

Онтологизация данных Европейского союза как переход от экономики данных к экономике знаний

О.В. Гринько, В.П. Куприяновский, О.Н. Покусаев, Ю.И. Волокитин, И.В. Понкин, Д.Е. Намиот, А.И. Редькина

Аннотация— Настоящая статья продолжает серию работ, посвященных применению формальных онтологий. Сегодня, в подавляющем большинстве случаев, значительная часть мировой информации существует в текстовой форме: деловые документы и т.п. Именно в такой исторической форме человек сохранял свои знания жил и работал. Сегодня текстовые документы перешли в цифровую форму, сохранив прежний внешний вид. Такое быстрое увеличение числа цифровых текстов вызывает рост потребности в текстовой аналитике и выдвигает вопрос о нахождении "умных" способов чтения и понимания текстов, и, в конечном счете, вопрос получения знаний из них. В работе рассматривается семантический веб, RDF и SPARQL. Подробно рассматривается технологическая платформа PoolParty Semantic Suite. Семантически и онтологически расширенные тексты открывают новую эру использования контента. Используя текстовую аналитику и технологии Web 2.0, огромное количество контента, разбросанного в различных формах в документах разных форматов, конвертируется в обогащенные, преобразованные и управляемые информационные фрагменты. Таким образом, машины не только оказываются способны обрабатывать длинные вычислительные строки символов и индексировать большие объемы данных, но они также могут хранить, управлять и извлекать информацию на основе их смысла и логических отношений между вещами в знаниях. Онтологии и семантика добавляют еще один слой в новом Интернет и могут показывать связанные вещи, факты и предметы вместо простого сопоставления слов.

Ключевые слова— онтология, знания, RDF.

Статья получена 11 октября 2018.

О.В. Гринько - ООО "Т-Система" (email: gov@t-systema.com)

В.П. Куприяновский - Национальный центр компетенций в области цифровой экономики, МГУ имени М.В. Ломоносова; Центр цифровых высокоскоростных транспортных систем РУТ (МИИТ) (email: vrpupriyanovsky@gmail.com)

О.Н. Покусаев - Центр цифровых высокоскоростных транспортных систем РУТ (МИИТ); Российская академия транспорта (email: o.pokusaev@rut.digital)

Ю.И. Волокитин - ООО ТАС (email: i18021958@gmail.com)

И. В. Понкин – РАНХиГС при Президенте РФ (email: i7@inbox.ru)

Д.Е. Намиот - МГУ имени М.В. Ломоносова; РУТ (МИИТ) (email: dnamiot@gmail.com)

А.И. Редькина - МГЮА им. О.Е. Кутафина (МГЮА). (email: juriste.ap@gmail.com).

I. ВВЕДЕНИЕ

Публикации про быстрое развитие онтологических проектов (формальных онтологий) [1-13] вызвали как значительный интерес у наших читателей, так и множество вопросов. В основном, они стали концентрироваться вокруг проблематики не столько приложений, сколько собственно данных, как таковых, обрабатываемых в онтологических приложениях. Но логика построения онтологических систем и работа с данными создают механизмы работы, во-первых, сильно между собою связанные (приложения и данные) и, во-вторых, зависящие от области использования или домена. Для того чтобы объяснить это необходимо рассказать о формальных описаниях, и авторы обычно начинают употреблять нотации на формальных языках, которыми далеко не все читатели владеют. Читатель перестает понимать смысл того, что авторы ему пытаются объяснить. Мы решили попробовать рассказать об этих явлениях цифровой экономики, не употребляя формализованных языков, заменив их наиболее выразительными рисунками и текстами. Так явление онтологизации данных и создание репозитория формализованных знаний становится повсеместным мировым явлением, с которым уже сталкиваются обычные пользователи интернета, которые активно пользуются онтологиями и WEB 2.0, может быть, даже не зная об этом.

На просторах интернета мы нашли союзников в этом подходе – две европейские компании: Ontotext и Semantic Web Company. Первая из них из Болгарии, а вторая из Австрии. Они успешно участвовали в десятках онтологических проектах на уровне ЕС, общее количество которых, как мы предполагаем, исчисляется тысячами. Ontotext и Semantic Web Company пошли ровно таким же путем пытаясь объяснить экономику данных ЕС и ее основу – онтологию простыми, по возможности, текстами и оригинальными рисунками.

II. ОНТОЛОГИЯ – КЛЮЧ К РАБОТЕ С ЦИФРОВОЙ ИНФОРМАЦИЕЙ

Сегодня, в подавляющем большинстве случаев, значительная часть мировой информации существует в текстовой форме: деловые документы,

правительственные документы, правовые акты, социальные медиа-потоки, клинические испытания, медицинские архивы, электронные письма, блоги. Именно в этой исторической форме человек сохранял свои знания жил и работал. Сегодня это быстро практически перешло в цифровую форму, сохранив, зачастую, прежний внешний вид (но не содержание!). Такое быстрое увеличение числа цифровых текстов (через Интернет и Интранет) вызывает рост потребности в текстовой аналитике и выдвигает вопрос о нахождении "умных" способов чтения и понимания текстов, и, в конечном счете, вопрос получения знаний из них. Много преобразований прошло с письменным словом - от самых старых сохранившихся надписей на глиняных табличках до настоящего поразительного количества документации, хранящейся в облачных системах (или других репозиториях). Но одна вещь осталась неизменной: информация, содержащаяся в наших текстовых источниках так же хороша, как наши способности и инструменты для ее извлечения и интерпретации (рисунок 1).



Рис. 1. Данных сегодня очень много (будет еще больше), разнообразие их велико – значит к ним нужен ключ (онтологический) (источник – Ontotext)

Читателю, вероятно, уже наскучил образ семантического веб-стека или веб-торта (его зовут и так), но, тем не менее, мы приводим его еще раз на рисунке 2. В свое оправдание мы можем только сказать, что это классика, и один из самых последних по детализации составляющих веб-стека, который нам нужен для изложения дальнейших соображений. И если на рисунке 1 показан замок, который надо открыть, чтобы попасть в удивительный мир цифровой экономики или экономики данных (официальная цель ЕС), то на рисунке 2, фактически, изображен ключ к этому цифровому миру.

Web 2.0 или Semantic Web базируется на основах оригинальной сети интернета и постепенно наполняется новыми онтологизированными данными. Данные (старые и новые) должны описываться метаданными. Эти метаданные будут идентифицировать данные, данные взаимосвязей и связывать данные с концепциями, чтобы машины могли их понимать.

Данные должны быть идентифицированы однозначно, и это делается с использованием универсального идентификатора ресурса (URI) или интернационализованного идентификатора ресурсов (IRI).

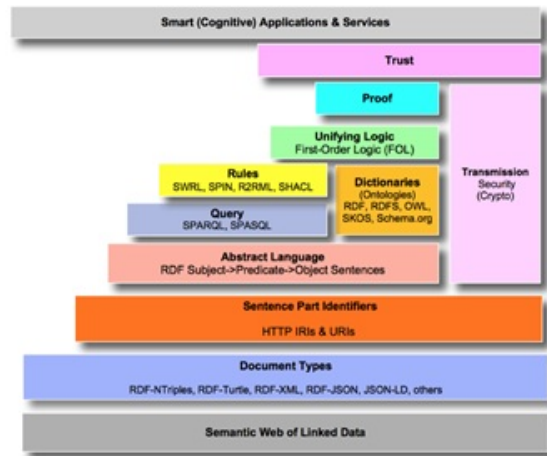


Рис. 2. Семантический веб-стек (он же Semantic Web Cake). (Источник- DBpedia и Idehen 2017)

Формат описания ресурсов (RDF) предоставляет модель данных. Значение добавляется на более высоком уровне с тем, что мы называем онтологиями. Другими словами, RDF задает синтаксис, тогда как онтологии определяют семантику, или, говоря проще, то, что называют по старинке семантическим вебом, логичнее назвать онтологическими.

Так же, как HTML является строительным блоком оригинальной сети данных «старого» интернета, RDF является строительным блоком информации Semantic Web или Web 2.0. Веб-контент может выставлять свою семантику путем встраивания операторов RDF в веб-страницы. Существует много способов сделать это: RDF, RDF-XML, RDF-JSON, JSON-LD, Microdata и т. д. Семантические данные, уже обработанные и сохраненные в формате RDF, могут быть запрошены. Так же, как MySQL существует для запроса реляционных баз данных, SPARQL - это язык для запросов к хранилищам данных, размеченным с помощью RDF. Учитывая семантику, правила могут помочь в применении логики и рассуждений.

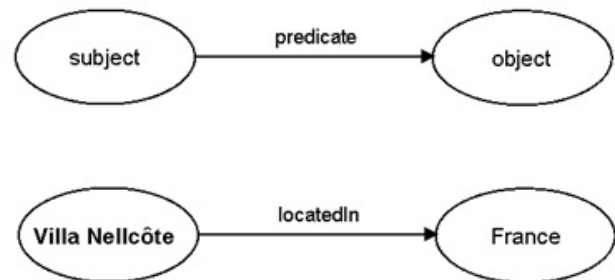


Рис. 3. RDF тройки или Triple: Subject, Predicate, Object (источник - DBpedia и Heitmann 2011).

Как следует из названия, RDF помогает нам описать любой ресурс, пока этот ресурс имеет уникальный

идентификатор. Другими словами, RDF помогает нам определять данные о других данных, то есть метаданные. RDF имеет три компонента: субъект, предикат, объект. Это утверждение о связи между объектом и субъектом (рисунок 3). Таким образом, тот факт, что Villa Nellcôte находится во Франции, можно выразить как Triple RDF. Все три части тройки выражаются как URI, литералы или пустые узлы. Когда мы объединяем много таких утверждений вместе, получаем то, что называется RDF Graph. Объекты и субъекты являются узлами графа. Предикаты образуют соединительные дуги. Например, мы можем заявить, что Франция находится в Европе, Париж является столицей Франции, в Париже проживает 2,2 миллиона человек. Каждый из них может быть выражен как RDF Triple. В совокупности они образуют начальный онтологический граф RDF.



Рис. 4. Semantic Web добавляет данные ссылок в Интернете и добавляет уровень значения в данные. (Источник- DBpedia и Petkova 2016)

RDF сам по себе не дает смысла данным. RDF - это модель данных, способ выражения отношений. Чтобы дать смысл, определены словари и онтологии. Они обычно пишутся в терминах классов, их свойств и отношений с другими классами. Например, тройка RDF может выразить, что Париж является столицей Франции, но для компьютера это все еще не имеет смысла. Словарь может определить, что столица - это тип города, город принадлежит стране, а страна - политическая единица. Это помогает компьютеру понять смысл контекста, хотя он никогда не сможет по-настоящему понять, как это делают люди. RDF Schema (RDFS) - простой словарь, в то время как язык веб-онтологий (OWL) является более мощным выразительным языком. Такое единение онтологий и онтологизированных данных читатель может видеть на рисунке 4, где, как нам кажется, удачно показана разница между Web 2.0 и прежним Web, только надо отметить, что эти два веба в жизни представляют собой единый интернет, но при этом уже по-разному работающий в разных странах и доменах.

