрабатываются сервисами, которые объединены в три базовые группы по функциональному признаку: а) для взаимодействия с каталогом данных, б) для запуска информационных посредников (роботов); в) для обеспечения взаимодействия между пользователями. Сервисы группы (а) обеспечивают запись,

редактирование, чтение из каталога, в том числе для целей поиска и навигации. Сервисы группы (б) решают задачи назначения и улучшения цен, облегчающие пользователям выбор условий для распространения своих массивов. Наконец, сервисы группы (в) поддерживают обмен информационными сообщениями между пользователями и системой: совершение юридически значимых действий (например, присоединение к пользовательскому соглашению), контакт пользователей для целей обмена данными, поддержка краудсорсинговых механизмов для обеспечения качества метаданных в каталоге. Ниже частично раскрывается функционал этих сервисов.

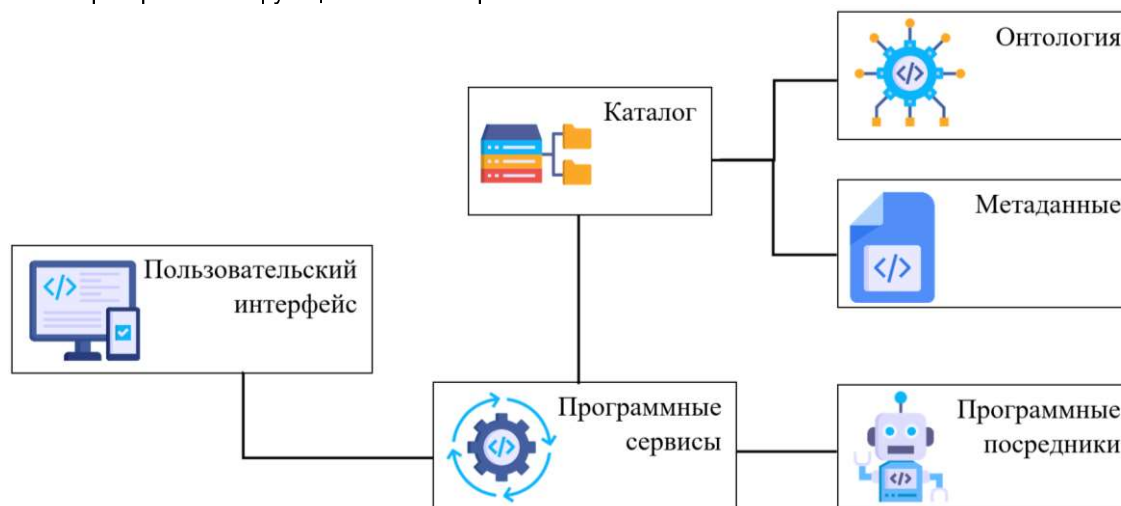


Рисунок 1. Ключевые компоненты информационной системы для поддержки рынка данных. Составлено автором с использованием изображений klišir art на Flaticon.

С каталогом метаданных, который представляет собой базу данных, связана онтология. База содержит расширенное описание доступных массивов данных – не только введённый пользователем текст для характеристики своего массива. Онтология помогает расширить описание в части области применения данных – решаемых с их помощью задач, построения социально-экономических моделей.

Далее описываются основные характеристики блоков в общих чертах.

Описание каталога

Экономические и социальные данные группируются по известным признакам. Основой для группировки могут служить, например, указания службы государственной статистики: демография, труд, промышленность, связь, транспорт и т.п. Причём по формам официальной статистики известно, какой показатель, собираемый на уровне наблюдений, участвует в расчёте того или иного агрегированного показателя. Уже существующий подход к группировке данных по характеризующим ими объектам подсказывает возможность построения онтологии социально-экономических данных для облегчения навигации между ними. О том, почему речь идёт именно об «онтологии», а не о «таксономии» данных, говорится ниже при описании сервиса, облегчающего навигацию пользователей по каталогам данных. Эта (перспективная) онтология предполагает описание детальных данных – на уровне индивидов, домохозяйств, работников, предприятий. Агрегированные данные также могут включаться в базы, чтобы исследователи имели возможность анализировать макроэкономические процессы. Но именно детальная и актуальная информация даёт известную точность анализа. Детальные данные в некоторых разрезах уже доступны исследователям, в том числе, через специализированные информационные системы или курируемые научными организациями массивы данных. С опорой на уже доступные данные базовая классификация, построенная на принятой методической основе, может быть расширена до онтологии через идентификацию сферы применения массивов.

