

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 4

Москва 2012

ОБЩИЙ РАЗДЕЛ

УДК 1 : [001.102 : 004]

Х. Р. Алошти (Иран)

Философский взгляд на информацию и информационную технологию*

Рассматриваются различные подходы к определению понятия «информация»: математико-физический, семантический, прагматический, герменевтико-экзистенциальный и подход с точки зрения англетики. Отмечается актуальность изучения истории философии с точки зрения информационной технологии (ИТ). Излагаются два противоположных взгляда на информационную технологию – инженерный и антропоцентрический. Кратко рассматриваются перспективные направления философии ИТ.

Ключевые слова: информация, информационная технология, философия, метафизика, герменевтика, англетика

Век, в котором мы живем, назван веком информации, и разговоры об информационном обществе, электронном обществе, сетевом обществе (обществе, охваченном сетью) и других похожих терминах можно услышать везде. Дискуссии, связанные с информационными технологиями и информационным обществом, происходят повсюду: от научных кружков и собраний до средств массовой информации, от научно-исследовательских журналов до журналов общей тематики. И действительно, не имея правильного представления о соответствующих явлениях,

понятии информации и их истинного понимания, как можно справиться с решением экономических, управленческих, политических, культурных и общественных проблем в современном мире?

Философия связана со временем, и если связь философии и времени будет разорвана, тогда её законы и предикаты становятся бездейственными и не будет источника воздействий. Структура информационной философии – это знак того, что философия жива и в ней рассматриваются сущность и значение информации, её динамика и вопросы, которые провоцируются информационными и коммуникационными технологиями. Данная статья состоит из двух разделов: пер-

* Перевод с фарси Абдувосида Юлдошева

вый раздел содержит обсуждение определения и сущности информации, второй посвящен философии информационной технологии.

1. ФИЛОСОФИЯ ИНФОРМАЦИИ

Ниже представлены некоторые распространенные подходы к определению понятия «информация», в том числе теории К. Шеннона, Р. Карнапа, Ф. Дретске, М. Бакланда и Р. Капулло.

1. Математико-физический взгляд на информацию

Математико-физический взгляд на информацию, по существу, возник благодаря работам инженеров-электронщиков, математиков, теоретиков систем, теоретиков в сфере кибернетики (Cybernetics), физиков и биологов. Специалисты каждой из перечисленных областей в распространении математико-физического взгляда на информацию играли роль, соответствующую своей научной области. В специализированных источниках этих областей термин «теория информации» («Theory of information») означает «математическую теорию информации», или «математическую теорию коммуникации», или «математическую теорию передачи данных», или, по мнению Уоррена Вивера, «математическую теорию инженерных аспектов коммуникаций» («The mathematical theory of the engineering aspects of communication») [1, с. 96]. Тот факт, что некоторые исследователи понимают под термином «теория информации» только математические аспекты передачи информации, стал причиной возникновения недоразумения.

Юджин Чиу и его коллеги Клод Шеннон, Норберт Винер, Уоррен Мак-Калок, Уолтер Питтс, Алан Матисон Тьюринг и Д. фон Ньютман считаются пионерами движения, которое было сформировано после Второй мировой войны и занималось объединением теорий, связанных со свойствами информации и ее обработкой и принадлежащих разным областям. Этих специалистов из разных областей связывал только один фактор – их знание о математической логике.

Математическая логика состоит в изучении законов абстрактного мышления. Можно сказать, что способность изложения одного процесса и его абстрагирования на высшем уровне помогла Шеннону сформулировать суть процесса коммуникаций при анализе информации относительно семантических измерений [2]. Эта абстракция была важным событием в электронной инженерии и дистанционной коммуникации, поскольку до 1928 г. коммуникации имели только технические аспекты и не были связаны с научными теориями. Объекты исследований коммуникаций с точки зрения типов средств и различия в способах передачи сигналов отличались друг от друга.