Основной результат такого развития - это изменение способов работы с онтологизированной информацией в Web 2.0 в тех доменах, где этот процесс прошел (рисунок 5), и реальная схема парадигмы работы с

данными как таковыми. Semantic Web не делает машины интеллектуальными в смысле искусственного интеллекта или машинного обучения (AI / ML).

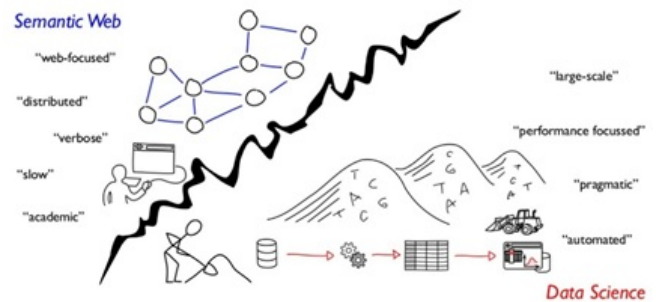


Рис. 5. Разница между семантическим Web и наукой о данных. (Источник- DBpedia и Lampra 2018, слайд 3).

Вместо того, чтобы просить машины понимать людей, мы помогаем машинам решать четко определенные проблемы с четко определенными данными с помощью четко определенных операций. Благодаря многим технологиям AI/ML продвинулся вперед и позволил машинам видеть, слышать и говорить. В середине 2010-х годов появились приёмные помощники, чат-боты, приложения для компьютерного зрения и многое другое. Это стало возможным благодаря доступности данных для обучения алгоритмов ML. Хотя некоторые алгоритмы ML требуют, чтобы некоторые данные были помечены или помечены для обучения, нет необходимости добавлять семантические метаданные ко всем данным в виде Semantic Web. Semantic Web дополняет подходы AI/ML. Чатботы и умные помощники будут использовать онтологии и обработанные данные просто из-за удобства работы с ними. Семантический Web может добавлять фоновые знания в системы AI/ML, особенно в тех областях, где данных недостаточно. Логично и понятно, что AI / ML применяют онтологии для концептуализации знаний домена для Semantic Web.

Онтология изначально имеет метафизический смысл, но в информатике это относится к формальному описанию знания. Так, уже сегодня, описаны концепции и их отношения в определенных доменах Web 2.0. Определены классы, атрибуты и отношения, такие как ограничения, правила и аксиомы. Они представляют знание этого домена, базирующееся в новом интернете. Причиной этой смены парадигмы и серьезных усилий является ограничения возможностей экономического развития в том, что сегодня называется цифровой экономикой или экономикой данных. Старая парадигма наглядного представления знаний в виде «документов» и связей между ними показана на рисунке 6.

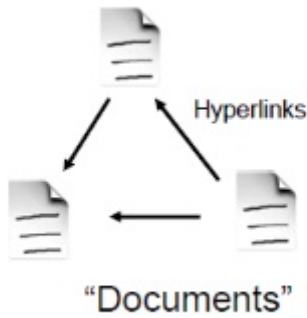


Рис. 6. Как связаны «документы» в «старом интернете» (источник – STI Innsburg)

Веб-документы в «старом интернете» (рисунок 6) состоят из основных элементов и это:

1. Имена (URI);
2. Документы (ресурсы), описываемые HTML, XML и т. д.;
3. Взаимодействие через HTTP. Ссылки (точнее гиперссылки) между документами или якорями в этих документах позволяют устанавливать их близость в поиске.

Есть два основных недостатка такого подхода:

1. Эти ссылки не являются типизированными;
2. Поэтому поисковые системы Интернета не работают со сложными запросами.

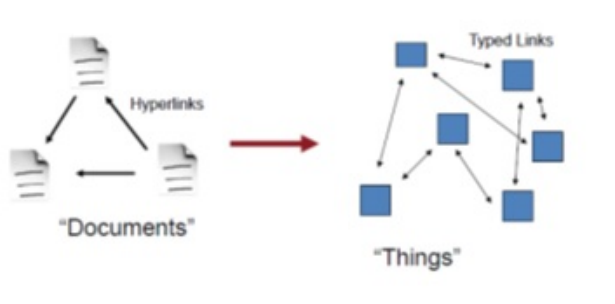


Рис.7а. Переход к семантическому интернету (источник – STI Innsburg)

На рисунке 7 мы попробовали показать переход к семантическому интернету, который еще называют Web of Data. Вместо понятия документ вводится более абстрактное понятие «вещь», которое привязывает новый интернет к физическому миру. Применение такого подхода позволяют реализовать:

- Связи между произвольными вещами (например, лицами, местоположениями, событиями, зданиями и т.п.);
- Структура данных в Интернете - страницы становятся ясными
- Вещи, описанные в Интернете, называются страницами и получают URIs
- Связи между вещами сделаны явным и набираются в системы онтологических знаний.

Так появляются связанные данные о вещах, которые касаются использования технологий Semantic Web для того, чтобы публиковать структурированные данные в Интернете и устанавливать связи между данными о вещах и источниками знаний. Как «работают» такие связи между знаниями о данных, относящимся к вещам в распределенном мировом Web 2.0 можно увидеть на рисунке 7. Очень важно, что связи между знаниями описываемыми основной онтологией предметной области или решаемой задачи с другими онтологиями или совместное использование знаний в Web 2.0 (рисунок 8) позволяют использовать фактически напрямую ранее корректно созданные онтологии, что приводит к удивительным по времени и экономике результатам, о которых написано в [1-13], хотя сама по себе онтологизация знаний не является простой и дешевой.

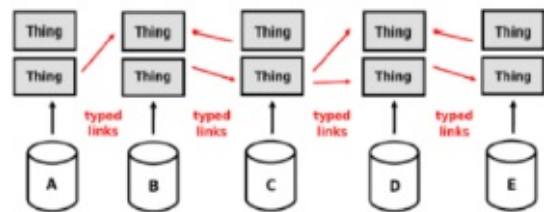


Рис. 7b. Связи между знаниями о данных относящимся к вещам в распределенном мировом Web 2.0 (источник – STI Innsburg)

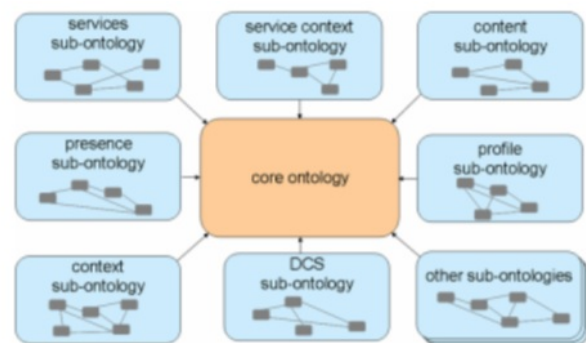


Рис. 8. Связи между знаниями, описываемыми основной онтологией предметной области или решаемой задачи с другими онтологиями или совместное использование знаний в Web 2.0 (источник – STI Innsburg)

История развития нового интернета формализованных онтологий давно уже вышла из области экспериментов и стала промышленным явлением экономических успехов внедрения формальных онтологий. Так уже в 2007 году сообщалось, что нефтегазовая промышленность использует RDF / OWL для объединения данных из разных источников и стандартизации обмена данными, совместного использования и интеграции между партнерами или приложениями. Также стало возможным совместное управление знаниями. Этот процесс использования механизмов нового интернета

стал быстро развивающимся стандартом ISO 15926 имеющим неформальное название «Нефтегазовая онтология». В этом году были приняты и опубликованы два новых приложения к нему: ISO / TS 15926-12: 2018 и ISO / TS 15926-13: 2018. Их содержание, соответственно: онтология интеграции жизненного цикла, представленная на веб-языке онтологии (OWL) и онтология для интеграции промышленных данных на протяжении всего жизненного цикла. Онтологии представлены в формате Web Ontology Language (OWL). Цена следования этому стандарту - примерно двукратное сокращение времени освоения нефтегазовых месторождений.

В апреле 2010 года Facebook запустил Open Graph, который веб-издатели могут использовать для интеграции своих веб-страниц в социальный граф Facebook. Это позволяет Facebook понять, что нравится пользователю, давать персональные рекомендации или подключать пользователей со схожими интересами. В этом случае была принята упрощенная форма RDF. Во время чемпионата мира по футболу 2010 года сайт BBC использовал семантические веб-технологии для динамического отображения контента. Были использованы запросы SPARQL и рассуждения OWL 2 RL. С успехом этого проекта, в январе 2013 года, BBC привержена разработке Linked Data Platform для обеспечения динамической семантической публикации. Сайт BBC с 2008 года также был ранним примером использования семантической сети, а сегодня так работают многие правительственные учреждения.

III. КАК ПОПАДАЮТ ЗНАНИЯ В WEB 2.0

Во все большей цифровой среде, где мир живет и работает уже сегодня, работы с текстами имеют решающее значение. Если мы будем использовать тексты с максимальной производительностью и минимальными затраченными усилиями, мы должны рассмотреть то, как мы кодируем машиночитаемую информацию в них. Только тогда компьютеры могут помочь нам с огромным разнообразием данных и в увеличении скорости обработки и использования текстовых потоков. Программные агенты (боты), читающие наши тексты, могут не быть среди общих читателей в обычной библиотеке, как мы представляем себе, когда думаем о представлении, хранении и повторном использовании информации или там, где они должны быть по нашему мнению. Эффективное исследование, быстрое просеивание информации открытие фактов в эпоху данных немислимо без помощи алгоритмов и формализованных онтологических языков как инструментов их описания. На их совокупные с техническими средствами (вычислительными ресурсами, электронную память, средствами связи и отображения) способности обрабатывать огромные источники информации и представить нам результаты, относящиеся к нашим потребностям и, зависит качество огромной части информации, из которой мы извлекаем знания (в том числе, в формальных онтологиях), которые мы

обнаруживаем, и, в большинстве случаев, важно понимание возможностей практических результатов, к которому мы приходим.

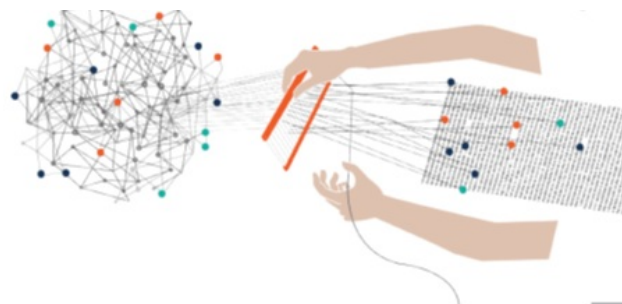


Рис. 9. Вплетание онтологических и семантические цифровых «заметок на полях» в тексты (источник – Ontotext)

Получение высококачественных, структурированных и управляемых машиной фрагментов информации из текстов - это возможности для построения текстовой аналитики знаний, открытых в них. Чтобы раскрыть соответствующую информацию, мы теперь должны взглянуть на процесс чтения заново, с точки зрения, по своей сути - как на процесс извлечения информации из любой системы кодирования, которому компания Ontotext [18] нашла очень образное представление (рисунок 9). «Цифровое ткачество» текста и данных вместе, естественно, связано с необходимостью получения хорошо организованной доступной информации из неструктурированных гетерогенных текстовых источников. Применяемый правильно, использующий семантические веб-технологии, процесс подготовки текстовой аналитики объединяет элементы данных и документов, чтобы помочь организациям отвечать и соответствовать одной из самых больших проблем в современной цифровой среде: работе с неструктурированными данными. Когда данные и текст взаимосвязаны, контент становится высоко управляемым, простым в использовании и правильно структурированным [18].