Привязка данных к онтологии – один из способов облегчить пользователям поиск необходимых им для работы массивов. Подход навигации через систему классификаторов давно известен и хорошо зарекомендовал себя. Но современные поисковые машины ориентируются на соответствие контента запросам пользователя. Причём вербальные запросы – уже сложившаяся и хорошо зарекомендовавшая себя практика. Однако для вербальных запросов нужен набор ключевых слов, ассоциированных с данными. Стандартный подход – описание массива данных пользователем на этапе их регистрации в системе – не исключает обогащение этого описания метаданными, извлечёнными из публикаций. На этапе загрузки сведений о доступном пользователю массиве данных существует возможность загрузки полных текстов тех публикаций, где описано использование этих массивов. Далее, известен сервис, который извлекает ключевые слова из публикаций и из контекста цитирований [Papirov, 2021]. Причём контекст цитирования может обозначить сам пользователь. Таким образом, метаданные массивов можно дополнить ключевыми словами публикаций и – более точно – ключевыми словами, которые обозначают контекст цитирования. И здесь уместно вернуться к тому, почему в отношении массивов данных

вводится именно термин «онтология», а не «таксономия». Связь между объектами даёт для этого все основания. С одной стороны, имеются социально-экономические процессы или явления, полные тексты с описанием использования массивов данных. С другой стороны – сами массивы. И онтология устанавливает случаи использования данных – отношения между объектами.

Организационный механизм загрузки данных

Говоря о наполнении каталога, следует отметить несколько принципиальных пунктов, которые позволяют избежать некоторых организационных трудностей. Во-первых, загружаются не сами массивы данных, а только их метаданные. Сами по себе массивы могут исчисляться большим количеством требуемой для хранения физической памяти, что уже создаёт трудности. Но загрузка своего массива в библиотеку также вызывает у пользователя справедливый вопрос: как контролировать доступ к данным? Владельца каталога можно заподозрить в бесконтрольном распространении массива данных и даже в его модификации. Конечно, специальные технические средства могут предоставить некоторые гарантии по недопущению таких действий, но целесообразно ограничиться хранением в каталоге только метаданных. Это не только позволит сэкономить ресурсы хранилища, но также освобождает администраторов каталога от подозрений во вторичном интересе со стороны пользователя: целью системы является распространение сведений о массиве, а не массива данных как такового. Пользователь остаётся в своём праве распоряжаться имеющимся у него массивом по своему усмотрению.

Во-вторых, пользователь загружает сведения о своём массиве данных самостоятельно. Такой шаг позволяет избежать возможной юридической претензии за незаконное распространение объектов интеллектуальной собственности – базы данных, о чём говорилось выше, и метаданных. Последнее может показаться абсурдным, но прецедент создан: научные электронные библиотеки признаются судом виновными в распространении аннотаций к научным статьям как самостоятельных объектов авторского права, при отсутствии специального соглашения с автором². Таким образом, механизм загрузки должен предусматривать самостоятельную загрузку метаданных пользователем с обязательным присоединением к пользовательскому соглашению.

В-третьих, каталог данных предлагает пользователю рекомендательную услугу по назначению цены за доступ к имеющемуся у него массиву. О таком механизме сказано ниже, но важно отметить, что пользовательское соглашение также должно включать информирование о рекомендательном характере цены, предостережения по её изменению, и окончательное решение остаётся за пользователем. Напротив, при взаимодействии двух пользователей – именно пользователей, а не человека с машиной – достигнутые ими предварительные соглашения освобождаются от такой оговорки и могут быть использованы сторонами для обязывания к сделке на достигнутых договорённостях в судебном порядке, если такое потребуется.