До 1940 г. были изобретены следующие средства коммуникаций: телеграф (1830), телефон (1870), беспроводной телеграф (1890), радио (1900), телевидение (1930), телетайп (1930). Различие между этими средствами в первую очередь заключалось в спосо-

бах передачи сигналов из одной точки в другую. Однако исследователи воспринимали звук и изображение в этих средствах как две разные вещи и не признавали для них ничего общего, а объединяющее понятие «информация» в этом контексте ещё не появилось. Соответственно, каждое из перечисленных средств коммуникаций представлялось уникальным и независимым от других, и невозможно было представить их объединение в одну форму и структуру. Работа, проведенная Шенноном, заключалась в том, что он сформулировал понятие информации [3, с. 28] и разработал абстрактное основание в деле передачи кодов и сигналов [2].

В 1924 г. Неквист опубликовал статью, в которой изложил способ передачи письменных кодов посредством одного телеграфного канала с максимально возможной скоростью и без изменения формы [3, с. 29]. Естественно, в то время Неквист не говорил об информации, но сообщил о скорости передачи сообщения (Speed of transmission of intelligence). По мнению Шеннона, понятие сообщения похоже на понятие информации.

Действительно, Неквист положил начало пониманию необходимости отделить содержание сигнала от информации, заключенной в сообщении. Он предложил логарифмическое правило, определяющее максимальный объем сообщения, который может быть передан по одной телеграфной линии [2]:

$$W = K \log m,$$

где W – скорость передачи сообщения, K – константа, m – количество существующих значений (Current values).

Работу Неквиста развивал Р. Хартли. Он стремился создать некую теорию, охватывающую основные средства его времени. И он пришел к выводу, что между значением и информацией в одном сообщении есть разница. При этом Хартли интересовали только возможности одной системы при передаче последовательностей символов [4]. Для объяснения количества передаваемой информации он обобщил логарифмическое правило Неквиста:

$$H = \log s^n,$$

где H – количество передаваемой информации, s – количество возможных сообщений, n – количество символов в одной передаче.

Для определения количества информации Хартли действовал так: он предполагал, что каждый символ сообщения можно выбирать S путем; таким образом, если у нас есть сообщение длиной один символ, мы можем определить количество S^1 другого сообщения [5, с. 11 – 12]. Здесь Хартли описывает количество информации в виде логарифма количества определяемых сообщений. Исходя из этого, мы имеем следующее:

$$H(s^1) = \log(s^1) = 1 \log s,$$

где $H(s^1)$ – количество информации (Amount of information), 1 – длина сообщения, s – количество возможных выборов и s^1 – количество сообщений.

Таким образом, для сообщений длиной 1 (символ) получим следующее:

$$H(s^1) = \log s$$

Или можно так:

$$H(s^1) = 1 H(s^1).$$

Это означает, что количество информации любого сообщения длиной 1 (символ) равно 1 количеству информации сообщения длиной 1. Из этого уравнения можно получить другое уравнение:

$$f(s^1) = 1f(s).$$

Функция, которая применяется в приведенном выше уравнении, будет вычисляться следующим образом:

$$f(s) = \log s.$$

Последнее уравнение описывает количество информации Хартли. В этой функции логарифм с увеличением количества символов S обеспечивает увеличение количества информации. Основание логарифма произвольное. Если использовать натуральный логарифм, т.е. если основание логарифма $e = \sim 2.71828$ (число Непера), единицу информации называют нат (естественной единицей). Если основание логарифма число 2, единицу информации называют бит (двоичным числом). Ясно, что связь бита и ната выглядит следующим образом:

$$1 \text{ nat} = 1.44 \text{ bit}$$

Обычно в качестве основания логарифма принимают число 2, так как определение единицы информации в качестве величины, которая сокращает нашу неосведомленность о теме наполовину (в связи с выбором между двумя опциями), является и результатом в работе, и логичным [6, с. 36].

Н. Петрович предложил простую иллюстрацию количества информации Хартли:

- Предположите, что у вас 2 чашки и один из ваших друзей спрятал монету под одной из них. Как определить, где спрятана монета?

Ясно, что определить местонахождение монеты можно с помощью одного вопроса и одного ответа (одной единицы информации):

- Монета спрятана под первой чашкой?
- Нет.

Сейчас можете поднять вторую чашку – монета находится под ней.

- Теперь предположите, что у вас 4 чашки и ваш друг спрятал монету под одной из них. Сколько единиц информации вам потребуется, чтобы найти монету?

В этом случае перед вами стоит выбор из четырех выводов, вероятность которых одинакова. И вам необходимо 2 единицы информации:

- Монета спрятана под первой или второй чашкой?
- Ни под одной из них.
- Монета спрятана под четвертой чашкой?
- Нет.