Онтологическая (семантическая технология) определяет и связывает данные в Интернете или внутри предприятия, развивая языки, чтобы выражать богатые, самоописывающиеся взаимосвязи данных в форме, которую могут обрабатывать машины. Таким образом, машины не только оказываются способны обрабатывать длинные вычислительные строки символов и индексировать «тонны» данных, но они также могут хранить, управлять и извлекать информацию на основе их смысла и логических отношений между вещами в знаниях. Онтологии и семантика добавляют еще один слой в новом Интернет и могут показывать связанные вещи, факты и предметы вместо простого сопоставления слов.

Онтологическая (Семантическая) технология используется для определения и связывания данных (в Интернете или внутри предприятия) путем разработки языка для выражения богатых взаимосвязей данных в форме, в которой машины могут это обрабатывать.

Таким образом, машины не только способны обрабатывать длинные вычислительные строки символов индексировать «тонны» данных, но они также могут хранить, управлять и получать информацию по смыслу и логическим отношениям. Основное различие между онтологическими (семантическими) технологиями и другими технологиями для данных, реляционных баз данных, например, заключается в том, что семантическая технология касается значения, а не структуры данных (рисунок 10).

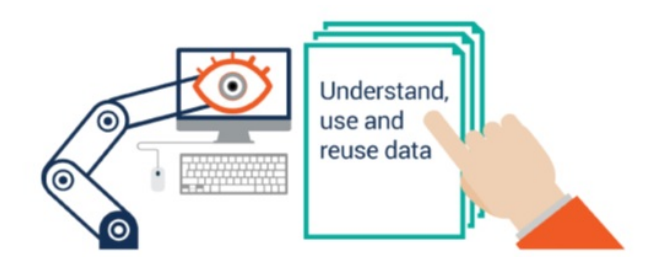


Рис. 10. Необходимо чтобы компьютер (машина) понимала использование и повторное применение информации (источник – Ontotext).

Созданная онтология знаний, превращение текстов описывающих вещи в части данных и их дальнейшая взаимосвязь состоит из ряда методов и процессов, ключевое значение которых - извлечение семантической и онтологической информации вместе с их аннотациями, используя различные алгоритмы для анализа свободного текста, куски его преобразуются в структурированные взаимосвязанные элементы данных то вещах. Это позволяет организациям легко и эффективно использовать информацию, отслеживать и понимать отношения в разрозненных текстовых источниках, находить актуальную информацию, обнаруживать факты. Извлечение структурированных данных из неструктурированных текстов, собственно так ее можно и определить, но суть этого процесса - это процесс ткачества данных как переход в онтологический граф [18] (рисунок 11).



Рис. 11. Процесс ткачества данных как переход информации и знаний о вещах в онтологический граф (источник – Ontotext)

Как существенная часть процесса онтологизации, связанные с этим методы обработки данных не только обогащают тексты данных, но также возглавляют создание перспективных, тесно взаимосвязанных текстовых цифровых активов, которые, по сути, превращаются в цифровые онтологизированные знания,

базирующиеся на связанных данных.

Связанные данные (Linked Data) разбивают информационные силосы, которые существуют, но между различными форматами и разрушают преграды между различными их источниками [18].

Это еще одно кардинальное достижение онтологий – прозрачность или совместимость информации. Связанные данные проще и эффективнее обеспечивают интеграцию данных и просмотр с помощью комплексированных данных, из-за стандартов, которым они придерживаются.

Проекты Инициативы Европейской платформы (IoT-EPI) определяют эту тему как “Интернет вещей и платформ для подключенных смарт-объектов и имеющие целью реализации систем IoT, расширенных в сеть платформ для подключенных устройств и объектов, поддерживающих интеллектуальные среды, предприятия, служб и лиц с динамической и адаптивной конфигурацией возможностей” [24].

Конкретные области фокуса исследовательской деятельности являются архитектуры и семантическая функциональная совместимость, которые надежно охватывают различные варианты использования. Целью является создание инфраструктуры с динамической конфигурацией и интеграционные платформы для подключенных интеллектуальных объектов, охватывающих несколько технологий и множество интеллектуальных артефактов. Экосистема IoT-EPI была создана с целью увеличения влияние европейских исследований и инноваций, связанных с IoT, в том числе, семи европейских перспективных проектов на платформах IoT: AGILE, BIG IoT, INTER-IoT, VICINITY, SymbIoTe, bIoTore и TagItSmart.

Это техническое исследование [24] дает представление о совместимости в платформах и экосистемах IoT, созданных и используемых IoT-EPI, охватывая аспекты взаимодействия, проблемы и подходы, которые справляются с интероперабельностью в существующих IoT и представляет некоторые идеи относительно будущего совместимости в этом контексте. В ней представлены возможные решения, и возможная архитектура платформы взаимодействия IoT. С другой стороны это отличный пример проведения открытой политики в научной сфере – работа [24] была подготовлена и издана в 2018 солидном научном издательстве River Publishers и находится в открытом доступе.

После стандартизации данные предприятия затем превращаются в интеллектуальные части данных, которые легко соединяются с множеством других данных.

IV. СВЯЗАННЫЕ ДАННЫЕ

Связанные данные - это термин, обозначающий структурированные данные, которые взаимосвязаны и подключены через открытые стандарты W3C, или онтологизированная информация, о которой мы говорили выше. Технологии связанных данных (для более подробного объяснения отсылаем к рисунку 2)

уникальны, поскольку они позволяют нам (через наших программных агентов) путешествовать по различным наборам данных, находить, делиться и легко и эффективно интегрировать информацию.

Связывание данных с выбранными фрагментами текстов является ключом к открытию контента для более эффективного доступа, использования и управления им.

Когда все нужные данные онтологизированы, подключены нужными связями и объединены разрозненные источники, появляется точка зрения единого пространства домена. Можно видеть отношения между всеми видами необходимых записей и корпоративным содержанием домена. В части организации это может быть набор: от маркетинговой и сбытовой документации до внутренних бизнес-записей, сервисов, цепочек поставок и информации о товаре.

При интегрировании данных в Web 2.0 текстовые источники получают еще один слой значений и оказываются собраны в сборник машиночитаемого корпоративного контента до такой степени, что соединенные точки становятся соединенными онтологическими узлами [18] (рисунок 12).

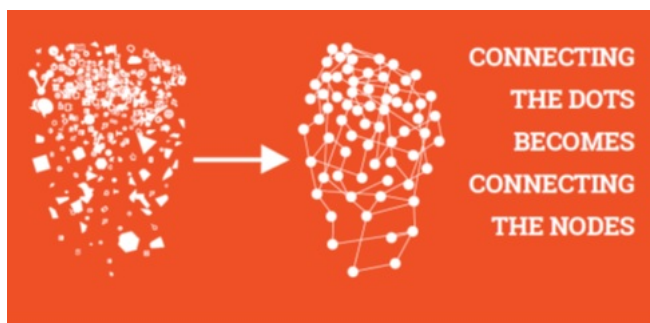


Рис. 12. Соединенные точки становятся соединенными онтологическими узлами (источник – Ontotext)

В парадигме онтологической (семантической) технологии создание цифровые "заметки на полях" (*marginalia* маргиналов) для машин для чтения и обработки путем добавления семантических метаданных, связанных со всем документом или конкретными частями внутри него. Семантические описания могут быть добавлены в любой текст - веб-страницы, регулярные (не веб-документы), текстовые поля в базах данных и т.д. Оставляя цифровые заметки «margin» для новых читателей ботов появляется на блоке информации, чтобы использовать онтологии или формализованные знания, и это добавляет еще один уровень - онтологии. Последнее означает, что появляются текстовые источники, которые значительно улучшают нашу эффективность в доступе, использовании и повторном использовании их. Семантически аннотированные тексты богаты машинно-обрабатываемыми соединениями, то есть в контексте и ссылках, поставляемый в читаемой форме для компьютеров. Ибо, если мы хотим, чтобы алгоритмы понимали текстовые источники, мы должны предоставить этим машинным читателям ссылки на

понятия с однозначным определением, которое имеется в виду. Учитывая это концептуальное понимание смысловой аннотации, давайте теперь рассмотрим детали этих процессов.

Чтобы добавить некоторые технические подробности к вышеупомянутому концептуальному пониманию, семантической или онтологической аннотации или семантического тегирования надо сказать, что это привязка имен, атрибутов, комментариев, описаний и т. д. осуществляется к целому документу, описывающему вещь, фрагментам документа, фразам или словам. Она предоставляет дополнительную информацию (метаданные) о существующем фрагменте текста. По сравнению с тегами, которые добавляют релевантность и точность полученной информации, семантические аннотации идут на один уровень глубже: они обогащают неструктурированные или полу-структурированные данные контекстом, который далее связан со структурированным знанием домена [18] (рисунок 13). Он позволяет получить результаты, которые явно не связаны с исходным поиском (именно так сегодня работает интернет).

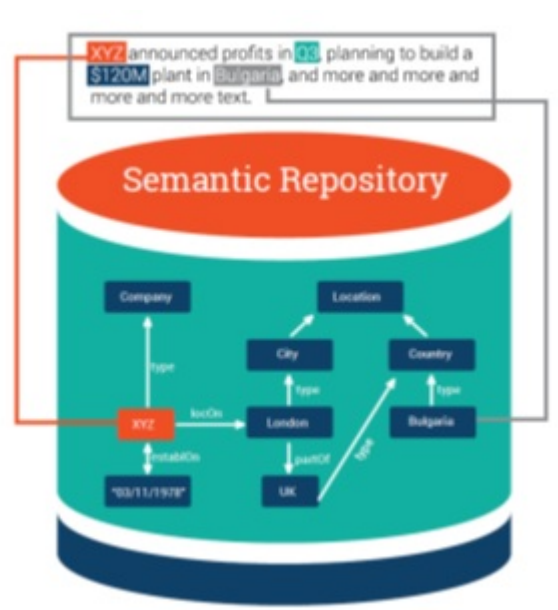


Рис. 13. Контекст, который далее связан со структурированным знанием домена через семантический репозиторий (источник – Ontotext)

Другой пример: посетитель прибывает на сайт газеты и хочет получить информацию о руководителе банка в Азии. Семантическая технология позволяет веб-сайту возвращать гораздо более сложный набор результатов из исходного поискового запроса. Поскольку система имеет понимание отношений, определяющих банковских управляющих в целом (через онтологию), она может использовать всю базу данных опубликованного текстового контента в более сложный способ, захват отношений, которые были бы упущены только компьютерным анализом. Как это происходит, мы покажем в нескольких шагах:

Текстовый анализ

Mr. Haruhiko Kuroda governor of Bank of Japan announced 0.1 percent growth...