Краудсорсинг для повышения качества каталога

Загрузка пользователями сведений об имеющихся у них массивах данных – это первый шаг на пути качества сервиса, который исследователи получают в своё распоряжение. Открытые массивы данных имеют известные недостатки. Во-первых, наблюдения могут публиковаться «как есть», то есть с выбросами или некорректными данными. Во-вторых, данные могут распространяться в неудобном для обработки формате. В-третьих, могут иметь место ошибки в описании и уточнение области применения. Проверка пользователем и использование в анализе, таким образом, свидетельствуют о том, что часть работы по повышению качества данных уже выполнена. В случае, когда сохраняется потребность в сырых – необработанных результатах наблюдений – следует обращаться к первоисточнику. Таким образом, ссылка на первоисточник также составляет метаданные массива, учтённого в системе.

Второй шаг к более качественному сервису – устранение ошибок в написании и обогащение метаданных со стороны потребителей массивов. Если пользователь заметил ошибку в описании данных, ему предлагается возможность подать соответствующую заявку на исправление. Причём заявка становится заметна всем пользователям системы, и её либо одобряет автор загруженного массива, либо участники специальной отметкой о полезности данной заявки повышают её валидность. Это один из вариантов работы краудсорсингового механизма, но могут быть и другие реализации. Также пользователи могут загрузить свои статьи об использовании массива данных, что, далее, через сервис для определения контекста цитирования, приводит к расширению и изменению веса ключевых слов, ассоциированных с конкретным массивом данных. Перспективными для системы (и полезными для других пользователей) также могут оказаться отметки потребителей массивов о проблемах дрейфа данных, дублирования массивов, индикация рисков открытого массива данных.

Назначение цен за использование данных

² Можно указать, например, на постановление Девятого арбитражного апелляционного суда № 09АП-26164/2023 № 09АП-24271/2023 от 18.05.2023г. по делу № А40-64528/22, где приводится аргументация такого решения. Также можно указать на постановление Суда по интеллектуальным правам от 28.09.2023г. по делу № А40-269679/2021, в котором размещение названий и аннотаций не приравнивается к нарушению исключительных прав. При этом, что характерно, информационные цели и цитирование не сыграли своей роли в доказательстве, в другом – сыграли. То есть опасения в судебном преследовании за размещения информационных материалов о массивах данных не являются обоснованными.

Массивы данных могут распространяться за вознаграждение. Загружая информацию об имеющихся данных, пользователь сам назначает цену за доступ к ним. Она может быть нулевой. Может иметь некоторую положительную оценку. По мере накопления в системе информации о сделках/ контактах между пользователями становится возможной ценовая дискриминация: назначение разных цен разным группам потребителей данных [Козырев, 2011]. В случае заинтересованности пользователь может воспользоваться функцией составления ценового меню для пользователей разного типа, сигнализируя о качестве массива данных, скрываемого за тем или иным сигналом (с соответствующей ценой). Термин «сигнал» в отношении массива данных остаётся неясным без развёрнутого комментария. В следующем абзаце понятие «информации» не привязано к какому-либо из упомянутых в начале статьи определений. В общем, подразумевается уточнение сведений об объекте, устранение неопределённости, и даже обывательское представление о термине приемлемо для понимания сути сигналов в процессе передачи информации.

Известен протокол передачи информации, который предполагает, что цена является для покупателя сигналом о качестве того, что за ней скрывается. В случае с информацией это более чем уместно по сравнению с «частными благами» или материальными товарами. Последние можно как-то измерить и оценить перед покупкой. С информацией дело обстоит сложнее: раскрытие информации часто подразумевает её передачу. Поэтому, чтобы «направить» покупателя именно к той информации, которую он ищет, в теории вводятся сигналы. Причём эти сигналы могут отождествляться с ценой (цена при этом и является сигналом о содержании информации), но могут передаваться за отдельную плату. Соответственно, приобретая сигнал, потребитель устраняет неопределённость относительно состава искомой информации.

Простейший сценарий появления сигналов о качестве массивов данных – это распространение ограниченных фрагментов. Безусловно, описание массива в системе включает часть сведений и соответствует принятой онтологии. Но предположим, пользователю нужен не весь массив наблюдений не по всем показателям, а только их части. Например, из 50 имеющихся показателей пользователю нужно только 15, и он согласен приобрести их только в том случае, если вариация значений показателя под номером 6 по всем наблюдениям окажется достаточной для его целей. Соответственно, речь может идти о доступе к наблюдениям по одному-двум показателям, а затем (в случае положительного тестирования) – к наблюдениям по 15 показателям. Соответственно, возникает вопрос о том, как оценивать доступ к фрагментам массивам данных при повторяющихся взаимодействиях. И в этом вопросе также есть достижения, на которые можно опереться [Неволин, Соколов, 2013].