Становится ясно, что монета находится под третьей чашкой.

- Если чашек 8 и монета спрятана под одной из них, тогда вам потребуется 3 единицы информации.

Таким образом, можно получить следующую таблицу (см. [6, с. 38]):

Соотношения между количеством возможных выборов и количеством информации

Количество возможных выборов	Количество единиц информации
Выбор 1 пути из 2	1
Выбор 1 пути из 4	2
Выбор 1 пути из 8	3
Выбор 1 пути из 16	4
Выбор 1 пути из 32	5
Выбор 1 пути из 64	6

Таким образом, если количество единиц информации обозначить как H и количество выборов как S , то согласно табл. 1 получим следующее:

$$S = 2^H.$$

Запишем логарифм этого выражения:

$$H = \log_2 S.$$

Это – то самое уравнение Хартли.

В изложенной выше теории информация описана в пределах вероятности, а между количеством информации и количеством возможных выборов есть прямая связь. Такова же и суть теории Шеннона [3, с. 33].

Описание информации, предложенное Хартли (информация в качестве меры количества наших выборов в представлении сообщений), было принято от Шеннона и поддержано. Задача Шеннона – это задача улучшения отношения «сигнал – шум» (Signal-to-noise ratio), и в принципе его теорию количества информации можно рассматривать в этом же аспекте. Шеннон считает, что термины «сигнал», «сообщение» и «информация» имеют почти одинаковые значения, и использует их как синонимы. Термины «сообщение» и «информация» он употребляет без учета семантики и прагматики, т.е. отделяется от смысла и рассматривает их только как коды и физические сигналы. Такие понятия, как информационный источник, передатчик и приемник, в схеме Шеннона указывают только на технические инструменты, но не на человеческие факторы.

Шеннон определяет коммуникацию в инженерии как точную регенерацию сообщения, переданного из информационного источника, с помощью приемника [1, с. 3]. Проблема коммуникации в инженерии заключается в том, что коды, передаваемые из информационного источника, должны регенерироваться в точности или с очень высокой приближенностью. Понятие количества информации, по мнению Шеннона, тоже объясняет меру вероятности выбора кодов посредством передатчика из информационного источника. Шеннон полагал, что информация в простейших случаях измеряется с помощью логарифма количества возможных выборов из символов. Задача Шеннона заключается в следующем: если один информационный источник посредством успешного выбора символов из законченной совокупности генерирует сообщение и вероятность появления одного символа зависит от предшествующего выбора, то каково количество информации, связанной с этим источником?

Шеннон ответил на этот вопрос описанием информации с помощью понятия энтропии (Entropy).

Он понимал: для того чтобы заниматься фундаментальной теорией коммуникационной системы, необходимо дать точное определение информации, такое определение, согласно которому информация – это физический параметр (physical parameter), который можно измерять. В результате он сформулировал энтропийное определение информации: чем больше энтропия, т.е. чем больше вариантов нашего выбора, тем больше информации.

В энтропийном определении информации Шеннона очевидно влияние работ Хартли [2]. Энтропия – понятие, которое применялось ранее только для систем, стремящихся к термодинамическому равновесию (т.е. к максимальному беспорядку в движении их составляющих, к увеличению энтропии), оно означает количество выборов и, следовательно, является фактором нашей нерешительности в определенном состоянии системы. Если в состоянии x мы имеем n -е количество одинаково вероятных результатов, наша неопределенность (=энтропия) в отношении того, какое событие случится, считается следующим образом:

$$H(X) = \log_2 n \text{ (bits)}.$$

И это – понятие информации Хартли.

Но если все вероятности не равны, энтропия (=количество информации) считается следующим образом:

$$H(X) = -\sum P_i \log_2 P_i \text{ (bits)}.$$

И это – количество информации Шеннона [5, с.17].

Разумеется, после Шеннона были и другие исследователи (например, Н. Винер), которые приложили усилия для описания понятия информации с физической точки зрения и, безусловно, достигли в этом успеха.

2. Семантический взгляд на информацию

В исследованиях по теории семантической информации (Theory of semantic information) можно выделить три направления, главными представителями которых являются К. Шеннон, Й. Бар-Хиллел и Р. Карнап, Ф. Дретске.