Рис. 14. Разметка текста (источник – Ontotext)

После того, как слова помечены, объекты затем могут быть распознаны и иметь информацию из разных источников, связанных с ними. В этом случае система знает, что «Харухико Курода» - это тип человека (рисунок 14). Это процесс можно себе представить как извлечение концепта знаний (рисунок 15).

Concept Extraction

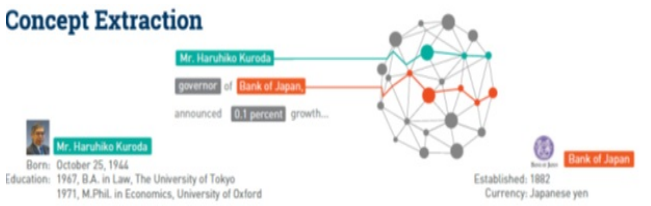


Рис. 15. Процесс извлечения онтологического концепта (источник – Ontotext)

Для того, чтобы система поняла, что «руководитель (governor)» - это «работа», существующая в субъекте «Банк Японии» должно существовать правило, которое утверждает это как абстракцию. Это называется онтологией (представляйте в этом процессе онтологию как о книгу правил: она описывает мир, в котором существует исходный материал или вещь). Рассказывая компьютеру о том, какие это элементы данных и как эти отношения могут быть оценены автоматически, становится возможным обрабатывать сложные фильтры запросов и поисковые операции. Используя тот же пример, система может создавать формальные, машиночитаемые отношения между Харухико Курода, его ролью руководителя и Банком Японии (рисунок 16).

Relationship Extraction



Рис. 16. Процесс извлечения онтологических отношений (источник – Ontotext)

Когда у нас есть аннотации, связанные с онтологией и знаниями, мы можем начать делать вывод отношений между субъектами в системе, которые не были напрямую связаны человеческим действием, организуя процессы индексирования и запоминания (рисунок 17).

Индексирование и запоминание

Indexing and Storing



Рис. 17. Процесс индексирования и запоминания (источник – Ontotext)

Другой пример: посетитель прибывает на сайт газеты и хочет получить информацию о губернаторах банка в Азии. Семантическая технология и онтология позволяет веб-сайту возвращать гораздо более сложный набор результатов из исходного поискового запроса. Поскольку система имеет понимание отношений, определяющих банковских управляющих в целом (через онтологию), она может использовать всю базу данных опубликованного текстового контента в более сложных способах, захват отношений, которые были бы упущены одним компьютерным анализом. Эта технология позволяет вывести отношения, которые конкретно не указаны в источнике материала: потому, что система знает, что Харухико Курода является руководителем Банка Японии, он способен работать с другими сотрудниками Банка Японии, что он работает в Токио, который находится в Японии, которая представляет собой набор островов в Тихом океане. Так контент превращается в цифровые онтологические знания [18] (рисунок 18).

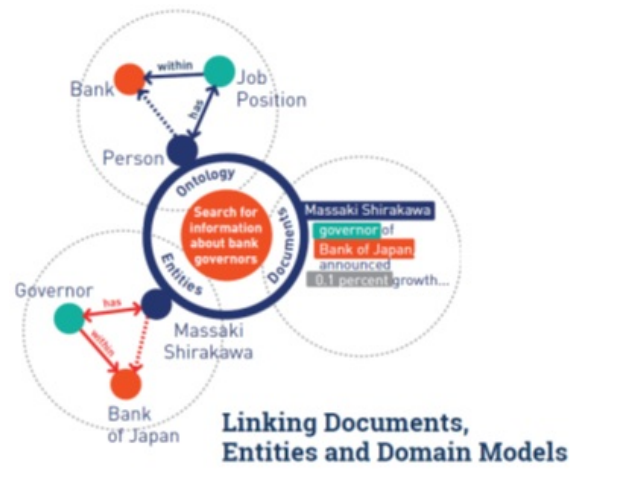


Рис. 18. Как контент превращается в цифровые онтологические знания (источник – Ontotext)

V. ИСПОЛЬЗОВАНИЕ ДАННЫХ СЕМАНТИЧЕСКОГО ИНТЕРНЕТА В ОНТОЛОГИЧЕСКИХ ПРИЛОЖЕНИЯХ НА ПРИМЕРЕ POOLPARTY SEMANTIC SUITE

PoolParty Semantic Suite - это технологическая платформа, предоставляемая компанией Semantic Web, о которой мы говорили выше. Основанная в ЕС компания принадлежит к ранним пионерам движения Semantic Web. Программное обеспечение поддерживает предприятия в области управления знаниями, аналитики данных и организации контента. Продукт использует

основанные на стандартах технологии, определенные W3C, что предотвращает блокировку поставщика. Клиентами являются, в частности, Boehringer Ingelheim, Credit Suisse, Европейская комиссия, REEEP, Kluwer Wolters и Группа Всемирного банка.

PoolParty Semantic Suite занимается успешно коммерциализацией технологий Semantic Web и поэтому мы решили использовать его данные в этой части статьи. В 2009 году вышел первый выпуск Semantic Software PoolParty. С тех пор продукт превратился из инструмента управления таксономией в многофункциональную платформу семантического и онтологического программного обеспечения, которая позволяет компаниям внедрять графы корпоративных знаний для интеграции структурированных и неструктурированных данных. Продукт быстро развивается благодаря сильному фокусу R & D и является неотъемлемой частью нескольких исследовательских проектов ЕС по инициативе Horizon 2020.

PoolParty Semantic Suite - это модульный и гибкий программный пакет. Он отличает девять модулей, которые могут быть индивидуально объединены в зависимости от бизнес-задачи:

- Управление таксономией и тезаурусом;
- Выделение текста и сущности;
- Управление онтологией;
- Концептуальная маркировка;
- Интеграция данных;
- Управление связанными данными;
- Семантический поиск;
- Рекомендательная система;
- Аналитика и визуализация.

Контентные активы получают семантически обогащенные теги и помещаются в контекст, сопоставляющий их с графом знаний. Это основа семантических приложений в качестве поисковых или связанных порталов данных. PoolParty Semantic Suite использует технологии Semantic Web, продвигаемые W3C. Основа информационной архитектуры строится с применением системы SKOS (Simple Knowledge Organization System), онтологий и принципов связанных данных. Любые данные, обрабатываемые в PoolParty, преобразуются в графики RDF и могут запрашиваться с помощью языка SPARQL.

Существенным преимуществом такого подхода является то, что проекты таксономии, разработанные в PoolParty, могут быть связаны с данными практически из любого репозитория. Компании получают выгоду от автоматического объединения объектов между ресурсами из семантического уровня на основе графа и ресурсами знаний из хранилищ данных, таких как системы управления документами. Поскольку спрос и требования для более интеллектуальных корпоративных приложений остаются без изменений, стек технологий Semantic Web становится все более привлекательным

для разработчиков и примерами успешного роста бизнеса в этом на направлении являются компании из ЕС Ontotext и Semantic Web Company

Управление онтологией PoolParty позволяет создавать графы знаний в своем собственном темпе и скорости пользователям практически без программистов. Так есть возможность создания своих собственных онтологий и пользовательских схем, повторно используя уже существующие онтологии, такие как FOAF, FIBO, schema.org и CHEBI, которые становятся приложением через исполнительную онтологию или невидимую трансляцию формализованных и проверенных описаний. Их можно применять к существующим таксономиям с легкостью получая практически готовые результаты (рисунок 19).

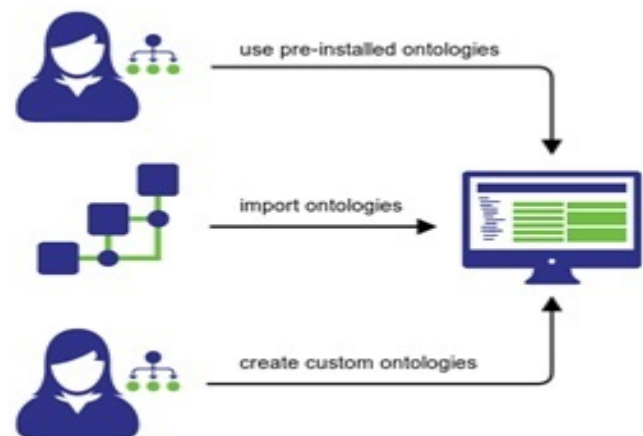


Рис. 19. Соединение различных онтологий для создания онтологии пользователя (источник - Semantic Web Company)

Предварительно установленная онтологическая библиотека включает в себя широкий спектр моделей знаний (FOAF, schema.org, FIBO, CHEBI, ...), которые можно легко расширить в соответствии с собственными потребностями. В PoolParty Semantic Suite можно легко загрузить существующие онтологии и повторно использовать их, чтобы создать свою собственную онтологическую схему. Есть возможности создавать в среде PoolParty Semantic Suite свои собственные классы, подклассы, типы отношений, такие как симметричные, обратные или направленные отношения, и многое другое (например, атрибуты: integer, boolean, float, date и т.п.). Программное обеспечение для управления тезаурусами и таксономиями в The PoolParty Taxonomy & Thesaurus Manager - это инструмент отличного уровня для создания и поддержки информационной архитектуры.

Менеджер пула PoolParty позволяет начинать работу с минимального обучения. Эксперты по предмету могут моделировать свои области знаний без поддержки ИТ. Программное обеспечение для управления таксономией PoolParty применяет SKOS (Simple Organization Organization System) и SKOS-XL. Стандарты W3C, обеспечивающие совместимость вашего тезауруса с другими таксономиями, графами знаний и API.

PoolParty позволяет импортировать существующие таксономии и тезаурусы (например, из Excel или XML). Вы также можете экспортировать их в другой стандартный формат. Как происходит сбор данных разных форматов под конкретную задачу показано на рисунке 20.

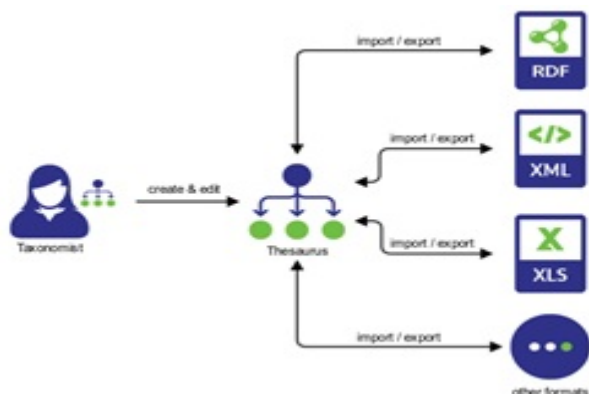


Рис. 20. Сбор данных разных форматов под конкретную задачу в PoolParty Semantic Suite (источник - Semantic Web Company)

Управление связанными данными через семантические и онтологические веб-технологии в PoolParty полагаются на взаимосвязанные данные, которое дает решение для разработки семантического слоя, чтобы привести в порядок цифровые активы. Связываясь с репозиториями данных, вы избегаете проектов миграции с интенсивным использованием ресурсов. Благодаря взаимосвязанным цифровым активам вы улучшаете свои приложения для поиска и персонализации. PoolParty Semantic Suite - простая в использовании и обучающая система. Чем больше вы его используете, тем лучше будет его алгоритм, так как это простое следствие формализованных через онтологии знаний. В PoolParty вы создаете граф знаний, который структурирует и представляет ваши приоритетные области знаний. С помощью Linked Data Management вы соединяете цифровые ресурсы по графу знаний. Активы могут быть связаны несколькими способами. В каждом сценарии связывания они будут иметь уникальный URI (= Uniform Resource Identifier). У вас есть инструменты, которые помогут вам в создании своей части онтологической (семантической) сети автоматически (рисунок 21).

