

Революция в обработке больших юридических данных

11.07.2023

СТАТЬИ

LEGAL TECH



Не правда ли, если бы статья называлась «Онтология для юристов», вы бы сразу пролистали журнал дальше. За последний месяц два раза при попытке включить в программы юридических конференций об инновациях тему стандартизации юридической работы на основе общепринятой онтологии / таксономии я слышал от юристов, что это крайне скучно. Данная тема действительно скучна, если подходить к ней исключительно с технической

точки зрения. Я лично наблюдал, как юристы засыпали на legaltech-конференции, когда специалист по онтологии начинал излагать техническую сторону данного понятия. Модные нынче темы искусственного интеллекта, больших данных и data science представляются гораздо более интересными для обсуждения, но в основе всех этих тем лежит онтология.

Заметьте, в юридической отрасли практически нет общепринятых стандартов: ни единого стандарта ISO для юристов (если не считать три ISO — 27001, 37001 и 22301), но они больше касаются узких моментов операционной деятельности, таких как информационная безопасность, управление непрерывностью бизнеса и предотвращение коррупции. Все это не может не удивлять, ведь юристы, как и другие профессионалы, пользуются (сознательно или нет) общеизвестными стандартами, взять хотя бы унифицированный размер листа бумаги для всех документов, и страдают, как все, от разнообразия коннекторов для любимых гаджетов, порой одного и того же производителя. В то же время это объяснимо, ведь «наша работа уникальна» — таков любимый аргумент юристов при любой попытке хотя бы частично автоматизировать их деятельность. И все же в юридической отрасли без стандартизации и общепринятой онтологии / таксономии не обойтись.

Зачем юристам стандартизация и онтология

Юридическая отрасль сложная и фрагментированная, с большим разнообразием правовых систем, терминологии и практики. Это может затруднять обмен информацией и сотрудничество различных организаций. Стандартизация и онтология — два подхода, которые могут помочь решить данную проблему: стандартизация нужна для разработки общих правил и процедур обмена информацией, онтология — для понимания того, какие формальные знания и представление можно использовать для рассуждения об информации.

При совместном использовании стандартизация и онтология создают общую основу для обмена юридической информацией и управления ею. Это повышает эффективность,

прозрачность и улучшает качество взаимодействия между юристами.

Представьте, что вы работаете в крупной юридической фирме. У вас продвинутая организация, в которой до мелочей налажена разметка всех дел и процессов на основе собственной таксономии. Теперь вы захотели обмениваться информацией с системой крупного клиента, которая использует другие разметки (теги). Например, вместо «оспаривания патентной заявки в споре об интеллектуальной собственности» эта система использует «подачу возражения на заявку на патент в споре о нарушении прав ИС». Оба описания идентичны по своей сути, но обмен данными между ними невозможен.

Вы, конечно, можете добавить теги системы клиента в свою систему или заставить его использовать ваши теги, однако это решит проблему только временно, до появления нового клиента с другой системой и онтологией. В отдельных странах имеются местные стандарты описания юридической работы, обычно создаваемые регуляторами, но в эпоху глобального бизнеса с системами автоматизации, внедряемыми на разных континентах, локальные стандарты не решают проблему в корне.

SALI Alliance

Для создания мирового стандарта описания юридической работы в 2017 г. три отраслевые ассоциации — Association of Legal Administrators, Legal Marketing Association и International Legal Technology Association — сформировали Standards Advancement for the Legal Industry Alliance ([SALI Alliance](#), или [SALI](#)). Альянс — некоммерческая организация, которая предоставляет свои разработки совершенно бесплатно по принципу open source, и каждый заинтересованный юрист может участвовать в создании онтологии. Членами SALI являются крупные юридические фирмы (Baker McKenzie, Clifford Chance, DLA Piper, White & Case, Greenberg Traurig и др.) и глобальные вендоры legaltech-решений (Thomson Reuters, LexisNexis, Wolters Kluwer, Inapp, Netdocuments, Litera, iManage и др.), онтологию альянса поддерживают клиенты и организации с мировым именем (Microsoft, Shell, Citigroup, CLOC, ACC и др.).

В марте 2022 г. SALI выпустила вторую итерацию своей онтологии Legal Matter Standard Specification (LMSS 2.0), которая содержит более 20 тыс. уникальных идентификаторов юридической работы. LMSS покрывает сущность юридической работы (substance of law) и ее операционную составляющую (business of law). При этом онтология не зависит от конкретной категории пользователей — она едина для юридических фирм, клиентов и вендоров технологических решений, что обеспечивает совместимость пользователей при обмене данными. Другими словами, LMSS позволяет собирать данные от всех участников рынка и проводить бизнес-аналитику. Управление 20 тысячами уникальных идентификаторов — большая работа, но тегирование проводится автоматически в фоновом режиме, теми решениями, которые внедряли LMSS.