2.1. Теория Шеннона и семантика

В центре внимания у Шеннона были ограничения одного канала при передаче знаков посредством телефонной линии, и поэтому в его теории важны только физические символы, а не значение. С точки зрения теории Шеннона строки «У меня хорошее чувство» и «Няме у оршеехо овствуч» содержат определенное количество информации. По этой причине Лангфорс (Langefors) полагает, что лучшее название для теории информации Шеннона – это теория передачи сигналов (Signal transmission theory) [7]. С момента опубликования статьи Шеннона другие исследователи указывали на этот аспект его теории. Даже Уивер, статья которого была опубликована вместе со статьей Шеннона в книге «Последние статьи о математической теории коммуникаций», напоминал об отсутствии вопроса значения в теории Шеннона.

Действительно, теория Шеннона не имеет связи с

вопросом значения, но некоторые ученые говорили о том, что в этой теории вопрос значения присутствует [8, с. 9 – 10]. В дискуссии о теории Шеннона вопрос значения связывается с вопросом избыточности (redundancy): в теории информации избыточность является некой мерой взаимосвязи сигналов, в теории Шеннона избыточностью называется часть сообщения, которая не является энтропией или информацией. Избыточность не является необходимой: если её нет, то, по существу, сообщение полное или может быть полным. Избыточность – это часть сообщения, которая не зависит от свободного выбора передатчика. Для говорящих на английском языке почти половина выборов обусловлена сущностью языка и правилами его применения. При этом избыточность можно использовать для устранения помех в канале коммуникации: тот факт, что около 50% английских слов являются излишними, дает возможность исправлять ошибки в сообщении, которое принято в канале, содержащем много помех [9, с. 74]. Чем больше избыточность, тем больше возможностей избавиться от помех. Однако избыточность имеет также общий характер, который связан с естественным языком. Такое значение избыточности изложено самим Шенноном при обсуждении вопроса приближения к естественному языку [1, с. 13 – 14]. На этом основании с помощью простых коллекций искусственных языков мы можем приблизиться к одному естественному языку.

2.2. Семантическая информация Бар-Хиллела и Карнапа

Первую попытку построения теории семантической информации предприняли Р. Карнап и Й. Бар-Хиллел. Они полемизировали с обобщениями Шеннона и подчеркивали, что математическая теория информации не имеет никакого отношения к содержанию передаваемых символов, вернее, она имеет дело только с частотой (Frequency) появления символов. Однако, к сожалению, торопливые ученые других направлений (выражение Бар-Хиллела и Карнапа) начали принимать и использовать понятия и термины, обсуждаемые в математической теории коммуникаций, т.е. там, где информация рассматривается относительно семантического измерения или прагматики, использовали понятия и терминологию математической теории информации. И это при том, что математический взгляд на информацию не может быть решением семантических или прагматических вопросов. Правда, заимствование терминологии математической теории информации в тех областях, которые занимались семасиологией и прагматикой, обычно было аллегорическим, однако такие аллегорические изложения понятий также приводят к искажению мышления и недоразумению, вводящим в заблуждение [10, с. 18].

Учитывая все вышесказанное, Бар-Хиллел и Карнап попытались сформулировать основные положения единой теории семантической информации. Определяющую роль в описании главных понятий этой теории играет содержание символов. Однако Бар-Хиллел и Карнап напоминают о том, что понятия

предлагаемой ими теории не должны быть поспешно отнесены к таким областям, как психология и социальные науки, поскольку семантическая теория имеет дело только с содержанием символов, без человеческого фактора (без интерпретационных аспектов), т.е. рассматривает их объективно и оставляет в стороне умственные факторы; в этой теории считается, что смысл независим от разума [10, с. 19].

Принимая такое объяснение информации, необходимо согласиться с двумя его следствиями, которые известны как парадокс Бар-Хиллела и Карнапа:

– во-первых, количество информации в математических и логических истинах (=аналитические суждения или тавтологии к объяснению Витгенштейна), которые подтверждаются во всех возможных положениях, равно 0, так как это необходимые суждения, вероятность которых равна 1;

– во-вторых, парадоксальные предикаты, которые не подтверждаются ни в одном положении, содержат максимум информации, поскольку их вероятность равна нулю [11, с. 475].