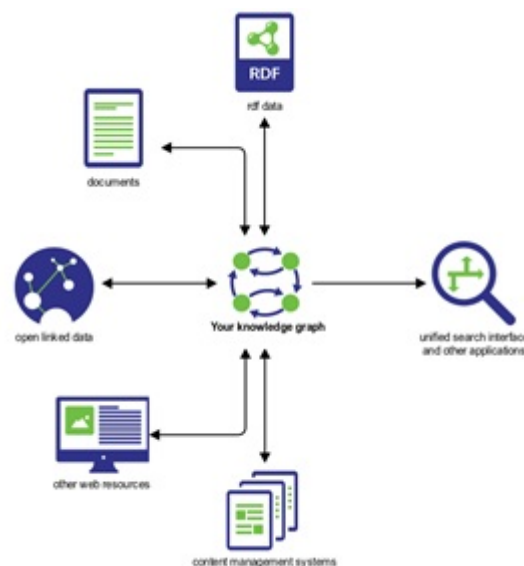


Рис. 21. Создании своей части онтологической (семантической) сети автоматически в PoolParty Semantic Suite (источник - Semantic Web Company)

VI. НОВЫЕ ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ КОНТЕНТА, КОТОРЫЕ ДАЕТ ОНТОЛОГИЯ

Семантически и онтологически расширенные тексты открывают новую эру использования контента. Используя текстовую аналитику и Web2.0 технологии, огромное количество контента, разбросанного в различных формах в документах разных форматов, преобразуются в обогащенные, преобразованные и управляемые части информации. Деятельность, связанная с просеиванием через крупные организации порождающие текстовые данные (например, правовые акты, документация предприятия, научные исследования) теперь можно масштабировать и автоматизировать (насколько это приемлемо). Алгоритмы вводят процессы управления рисками, обнаружения мошенничества, извлечения фактов и статистики, изучения связей, поддержание соблюдения стандартов, отслеживанием поведения потребителей и многие другие. Расширение сферы деятельности и обрабатываемые компьютером текстовые источники позволяют эффективно определять тенденции, выявлять закономерности, определять отношения.

В течение последних лет, открытые связанные данные (Linked Open Data - LOD) выросли всего от нескольких десятков, до более чем 10 000 баз знаний. Эти знания в основном поступают из разных областей, включая (но не ограничиваясь) публикации наук о жизни, социальных сетей, правительств, средств массовой информации, лингвистики. Связанные открытые данные и создание знаний из взаимосвязанных данных давно находятся в центре интересов ЕС. Результаты проекта LOD2 ЕС, состоявшего из нескольких проектов, были опубликованы еще в 2014 году [20].

Кроме того, что было создано в ЕС, облако LOD также содержит большое количество междоменных

(crossdomain) баз знаний, такие как DBpedia и Yago. Эти базы знаний обычно управляются децентрализованным образом и содержат частично перекрывающуюся информацию. Такой архитектурный выбор привело к тому, что знания, относящиеся к одному и тому же домену, были опубликованы как независимые объекты в облаке LOD. Например, информация о наркотиках можно найти в Diseasesome, а также в DBpedia и Drugbank.

Кроме того, были опубликованы некоторые базы знаний, такие как DBLP несколькими органами, что, в свою очередь, привело к дублированному контенту в LOD. Кроме того, большое количество гео-пространственной информации было доступно с ростом разнородной Сети данных в интернете.

Совместная публикация баз знаний, содержащих информацию, обещает стать феноменом все возрастающего значения с ростом числа независимых поставщиков данных. Обеспечение совместного использования баз знаний, публикуемых этими поставщиками для таких задач, как федеративные запросы, ответы на вопросы, связанные с кросс-онтологиями и интеграция данных чаще всего решаются путем создания связей между ресурсами, описанными в этих базовых знаниях.

В рамках этого тезиса мы отчетливо видим переход от изолированных баз знаний к обогащенным наборам связанных данных, где информация может быть легко интегрирована и обработана. Для достижения этой цели, как правило, предоставляются концепции, подходы и варианты использования, которые облегчают интеграцию и обогащение информации с другими типами данных, которые уже присутствуют в Сети связанных данных с акцентом, например, на осваиваемые сегодня в интернете гео-пространственные данные. Гео-пространственные данные имеют совсем иные форматы, чем тексты, но чрезвычайно важны в цифровой экономике.

Для онтологизации этих данных [21] вызовом, например, является отсутствие мер, которые используют географические данные для связывания гео-пространственных баз знаний. Частично это объясняется тем, что гео-пространственные ресурсы описываются как средства векторной геометрии. В частности, расхождения в детализации и измерение ошибок в базе знаний делают необходимым выбор соответствующих дистанционных мер измерений для гео-пространственных ресурсов.

Вторым вызовом, для этого типа данных является отсутствие автоматических средств Link Discovery (LD), способных работать с гео-пространственными базами знаний с отсутствующими и ошибочными данными.

Третья проблема, которую мы рассматриваем, заключается в отсутствии масштабируемых подходов LD для решения баз знаний больших гео-пространственных данных (их около 40% от объема всех больших данных).

Вместе с тем онтологизация гео-пространственных

ресурсов знаний - это очень знаковое направление.

Онтологическая (Семантическая) технология оказывает огромную помощь, когда она приходит к созданию, кураторству и использованию текстовых источников. Набор универсальных стандартов, разработанных и согласованных Консорциумом Всемирной паутины (W3C), а также смысловые технологии помогают предприятиям находить данные, вывести ссылки и извлечь знания из огромных наборов необработанных данных в различных форматах и из разных источников.

Использование семантических и онтологических технологий для текстовой аналитики выводит создание и доставку контента на новый уровень, где обнаружение информации выполняется гладко и без особых усилий. Организации разных размеров и доменов использовали этот подход для эффективного решения проблем объемов, разнообразия и несогласованности в документации и информации для того, чтобы представлять и создавать их системы управления контентом.

С помощью текстовой аналитики привязанные потоки фрагментированной, зашумленной и частично просматриваемой информации тщательно распутываются и вплетены в возможности интегрированного, последовательного и эффективного управления контентом. С помощью онтологий и текстовой аналитики, например, процессы управления знаниями и его обнаружения для предприятия упрощаются и тексты, развязанные от тирании неэффективности хранимого, управляемого и используемого (или, скорее, не используемого) контента. Обработанные и переплетающиеся в наших человеческих читаемых текстах, данные помогают правительствам и организациям знать контент (как внутренний, так и внешний) на необходимом уровне детализации и, следовательно, делать лучше в том, что они делают, быть больше и хорошо осведомленными для принятия решений, основанных на данных, и достаточно информированных для процветания в гиперсвязной среде.

Связанные и онтологизированные данные - это то, что должно быть для этого предусмотрено. Типичными приложениями онтологий и аналитики контекста в правительствах, организациях и на предприятиях ЕС сегодня являются:

- Бизнес-аналитика;
- Научное исследование;
- Классификация контента;
- Индивидуальные рекомендации;
- Анализ рисков;
- Обнаружение мошенничества;
- Анализ записей обслуживания клиентов;
- Соответствие требованиям;
- Соответствие нормам безопасности;
- Производство и доставка новостей.

Включение онтологий в данные в сочетании с привязкой этих данных к другим источникам добавляет ценность для контента и относительно малобюджетным способом. Методы онтологической аналитики позволяют представлять, записывать и извлекать информацию в общем формате и, таким образом, быть готовым к экономичному управлению и бесшовной интеграции.

Есть много отраслей, которые очень сильно выигрывают от онтологизации своих данных и добавления слоя семантик источникам. Среди областей, которые, в основном, выигрывают от достижений в области онтологической аналитики, находятся те, где знания в основном формальны:

Наука о жизни;
 Научные публикации;
 Издание СМИ и контента;
 Фармацевтика и здравоохранение;
 Банковское дело;
 Финансовые услуги и страхование;
 Правовые вопросы;
 Цифровые гуманитарные науки.

Примеры компаний, которые совершили прыжок в создании своего интеллектуального контента по информации компании Ontotext являются: FT, Bloomberg, Euromoney, JohnWiley & Sons, Oxford University Press, IET BBC, DK, AstraZeneca и другие.

Данные, извлеченные, например, из документов и связанные с выбранными фрагментами текстов, полезны как для создания, так и для повторного использования всех видов письменного контента. Применение методов текстовой аналитики для создания машиночитаемых текстовых источников существенно улучшает:

- Доступ к информации (посредством семантического поиска);
- Принятие решений (путем интеграции разрозненных и, казалось бы, несвязанных источников);
- Исследования и разработки (путем выявления скрытых отношений);
- Управление знаниями (путем объединения всей соответствующей информации);
- Обнаружение знаний (путем автоматического обнаружения ссылок на понятия и сущности);
- Создание и доставку контента (посредством связывания текста с большими данными).

Истина об успехах онтологической аналитике – это, по сути, то, что они опираются на официальные и формальные знания. Очень часто, люди занятые преследованием некоторого смысла как такового, пропускают основную точку в переплетении данных, и это делает информацию из источников управляемой посредством данных, а не смысла. Когда дело доходит до онтологической аналитики, самое важное понимание, которое не должно быть пропущено – это то, что этот

подход не является серебряной пулей для обнаружения знаний и дразнящего смысла из данных. Алгоритмы по-прежнему очень трудно понимают информацию в старом гуманитарном смысле так, как мы, человеческие читатели. Однако, со временем, электронные машины становятся лучше при заполнении четко определенных, измеримых, широко понятных задач, в которых интерпретация является вопросом вычисления. Вот почему превращение информационных источников в активы данных лучше всего применять в областях, где знание является явным и множественным, и двусмысленные интерпретации редки.

Онтологическая аналитика – это не что иное, как конкретный инструмент для принятия обоснованных решений на основе широкомасштабных информационных ресурсов который использует онтологическую (семантическую) технологию. Это так же просто и прямолинейно. В конце процесса появляется много и сложных наборов практик и процессов, лежащих в основе разработки и добычи полезных данных и ключевого процесса, называемого семантической аннотацией. Несколько таких примеров мы приведем далее.