Оптимизация роялти и устаревание данных

Говоря о платном доступе к данным, нельзя не упомянуть инструменты оптимизации цен для пользователей – покупателей данных. Контент имеет тенденцию к сокращению срока коммерческой жизни. Так, спрос на популярные произведения в Интернете «затухает» в течение довольно непродолжительного времени, и только отдельные произведения пользуются популярностью в течение многих лет. По литературным произведениям, например, разумный срок правовой охраны оценивается в 8-15 лет – гораздо короче установленных законом жизни автора и 70 лет после его смерти [Pollock, 2008]. Можно предположить, что такая же зависимость существует и для данных: более актуальные сведения ценятся сильнее тех, что описывают период десятилетней давности. Соответственно, цена за доступ к данным также может дифференцироваться не только по типу получателя, но и во времени. Следует ещё раз отметить – автор массива данных может указать безвозмездный вариант доступа. Но, если некоторые массивы доступны на платной основе, следует обсудить и то, как сделать этот – ценовой – барьер доступа разумным.

Предположим, что изначально доступ к данным планируется открывать за некоторую плату, которая остаётся неизменной в течение всего срока размещения. С учётом дисконтирования стоимость денежного потока ограничена, и можно говорить о том, как распределить эту величину между потенциальными пользователями в течение времени. Эту ситуацию можно моделировать через игру двух лиц – автора массива и обобщённого потребителя, которые при разных предпочтениях делят единичные ресурсы (ежегодные взносы за использование) в течение времени. Аппарат решения такой игровой ситуации предполагает решение с более высокими (относительно базового уровня) ценами в первые периоды и более низкими – в последние (вплоть до бесплатного доступа). Причём техника реализации алгоритма решения игры предполагает участие программы, имитирующей действия покупателя (при априорно заданных ожиданиях о его предпочтениях) [Неволин, 2012]. Соответственно, сервис оптимизации цен за доступ к данным предполагает наличие вероятностного распределения о темпах «устаревания» данных – изменения спроса с течением времени. И это распределение может уточняться по мере спроса на данные. К примеру, если спрос в начальные периоды оказывается ниже, автор массива может получить рекомендацию о снижении цены.

Существуют и другие механизмы организации платного доступа, основанные на понимании данных как клубного блага [Detering, 2001]. В таком случае следует подключать механизмы выявления истинных предпочтений через систему голосования с тем, чтобы участники «клуба» оплачивали приоритетный доступ к данным в течение некоторого предустановленного периода, по окончании которого данные

переходят в разряд общественного блага (а база данных – в категорию общественного достояния, говоря юридическими терминами) с открытым доступом всем желающим без взимания какой-либо платы.

Заключение

Формальная связь между программой и данными устанавливается при переходе к алгоритмическому определению количества информации. С практической (и экономической – в смысле извлечения выгоды) точки зрения эта связь означает более высокую ценность данных при наличии программных средств для их обработки. Описана концептуальная схема повышения ценности данных путём подключения специальных сервисов для их обработки. Затронуты вопросы именно экономической ценности данных в результате расширения возможностей по использованию накопленных массивов. Вопросы информационной безопасности оставлены в стороне, хотя действия злоумышленников могут серьёзно повлиять на ценность как самих массивов, так и продуктов на их основе. Безопасность действительно требует специальной проработки, но этот аспект смещает фокус с общей концепции на выделенный компонент. Скорее, он заслуживает отдельной статьи для подробного освещения.