С помощью LMSS теги присваиваются не только делам, но также проектам и документам в рамках дела. Каждое дело, проект или документ могут иметь множество тегов, как и выбираемая вами на «Озоне» рубашка. Вряд ли вы нашли бы нужную рубашку, если бы тегнули «мужская рубашка» — нужен контекст (цвет, размер, материал, бренд, страна происхождения, ценовой диапазон и т.д.). Примерно то же самое в юридическом бизнесе. Приведу несколько примеров того, как извлечение данных и структурирование на базе

единой онтологии дает возможность провести глубокий анализ дел, проектов и документов на стороне клиента и юридической фирмы.

Теги приведены здесь на английском, поскольку изначально онтология LMSS была разработана на английском, но уникальным идентификаторам, которые служат базой для тегов, можно присваивать названия на любом языке. Например, для patent law, Patentrecht или патентного права создан единый уникальный идентификатор, что обеспечивает совместимость систем на всех языках. На стороне вендоров LMSS может применяться для большого диапазона решений: matter management, legal research, e-billing, RFP, document management и др.

Сама SALI частично использует для LMSS стандартизацию, например предварительно определенные значения для локаций (ISO 3166- 2) и валют (ISO 4217), а также коды NAICS для определения отраслей промышленности.

Sali относится к категории Metadata Standards, на которой базируются все остальные категории стандартов. Примерами других юридических Metadata Standards являются [Noslegal](#) и Legal Issues Taxonomy ([LIST](#)) от Stanford Legal Design Lab.

Content Standards и Process Standards базируются на «Metadata Standards». Многим юристам в России известна инициатива стандартизации соглашения о неразглашения информации [oneNDA](#), к которой уже присоединилось большое количество игроков на рынке. В России создан аналог oneNDA под названием [SimpleNDA](#), адаптированный к реалиям отечественного рынка. Примеров для Process Standards в юридической индустрии пока нет, но это, по-видимому, вопрос времени.

Как уже отмечалось, стандарты SALI являются основополагающими для развития искусственного интеллекта, data science и программного обеспечения. Рассмотрим, как это выглядит на практике в области искусственного интеллекта. Стартовой точкой для применения стандарта является машинное обучение (Machine Learning, ML), когда алгоритм

обучается определенной специфике юридической работы или, как в случае с большими языковыми моделями, — более широкому применению. Дальше модели применяются уже на практике в сегменте ИИ под названием «Естественная обработка языка (Natural Language Processing, NLP), например при автоматической классификации документов, поиске юридической информации или анализе судебной практики либо потенциальных регуляторных рисков.

Внедрение SALI

SALI предлагает четыре ступени внедрения своего стандарта:

СТУПЕНЬ 1

ВАЛИДИРОВАНИЕ (VALIDATE)

Как собственная структура тегов соотносится с тегами SALI?

СТУПЕНЬ 2

РАСШИРЕНИЕ (EXPAND)

Добавьте теги SALI к своим.

СТУПЕНЬ 3

СОДЕЙСТВИЕ (CONTRIBUTE)

Внесите свой вклад в развитие SALI, чтобы пользователи использовали ваши теги

СТУПЕНЬ 4

ДОБАВЛЕНИЕ (ADD)

Добавьте новые отношения в граф знаний (Knowledge Graph – семантическая сеть, которая умеет выявлять взаимосвязи между разными объектами) для улучшения данных и их анализа.

Чтобы у вас не оставалось сомнений по поводу внедрения в свою работу стандарта SALI, приведу примеры положительного эффекта от этого.

Повышение эффективности. Стандартизация и онтология повышают эффективность юридических процессов, упрощая обмен информацией и сотрудничество между различными организациями. Юридическая фирма может использовать онтологию для автоматического обмена юридическими документами со своими клиентами, это экономит ее время и трудозатраты, помогает обеспечивать клиентов необходимой информацией.

Повышение прозрачности. Стандартизация и онтология способствуют прозрачности в юридической отрасли, облегчая понимание и отслеживание информации. Например, государственное учреждение может использовать онтологию для публикации юридических данных в машиночитаемом формате, это позволит общественности легко получить доступ к данным и понять их, что повысит степень прозрачности и подотчетности.

Улучшение сотрудничества. Стандартизация и онтология расширяют сотрудничество в юридической отрасли, представляя собой основу для обмена юридической информацией и управления ею. Юридическая фирма может использовать онтологию для сотрудничества с другими юридическими фирмами по сложному юридическому вопросу. Это гарантирует, что все будут работать с одной и той же информацией, а значит, облегчит сотрудничество и достижение успешного результата.

Снижение затрат. Стандартизация и онтология помогают сократить расходы в юридической отрасли за счет упрощения автоматизации задач и оптимизации процессов.

Так, юридическая фирма, использующая онтологию для автоматизации процесса классификации документов, экономит время и деньги. Кроме того, повышается точность процесса классификации.

Таким образом, какими бы скучными ни казались юристам стандартизация и онтология, в эпоху сбора и анализа большого объема цифровых данных они жизненно необходимы, так как поднимают юридическую работу на совершенно новый уровень. Игнорируя их, вы попросту не продвинетесь!



**Хольгер
Цшайге**

Генеральный директор "Инфотропик Медиа", Член правления ELTA, Член Advisory Board Global Legal Tech Hub

СТАТЬИ

LEGAL TECH