VII. ОНТОЛОГИЗАЦИИ ДАННЫХ И ПРИЛОЖЕНИЙ В ЕС

В работе [13] было сказано что размерность и успешность онтологических проектов ЕС, по видимому, одна из самых серьезных в мире. В 2018 году вышло, как мы полагаем, несколько знаковых документов подтверждающих этот тезис. Так, был опубликован Open Data Goldbook [22] для менеджеров данных ЕС, который можно свободно загрузить с веб-сайта Европейского портала данных. Он публикуется под открытой лицензией. В этом документе написано: «Не стесняйтесь распространять этот Goldbook», так как политика ЕС в целом направлена на минимизацию действия авторских в тех областях, где их применение приводит, в целом, к экономическим потерям. Именно поэтому ЕС является заказчиком многих исследований и свободно распоряжается их результатами. Открытые данные для повторного их использования с целью получения дополнительных выгод являются основой европейской экономики данных. К этому примыкают: открытые правительства, открытая наука и открытое образование, усиливая экономические эффекты экономики данных

Различные участники, так называемые персоны, играют разные роли, когда речь идет о разработке и реализации инициативы «Открытые данные». Кроме того, не все знают, с чего начать и не имеют четкой картины того, какие аспекты необходимо решать. Вступают в игру разные роли. Возможно, придется написать политику, другой, возможно, придется разработать портал, а другой может собирать данные [22]. Для решения различных ролей, участвующих в Open Data, был разработан Open Gold Goldbook, в котором было представлено, что и как в ЕС можно получить из отологизированных данных.

Приведем три рисунка, поясняющие исключительную

важность такого онтологизированного портала для 508 миллионов жителей ЕС и соседних государств. Так на рисунке 22 показаны преимущества повторного использования открытых государственных данных ЕС в совершенно разных областях (рисунок 24). То, как в целом организован процесс работы с данными на портале можно увидеть на рисунке 23.



Рис. 22. Преимущества повторного использования открытых государственных данных [22]



Рис. 23. Визуализация процесса онтологизации или Extract, Transform, Publish [22]



Рис. 24. Пример того, что есть на европейском портале из <http://www.europeandataportal.eu/> (Европейский портал данных [22])

Развитие Сети в соответствии с семантической парадигмой онтологий, обычно называемое семантическим Web, является существенным предварительным условием для определения новых

сценариев будущего Интернета, направленных на создание среды взаимодействия для распределенных и вездесущих сетей, где программные агенты, устройства и люди могут эффективно взаимодействовать как производители или потребители ресурсов и услуг. Это одно самых быстро развиваемых направлений сегодняшнего развития ЕС.

Интернет, и в частности Web, в настоящее время является основным источником в ЕС Закона и Права, поэтому развитие семантической сети в юридической области - это перспектива, представляющая особый интерес для специализированного характера правовой информации, сложность законодательного документооборота и особенностей потребностей пользователя. Такое развитие предусмотрено для воздействия на всех игроков сценария правовой информации: государственных администраций, издателей, юристов ученых, а также граждан ЕС. Сегодня на 100 % онтологизированы новые документы законодательной и судебной сферы ЕС, включая прецедентное право.

В этом состоянии развития Публикационное бюро Европейского Союза (ОП) является ключевым игроком, имеющим двукратную роль: отдел государственных учреждений и юридического издателя. Именно ОП ведет европейский портал о котором мы говорили и только после публикации на этом портале документы и правовые решения приобретают юридическую силу. В этих двух ролях, о которых мы говорили выше, ОП направлено на обеспечение информационных услуг (в частности в юридической информации), способных гарантировать доступность, ремонтпригодность и повторное использование правовых информационных ресурсов. Эти услуги уже есть и будут больше в ближайшем будущем, сильно затронутой эволюцией Сети, которая будет все более характеризоваться как поставщик семантически квалифицированных фрагментов онтологизированной информации. Эта эволюция представляет собой проблемы, решаемые в стратегических целях, которые должны быть достигнуты ОП в следующем десятилетии.

Бюро публикаций Европейского союза (Бюро публикаций) является межведомственным бюро, которое публикует и распространяет публикации учреждений и других органов Европейского Союза (см. Решение 2009/496 / ЕС, Euratom).

Бюро публикаций публикует официальный журнал Европейского союза на 23 языках (24, когда требуется ирландское рисунок 25) и производит (или совместно производит) общие публикации и информационные материалы о инициативах ЕС, а также о деятельности учреждений и других органов Европейского Союз.

Кроме того, в Бюро публикаций предоставляется ряд онлайн-сервисов, предоставляющих бесплатный доступ к информации о:

Законах ЕС (EUR-Lex);
 Публикациях ЕС (книжный магазин ЕС);
 Открытых данных ЕС (Портал открытых данных ЕС);
 Контактных данных официальных лиц ЕС на
 руководящих должностях (EU Whoiswho);
 Исследованиях и разработках ЕС (CORDIS);
 Государственных закупок (TED);
 Стилистических правилах и условных обозначениях,
 такие как макет и типография (руководство по стилю);
 Многоязычном тезаурусе ЕС (EuroVoc);
 Данным, используемым институтами ЕС (реестр
 метаданных).

Собственно это бюро только что издало второй, как мы полагаем, выдающийся документ по обсуждаемой тематике - Cellar Семантический репозиторий [23]. Там решается большой вопрос о том, как получать онтологии и соответствующие для развития приложений в условиях работы с таким количеством языков (рисунок 25). Как при этом обеспечить автоматическое пополнение онтологизированными данными всех этих разноязычных приложений в боевом режиме и многое другое в этом руководстве.

ISO 639-3 code	Language
bul	Bulgarian
ces	Czech
dan	Danish
deu	German
ell	Modern Greek
eng	English
est	Estonian
fin	Finnish
fra	French
gle	Irish
hrv	Croatian
hun	Hungarian
isl	Icelandic
ita	Italian
lav	Latvian
lit	Lithuanian
mlt	Maltese
nld	Dutch
nor	Norwegian
pol	Polish
por	Portuguese
ron	Romanian, Moldavian, Moldovan
slk	Slovak
slv	Slovene
spa	Spanish, Castilian
swe	Swedish

Рис. 25. Перечень языков, на которых происходит цифровая публикация данных на портале ЕС (источник - [23])

После принятия решения о повторном использовании данных Бюро публикаций (ОП или ОР) внедрило Cellar, как семантическое (онтологическое) хранилище для включения его публикаций для использования онтологических разметок и моделей. В настоящее время EUR-Lex, EUBookshop и некоторые другие включены в Cellar.

Cellar основан на семантических технологиях: в

рамках нескольких стандартов для совместного использования и повторного использования данных. Они нормализуют именованные ресурсы в контролируемых словарях, которые позволяют компьютерам разговаривать и связываться друг с другом.

В [23] представлена информация и примеры того, как обращаться к файлам контента и метаданным из Cellar, хранилища цифрового распространения ОП. Несмотря на то, что он открыт для всех граждан, этот документ направлен главным образом на участников и компании, которые хотят автоматизировать их доступ к публикациям ОП, то есть создать свои онтологические приложения. «Если Вы заинтересованы в обычном, но не автоматизированном доступе, вы можете найти лучше, более актуальным способом информацию на своих порталах [23]». То есть, Cellar предназначен для создателей приложений и половина этого документа просто написана на формализованных онтологических языках.

Вот цитата из [23], где мы поменяли только номер рисунка: «Семантическая сеть - это эволюция Всемирной паутины, которая позволяет автоматизировать задачи в Интернете, разрешая компьютерам разговаривать друг с другом. Для этого данные должны соответствовать общим форматам, структурам и стандартам, известным всем участвующим системам. Эта группа стандартов называется семантической сетью»

Таблица 1: Примеры троек в Cellar [23]

Subject	Property	Object
oj:30A_1952_001_R ()	oda:work_date_document	1952-12-30
oj:30A_1952_001_R	rdftype	oda:official_journal
colux:31952#0003	oda:resource_legal_published_in_official_journal	oj:30A_1952_001_R
colux:31952#0003	owl:nameAn	uri:doc/1952/3/oj

Работа с Cellar основана на формальном описании семантических (онтологических) троек (таблица 1) и свойствах или предикатах, это глагол, объясняющий, как объект связан с объектом. Свойство всегда является URI, и свойство также может быть объектом в тройке. Объектом в Cellar может быть либо объектом (выраженным с помощью URI), либо буквенным значением с его тип данных и объекты с URI также могут быть объектами в тройке.

Онтология в [23] трактуется как формальная модель, позволяющая представлять знания о конкретном домене. Онтология описывает типы вещей, которые существуют (классы), отношения между ними (свойства) и логическими способами этих классов и свойств могут использоваться все вместе (аксиомы).

RDF в [23] это структура описания ресурсов и семейство международных стандартов для обмена данными в Интернете так как RDF основан на идее определения вещей, использующих идентификаторы веб-интерфейса или URI HTTP и описание ресурсов с

точки зрения простых свойств и значения свойств. Это модель, в которой тройки (субъект, предикат и объект) кодируются. Некоторые словари и онтологии предопределены, например, RDFS, OWL или SKOS, все они используются в ОП. Внешне это выглядит как расширенный язык разметки (XML).

Архитектура Cellar с деталями в интерфейсах так же выглядит необычно (рисунок 26), как набор онтологий и словарей, но, пожалуй, стоит к таким архитектурам привыкать.



Рис. 26. Архитектура Cellar с деталями в интерфейсах [23]

Существует несколько уровней, на которых возможен доступ; см. таблицу 2 для высокого уровня классификации.

Таблица 2. Доступ к уведомлениям, содержанию и метаданным Cellar [23]

Table 4: Access possibilities to Cellar notifications, content and metadata

	RSS notifications	Metadata	Metadata and content
EUR-Lex (meant for humans)	Custom	Web services	EUR-Lex website
Cellar (meant for machines)	Full	Sparql endpoint	RESTful interface

EUR-Lex и OP Portal являются интерфейсами для хранилища Cellar. EUR-LexRSS может быть полезен, когда вам нужна сводка того, что публикуется. В нем есть несколько категорий на выбор. Спецификации см. На веб-странице EUR-Lex RSS (<http://eur-lex.europa.eu/predefined-rss.html>).

В самом начале этой статьи мы говорил о трудностях описания формальных онтологий, которые, по сути, являются частью математики, но мы попробуем дать читателю представление о метаданных ЕС. Метаданные предоставляют информацию о данных и, следовательно, обеспечивают доступ к цифровым объектам. Стандарты на них поэтому являются существенными так как многое в онтологиях происходит в контексте метаданных, поскольку они

обеспечивают общую структуру для «данных о данных» и помогают обеспечить совместимость систем и улучшить обнаружение и доступ к данным. Вот наиболее известные стандарты метаданных и онтологий, используемые в контексте Open Data и Cellar:

- Общий европейский информационный формат исследований (CERIF) - рекомендация ЕС государствам-членам и используется для всей исследовательской информации.
- Dublin Core (DC) - используется для описания веб-страниц.
- Словарь данных категории (DCAT) - используется для наборов данных в Интернете, также на основе DC.
- Всеобъемлющая сеть архивов знаний (CKAN) - используется на государственных сайтах открытых данных.
- Стандарт метаданных eGov (eGMS) - на основе DC.
- Inspire - используется для наборов данных с геопространственными координатами.