Литература

1. ГОСТ Р ИСО/МЭК 20546-2021 Информационные технологии (ИТ). Большие данные. Обзор и словарь. М., 2021.
2. Козырев А.Н. Цифровизация, математические методы и системный кризис экономической науки // Цифровая экономика. – 2019. – № 4(8). – С. 5-20.
3. Козырев А.Н. Моделирование НТП, упорядоченность и цифровая экономика // Экономика и математические методы. – 2011. – Т. 47, № 4. – С. 131-142.
4. Колмогоров А.Н. Три подхода к определению понятия «количество информации» // Проблемы передачи информации. – 1965. – Т.1 – Вып. 1. – с. 3–11
5. Неволин И.В. Экономические модели и недискриминируемый доступ к данным // Цифровая экономика. – 2023. – № S5(26). – С. 12-17.
6. Неволин И.В., Соколов Н.А. Назначение цен на информирующие сигналы при передаче ноу-хау // Обзорение прикладной и промышленной математики. – 2013. – Т. 20, № 2. – С. 181-182.
7. Неволин И.В. Оптимизация ставок роялти в лицензионных договорах на основе распределённых вычислений и сетевых технологий / диссертация на соискание ученой степени кандидата экономических наук – Москва, 2012. – 143 с.
8. Babaioff M., Kleinberg R., Paes Leme R. Optimal mechanisms for selling information // Proceedings of the 13th ACM Conference on Electronic Commerce. – 2012. – pp. 92-109.
9. Detering D. Ökonomie der Medieninhalte. Allokative Effizienz und soziale Chancengleichheit in den Neuen Medien / Münster: LIT. Zugl. – 2001. – 288 s.
10. Nevolin I, Kozyrev A. Developing CRIS Module for Technology Transfer // Procedia Computer Science, Rome, 13–15 мая 2014 года. – Rome, 2014. – P. 158-162.
11. Parinov S. Citation contexts as a data source for evaluation of scholarly consumption. Scientometrics 2021 Vol. 126, pp. 9249–9265.
12. Pollok R. Forever Minus a Day? Theory and Empirics Of Optimal Copyright Term // MPRA Paper No. 8887. - 2008. – 28p.

References in Cyrillics

1. GOST R ISO/MJeK 20546-2021 Informacionnye tehnologii (IT). Bol'shie dannye. Obzor i slovar'. M., 2021.
2. Kozyrev A.N. Cifrovizacija, matematicheskie metody i sistemnyj krizis jekonomicheskoj nauki // Cifrovaja jekonomika. – 2019. – № 4(8). – S. 5-20.
3. Kozyrev A.N. Modelirovanie NTP, uporjadochennost' i cifrovaja jekonomika // Jekonomika i matematicheskie metody. – 2011. – Т. 47, № 4. – S. 131-142.
4. Kolmogorov A.N. Tri podhoda k opredeleniju ponjatija «kolichestvo informacii» // Problemy peredachi informacii. – 1965. – Т.1 – Вып. 1. – с. 3–11
5. Nevolin I.V. Jekonomicheskie modeli i nediskriminiruemyj dostup k dannym // Cifro-vaja jekonomika. – 2023. – № S5(26). – S. 12-17.
6. Nevolin I.V., Sokolov N.A. Naznachenie cen na informirujushhie signaly pri peredache nou-hau // Obozrenie prikladnoj i promyshlennoj matematiki. – 2013. – Т. 20, № 2. – S. 181-182.
7. Nevolin I.V. Optimizacija stavok rojal'ti v licenzionnyh dogovorah na osnove rasprede-lennyh vychislenij i setevykh tehnologij / dissertacija na soiskanie uchenoj stepeni kandidata jekonomicheskikh nauk – Moskva, 2012. – 143 s.

*Неволин Иван Викторович, к.э.н., в.н.с. ЦЭМИ РАН (i.nevolin@cemi.rssi.ru)
ORCID: 0000-0002-8462-9011*

Ключевые слова

большие данные, экономика данных, экономика платформ, информационные системы, структура данных

Ivan Nevolin. Conceptual framework for platform design to support data markets**Keywords**

big data, economics of the data, economics of the platforms, information systems, data structure

DOI: 10.33276/DE-2024-02-03

JEL classification O30 – инновации, исследования и разработки, изменения в технике, права интеллектуальной собственности (общие вопросы); D4 – структура рынков, ценообразование, проектирование рынков.

Abstract

The value of the data relates to the programs processing them. This intuitive consideration could be supported by the formal grounds of an algorithmic approach to the definition of information. Proposed by Kolmogorov, this approach operates with information in terms of a program and a programming method that transform objects. This formalism suggests a natural way to increase the value of data by building a platform designed to connect datasets with the processing programs. The article describes a conceptual scheme of such a platform. Provided with software services for navigation through the datasets, pricing services and services to support price matching procedure the platform could facilitate data usage and influence their value.