VIII. РЕКОМЕНДАЦИИ ПО МЕТАДАНЫМ

Предоставление качественных метаданных представляет собой сложную, но необходимую практику. Фонд W3C разработал руководящие принципы и передовые методы поддержки держателей данных. Более того, взаимодействие с Европейским порталом данных имеет решающее значение. Это позволяет избежать дорогостоящих пешеходных переходов и сопоставлений между наборами данных. Следовательно, настоятельно рекомендуется использовать DCAT-AP. Чтобы обобщить, опубликуйте метаданные с данными, используя машиночитаемый формат и стандартные термины для определения метаданных. Кроме того, опишите общие характеристики набора данных с информацией о локальных параметрах, лицензии, происхождении и качестве. Рекомендации Open Data и Cellar: «всегда публикуйте свои метаданные как Связанные данные. Это увеличивает открытость и совместимость ваших наборов данных». Метаданные также позволяют подготовить необходимое качество информации через уже развитую систему цифровых онтологических проверок (рисунок 27).

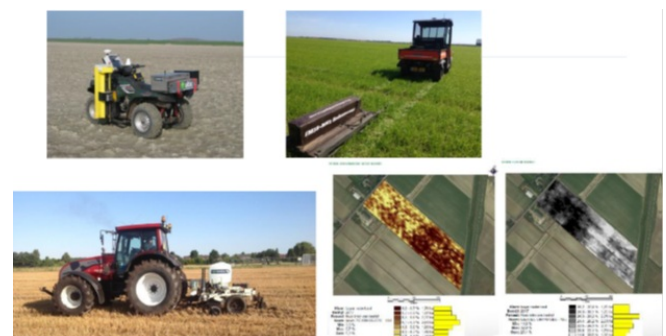


Рис. 27. Контрольный список набора данных [22]

Руководство по измерению [25] помогает правительствам, гражданскому обществу и исследователям понять, как оценивать деятельность открытых данных на основе принципов Устава открытых данных («Устава»). Он стремится пролить свет на часто непрозрачный и жаргонный мир открытого измерения данных.

Руководство по измерению - это анализ принципов Устава и их оценка, основанные на существующих открытых правительственных инструментах измерения данных - с уделением особого внимания обязательствам, которые можно измерить, обязательствам, которые невозможно измерить, и существующим пробелам (например, обязательства которые не были измерены).

Руководство по измерению сделано для правительств, гражданского общества и исследователей для понимания как можно измерить принципы Устава. Он обеспечивает анализ показателей, который включает в себя всеобъемлющие таблицы глобальных показателей (например, таблиц индикаторов) за каждый принцип устава:

- Для правительств руководство направляет наиболее важные идеи в этом разделе, в виде Управляющее резюме.
- Для гражданского общества и коммуникаторов таблицы индикаторов и наш анализ обеспечивают прозрачность о существующих инструментах измерения («Пять открытых инструментов оценки данных») и они измеряют. Это может помочь гражданскому обществу контролировать ход политики открытых данных на страны.
- Для исследователей руководство объясняет методологию сопоставления открытых индикаторов данных с Уставными обязательствами. Созданные таблицы индикаторов могут использоваться для сравнения существующих данных инструментов измерения и разработки новых показателей.

Руководство по измерению содержит информацию от экспертов по открытым данным и членов организаций, которые работают на открытых инструментах измерения данных. Анализ охвата пяти ведущие инструменты для измерения открытых данных - открытый барометр данных (ODB), глобальные данные Open Data Index (GODI), Open Data Inventory (ODIN), Open Полезные повторно используемые данные правительства (OURdata), и European Open Data Maturity Assessment (EODMA) - показывает, что только части основных принципов Устава и их компонентов; или что некоторые обязательства могут быть измерены в будущем. Однако некоторые концепции Устава слишком широкие (например, «высококачественные данные», «удобство использования широким кругом пользователей») или отсутствие общих, что затрудняет поиск общего индикатора.

Руководство по измерению также описывает, как существующие индикаторы метрируют ключевые концепции открытых данных. Важно отметить, что не

все аспекты обязательства четко определены. Несколько способов в настоящее время существуют определенные обязательства. Необходимо определить некоторые обязательства и измеряется по каждой стране для включения местного контекста.

Системы и практическая работа ЕС в части LOD опирается на хартию открытых данных - это сотрудничество правительств и экспертов, созданных в 2015 году с целью открытия данных. Вместе с гражданским обществом и экспертами по всему миру они установили принципы публикации данных. К этому движению присоединилось более 70 правительств и организаций. Шесть принципов есть в этой хартии:

- Открыть по умолчанию;
- Своевременный и всеобъемлющий;
- Доступный и удобный;
- Сопоставимые и совместимые;
- Для улучшения управления и участия граждан;
- Для инклюзивного развития и инноваций.

В Руководстве по оценке устаревших данных дается обзор существующих инструментов измерения для каждого принципа Устава данных. Это дает возможность сравнить различные индикаторы, которые используются в пяти крупнейших инициативах по измерению открытых данных - Европейской оценки данных о стоимости данных в Европе, Барометре открытых данных Web Foundation, Глобальном индексе открытых данных Open Knowledge International, Open Data Watch's Open Data Inventory и OURData Index OECD.

Портал открытых данных Европейского союза (EU ODP) обеспечивает доступ к расширенному спектру данных из институтов Европейского союза (ЕС) и других органов ЕС. Можно свободно использовать и повторно использовать эти данные в коммерческих или некоммерческих целях. Предоставляя легкий доступ к данным, ЕС стремится помочь внедрить их в инновационное использование и разблокировать их экономический потенциал. Портал также предназначен для того, чтобы сделать институты ЕС и другие органы более открытыми и подотчетными. Всем учреждениям ЕС предлагается публиковать свои данные по мере возможности. Это означает, что данные могут быть повторно использованы бесплатно и без каких-либо ограничений авторского права.

IX. ЗАКЛЮЧЕНИЕ

В обществах XXI века информационная грамотность, доступ и использование знаний становятся предпосылкой для того, чтобы люди активно участвовали в социальной, экономической, культурной и политической жизни. Информационная грамотность (Information literacy -IL) сегодня рассматривается не как специализированный набор навыков, должна считаться

фундаментальной компетенцией, такой, как способность читать, писать и вычислять. Информационная грамотность стала необходимой для участия в жизни общества во всем мире, поэтому ЮНЕСКО даже считает это «основным правом человека» [26].

Эти компетенции (ИЛ) становятся все более необходимыми в условиях быстрого развития цифровой экономики. Вот сценарий [27] 2040 года: «Открытая наука, открытые инновации и открытое образование преобладают в Европе и во многих частях мира. В основе новой системы знаний Европы лежат кластеры хорошо финансируемых, всемирно известных (высших) учебных заведений и общественных исследовательских организаций в тесном партнерстве с широким кругом участников на местном, региональном, национальном и европейском уровнях. Высокий уровень государственных инвестиций создал множество открытых цифровых инфраструктур для обмена информацией, данными, знаниями и опытом в разных странах.

Инновации в промышленности являются частью процесса совместного создания знаний, связанных с различными глобальными и местными «сообществами практики», например, вопросами устойчивого развития. Глобальные и профессионально управляемые человеческие рабочие облака составляют 40% и более всех команд в области науки и бизнеса. Те, которые развивались в Европе, включают в себя европейские ценности уже в своем дизайне, создавая рабочую среду, которая обеспечивает совместное и открытое производство знаний. Эксперты, новаторы, университеты, промышленность и многие новые участники - НПО, фонды, люди всех возрастов - работают вместе в быстро меняющихся сетях для решения местных, региональных и глобальных задач.

Использование вездесущей оцифровки и наукоемкой информации во всех аспектах жизни в Европе поддерживается всесторонней грамотностью данных. Начальное и среднее образование и непрерывное обучение были полностью преобразованы и внесли важный вклад в создание общества знаний.

Инновации более чем когда-либо зависят от способности школ и университетов наделять студентов необходимыми цифровыми, интеллектуальными и поведенческими навыками, необходимыми для использования возможностей, предлагаемых новой системой знаний. Этот процесс нужно начинать рано. Научная коммуникация вместе с инструментами моделирования, виртуальной реальностью и расширенной реальностью позволяет ученикам начальной школы стать высококвалифицированными разработчиками сложных проблем реального мира.

Общие технологии, такие как искусственный интеллект и распределенная книга (блокчейн), обеспечивают новую инфраструктуру для представления, доступа и использования данных, информации и знаний. Эти инфраструктуры были

сформированы за счет государственных инвестиций в исследования и инновации с самого начала и, таким образом, разработаны в соответствии с принципами открытости и прозрачности. Это усиливает их авторитет и признание в обществе в качестве инструментов для решения даже этических противоречий.

На социальном уровне конфликты и этические споры стали более распространенными, поскольку технологии, основанные на знаниях, расширяются до областей человеческого существования и деятельности, которые ранее считались «неприкосновенными». Внедрены новые процедуры преодоления этических и политических дилемм. Широкое участие в процессах принятия решений и принятия решений облегчается благодаря инфраструктуре цифровых знаний.

Открытая, совместная система знаний обещает быть важным вкладом в европейское общество и экономику. Для промышленности и предпринимательства интегрированная система знаний открывает новые возможности для диверсификации и ускорения инноваций. Благодаря знаниям, связанным с инновациями, доступным для всех участников, предпринимательство станет для многих вариантом.

Открытая система знаний создаст важные источники экономических доходов посредством услуг, связанных со знаниями. Это также повысит эффективность использования ресурсов в производстве, потреблении и администрировании и позволит децентрализовать циркулярную (круговую) экономику, где добавленная стоимость останется в Европе.

Улучшение управления знаниями и совместной работы является приоритетом для преодоления менталитета силосов и подключения синергизма между портфелями [28]. В Европе взят курс на трансформацию экономики данных в экономику знаний и сегодня это уже приводит к следующему:

- 1) внедрение горизонтального уровня управления знаниями в организационной структуре, мобилизации научных компетенций из разных директив вокруг целей политики Комиссии ЕС;
- 2) отстаивание внедрения новых методов и платформ сотрудничества, а также разработка программы профессиональной поддержки управления знаниями;
- 3) организации и университеты ЕС начинают превращаться из традиционных исследовательских организаций в ведущего мирового лидера в области онтологизированных знаний для их многократного применения, как в виде совместимых онтологий, так и онтологизированной информации.

Огромную роль в этом процессе играет наука [29,34], однако сам процесс онтологизации влияет практически на все аспекты сегодняшней и конечно завтрашней жизни [30-33]. Следуя объявленным выше принципам наглядности, мы попробовали представить это процесс на рисунке 28.



Рис. 28. В информационном потоке сегодняшнего дня роль JRC в управлении знаниями - это поддержка разработки политики ЕС, которая может быть фундаментальной (источник – [28]).

В работе [13] было отмечено, что масштабные онтологические преобразования в ЕС начались с онтологии права. Лон Л. Фуллер выделяет «восемь способов потерпеть неудачу в попытке создания закона; им соответствуют восемь видов юридического совершенства, к которым может стремиться система норм» [34]. В действительности, и способов больше, и, что более важно, много больше детерминантов высокого нормативного качества.

В условиях экспоненциального роста сложности нормативно-правового пространства, как в ЕС, так и в России, возникает необходимость поиска новых автоматизированных способов управления нормативными массами, массивами, правовыми отраслями (классическими и комплексными), подотраслями, институтами (отраслевыми и межотраслевыми) и субинститутами, нормативными горизонтами, нормативно-правовыми векторами (развивающимися направлениями правового регулирования), нормативно-правовыми режимами.

Современные процессы упрощения законодательства[35], омологирования нормативно-правового дизайна[36], редуцирования энтропии в законодательстве наталкиваются на то, что нормативная правовая «вселенная» всё более превращается в классическую сложную открытую динамическую систему[37], которая все больше охватывает цифровую экономику.

И на сегодня это пока один из сложнейших вопросов: возможна ли релевантная машинная формализация права без явного ущерба для эффективности права.

Сегодня термин «онтология» (применительно к сфере права) интерпретируется и объясняется следующими способами:

1) «онтология – технология (и результат её задействования) всеобъемлющей и подробной формализации некоторой области знаний с помощью концептуальной схемы», в том числе, «посредством задействования формализованных онтологических языков и инструментов, которые уже стали мировыми стандартами»[7];

2) онтология – средство конструирования и/или репрезентации правовой реальности и правового

универсума[38];

3) онтология – аппроксимированная к условно-идеальному правовая форма;

4) «правовая онтология – учение о мире права и формах его существования»[39];

5) специфические формы (или конкретная форма) правовых норм, правовых феноменов, правовых процессов и правоотношений, фреймированные определёнными нормативными правовыми, иными регламентными[40,41], деонтологическими (ценностно-нормативными) и другими порядками [42,43].

Именно текст правовой нормы и (на несколько более высоком уровне) текст нормативного акта выступают в числе основных онтологических единиц в онтологии права.

Основанная на чётком атрибутировании топологизация сегментов, иерархизаций и интерреляций внутри массивов нормативного правового регулирования и технического нормативного регулирования, маршрутизация изменений в праве – всё это (при правильном задействовании) позволяет выстроить систему управления онтологиями в праве и без этого трудно рассчитывать на успех цифровых преобразований в России.

БИБЛИОГРАФИЯ

- [1] Соловьев А. И., Куприяновский В. П., Соловьев С. А. Single digital market of the European Union: current state and development trends //International Journal of Open Information Technologies. – 2017. – Т. 5. – №. 10.-С.47-54
- [2] И.А. Соколов и др. Искусственный интеллект как стратегический инструмент экономического развития страны и совершенствования ее государственного управления. Часть 1. Опыт Великобритании и США //International Journal of Open Information Technologies 2017 – Т. 5.- №. 9.-С. 57- 75.
- [3] И.А. Соколов и др., Искусственный интеллект как стратегический инструмент экономического развития страны и совершенствования ее государственного управления. Часть 2. Перспективы применения искусственного интеллекта в России для государственного управления //International Journal of Open Information Technologies 2017 – Т. 5.- №. 9.-С. 76- 101.
- [4] Соколов И. А. и др. Государство, инновации, наука и таланты в измерении цифровой экономики (на примере Великобритании) //International Journal of Open Information Technologies. – 2017. – Т. 5. – №. 6.
- [5] Drozhzhinov V. et al. On strategic approach to the formation of the US digital government //International Journal of Open Information Technologies. – 2017. – Т. 5. – №. 4. – С. 29-54.
- [6] Куприяновский В. П. и др. Технологии трансграничных цифровых сервисов в ЕС, формализованные онтологии и блокчейн //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 7.
- [7] Volokitin Y. et al. On problems of the digital economy and formalized ontologies //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 6. – С. 87-96.
- [8] Sokolov I. et al. The digital economy of Western Australia-smart mining, oil, gas enterprises, railways, seaports, and formalized ontologies //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 6. – С. 44-62.
- [9] Kupriyanovsky V. et al. Formalized ontologies and services for high-speed and digital railways //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 6. – С. 69-86.
- [10] Sokolov I. et al. Modern EU research projects and the digital security ontology of Europe //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 4. – С. 72-79.
- [11] Kupriyanovsky V. et al. Semantics, metadata and ontologies in smart city applications-new BSI standards //International Journal of Open Information Technologies. – 2017. – Т. 5. – №. 6. – С. 94-108.

- [12] Klimov A. et al. BIM and engineering formalized ontologies on the European digital railway in the EULYNX-data economy //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 8. – С. 38-65.
- [13] Kupriyanovsky V. et al. On the effects of formalized ontologies in the data economy-the EU experience //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 8. – С. 66-78.
- [14] Handbook on European data protection law .2018 edition. © European Union Agency for Fundamental Rights and Council of Europe, 2018
- [15] Interoperability and fundamental rights implications Opinion of the European Union Agency for Fundamental Rights Vienna, 11 April 2018. © European Union Agency for Fundamental Rights and Council of Europe, 2018
- [16] Hate crime recording and data collection practice across the EU © European Union Agency for Fundamental Rights and Council of Europe, 2018
- [17] Under watchful eyes: biometrics, EU IT systems and fundamental rights EU © European Union Agency for Fundamental Rights and Council of Europe, 2018
- [18] MULTIANNAL EUROPEAN E-JUSTICE ACTION PLAN 2014-2018. NOTICES FROM EUROPEAN UNION INSTITUTIONS, BODIES, OFFICES AND AGENCIES ,COUNCIL,(2014/C 182/02)
- [19] General Data Protection Reform (GDPR) http://ec.europa.eu/justice/data-protection/reform/index_en.htm
- [20] Sören Auer, Volha Bryl, Sebastian Tramp (Eds.) Linked Open Data – Creating Knowledge Out of Interlinked Data. Results of the LOD2 Project © The Editor(s) (if applicable) and the Author(s) Springer 2014
- [21] Mohamed Ahmed Mohamed Sherif, Automating Geospatial RDF Dataset Integration and Enrichment , dissertation, Universität Leipzig, Fakultät für Mathematik und Informatik , 2016
- [22] Open Data Goldbook for Data Managers and Data Holders. Practical guidebook for organizations wanting to publish Open Data. EU ,Last update: January 2018
- [23] Cellar The semantic repository of the Publications Office Manuscript completed in June 2018. © European Union, 2018
- [24] ADVANCING IoT PLATFORMS INTEROPERABILITY ©2018 River Publishers
- [25] Open Data Charter Measurement Guide. Open Data Charter, Open Knowledge International, World Wide Web Foundation MAY 2018
- [26] Deliverable 2.2: Updated curricula and prototypes for adaptive training support and introductory MOVING MOOC for community building © MOVING Consortium, 2018
- [27] Towards a New Knowledge System Targeted scenario N°19 Glimpses of the future from the BOHEMIA study © European Union, 2018.
- [28] Knowledge Management for Policy .© European Union 2017 [34] Putting science at the heart of European policymaking © European Union, 2018
- [29] Brussels, 18.10.2016 C(2016) 6626 final COMMUNICATION TO THE COMMISSION Data, Information and Knowledge Management at the European Commission {SWD(2016) 333 final}
- [30] Куприяновский В. П. и др. Information technology in the university system, science and innovation of the digital economy on the example of the UK //International Journal of Open Information Technologies. – 2016. – Т. 4. – №. 4. – С. 30-39.
- [31] Kupriyanovsky V. et al. On decision-making in the digital economy. UK Experience //International Journal of Open Information Technologies. – 2017. – Т. 5. – №. 4. – С. 63-73.
- [32] Ярцев Д. И. и др. Экономика стандартизации в цифровую эпоху и информационно-коммуникационные технологии на примере Британского института стандартов //International Journal of Open Information Technologies. – 2016. – Т. 4. – №. 6.
- [33] Sokolov I. et al. State, Innovation, Science and Talents in Measuring the Digital Economy (UK Case Study) //International Journal of Open Information Technologies. – 2017. – Т. 5. – №. 6. – С. 33-48.
- [34] Фуллер Л.Л. Мораль права: Пер.с англ. – М.: ИРИСЭН, 2007. – 308 с. – С. 56
- [35] Понкин И.В. Упрощение законодательства как инструмент «новой» модели публичного управления // Административное право и процесс. – 2014. – № 4. – С. 8-12
- [36] Барциц И.Н. Конституционный дизайн: Образ государства и образ эпохи. – М.: Дело, 2018. – 60 с.
- [37] Понкин И.В. Энтропия, неэнтропия и порядок в публичном управлении и в праве // Право и образование. – 2016. – № 9. – С. 11–20.
- [38] Понкин И.В. К вопросу об определении понятия «юридическое пространство» в контексте публичного управления // Власть. – 2014. – № 1. – С. 105–108
- [39] Гаджиев Г.А. Онтология права: Критическое исследование юридического концепта действительности. – М.: Норма – ИНФРА-М, 2013. – 320 с. – С. 123, 278
- [40] Понкин И.В. Автономный внеправовой нормативный порядок в сфере религии и защита государством религиозных чувств и достоинства личности верующих // Религия и право. – 2014. – № 3. – С. 8–11.
- [41] Понкин И.В., Понкина А.И. К вопросу о понятии и особенностях автономного внеправового нормативного порядка в области спорта // Вестник Пермского
- [42] Понкин И.В. Теория государственного управления: содержание понятия «порядок» // Административное право и процесс. – 2016. – № 11. – С. 8–10.
- [43] Понкин И.В. Энтропия, неэнтропия и порядок в публичном управлении и в праве // Право и образование. – 2016. – № 9. – С. 11–20. Университета. Сер. «Юридические науки». – 2016. – № 1. – С. 28–34

The ontologization of European Union data as a transition from a data economy to a knowledge economy

Oleg Grinko, Vasily Kupriyanovsky, Oleg Pokusaev, Yuri Volokitin, Igor Ponkin, Dmitry Namiot, Alena Redkina

Abstract— This article continues the series of papers on the use of formal ontologies. Today, in the overwhelming majority of cases, much of the world's information exists in textual form: business documents, etc. It was in this historical form that a person kept his knowledge lived and worked. Today, text documents are digitized, retaining the same appearance. Such a rapid increase in the number of digital texts causes an increase in the need for text analytics and raises the question of finding “smart” ways to read and understand texts, and, ultimately, the question of obtaining knowledge from them. The paper deals with the semantic web, RDF and SPARQL. It presents a detailed discussion of the technology platform PoolParty Semantic Suite. Semantically and ontologically extended texts open a new era of content use. Using text analytics and Web 2.0 technologies, a huge amount of content scattered in various forms in documents of different formats is converted into enriched, transformed and managed information fragments. Thus, machines are not only able to process long computational strings of characters and index large amounts of data, but they can also store, manage, and retrieve information based on their meaning and logical relationships between things in knowledge. Ontologies and semantics add another layer to the new Internet and can show related things, facts and objects instead of just matching words.

Keywords— ontology, knowledge, RDF.