Это противоречит здравому смыслу, и поэтому Бар-Хиллел и Карнап исключают из определения информации условие подтверждения. Они объясняют такой шаг следующим образом: «Наверное, вначале выглядит удивительным, что один парадоксальный предикат, который ни один разумный человек не примет, считается носителем большей информации. Поэтому следует принять во внимание, что семантическая информация не является обязательно истинной. Следовательно, тот ложный предикат, который может дать нам больше информации, в том значении, которое мы имеем в виду относительно информации, имеет высокий уровень информативности. Является ли информация, которую несет такой предикат, истинной или ложной, имеет ли она практическое значение, нас не касается» [12, с. 229].

2.3. Семантическая информация Дретске

В теории семантической информации Дретске информация не требует интерпретационного процесса, хотя это и является главным условием приобретения знания. Дретске подчеркивает три условия, которым должно отвечать каждое определение информации:

1) Сигнал несет тем больше информации о S, чем больше S указывает на то, что S – это F.

2) S – это F.

3) Количество информации, которое сигнал несет о S, – это то количество, которое генерировано посредством того, что F – это S (а не посредством того, что G – это S) [13, с. 63 – 64].

В соответствии с теорией Дретске информация всегда зависит от фундаментальных знаний (K); это нечто такое, что необходимо для знания [13, с. 80 – 81]. Информация – это нужная вещь для приобретения знания. Не бывает ложной информации (False information), а то, что некоторые называют ложной информацией, на самом деле есть семантический контент (Semantic content), который не является истинным. Действительно, информация – это то, что способно принести знания, а так как

знание требует истинности, информация также требует истинности [13, с. 45].

По мнению Дретске, информационный поток (Information flow) основан на ксерографическом принципе (Xerox principle). Ксерографический принцип Дретске можно объяснить так: «Если A содержит такую информацию, которую содержит B, и B содержит такую информацию, которую содержит C, тогда A содержит такую информацию, которую содержит C» [13, с. 57].

В целом теории Дретске можно оценить как пограничную между двумя взглядами на информацию – семантическим и познавательным (прагматическим). Понятие информации, по мнению Дретске, в основном связано с познавательными системами. Отношение между знанием и информацией является рекуррентным (Recursive), но оно не циклическое. Для информации о том, что S – это F, необходимо иметь знания относительно S (не имея знания о том, что S – это F). С другой стороны, информация о том, что S – это F, станет причиной того, что K поверит, что S – это F. Знание – это вера, порожденная информацией (Information – produced belief) [13, с. 91 – 92].

3. Прагматический взгляд на информацию

К числу самых известных исследователей, которые отделяются от традиции Шеннона и пытаются дать разъяснение понятию информации в связи со структурой познания (знания) человека, принадлежат Майкл Бакланд (Michael Buckland), Николас Белкин (Nicolas Belkin), Бертран Брукс (Bertram Brooks) и Питер Ингверсен (Peter Ingwersen).

Существование разных определений информации создает путаницу. Для того чтобы избежать ее, М. Бакланд предлагает прагматический подход: он стремится представить, классифицировать и описать основные применения слова «информация» и таким образом прояснить его значение. Бакланд различает три типа информации:

1. Информация в качестве процесса (Information-as-process): информация в этом смысле указывает на изменение в состоянии знания (познания) человека. Здесь слово «информация» употребляется в значении действия информирования, или действия извещения, или действия осведомления. Говорить о чем-либо кому-либо или сказанное о чем-либо – это и есть информация в качестве процесса.

2. Информация в качестве знания (Information-as-knowledge): в этом значении информация указывает на то, что познается в процессе осведомления. Информацию, уменьшающую неопределенность, можно считать информацией в качестве знания.

3. Информация в качестве предмета (Information-as-thing): информация в этом значении указывает на предметы, которые содержат информацию (являются информативными). Данные, документы и все, что можно назвать объективным знанием (на основании теории трех миров К. Поппера), являются видами информации в качестве предмета [14, с. 351].

Бакланд признает два важных различия: 1) между материальными (tangibles) и нематериальными

(intangibles) и 2) между сущностью (entity) и процессом (process). На основании этого он предлагает следующую таблицу:

Таблица 2

Классификация информации с точки зрения М. Бакланда

	Информация	
	материальная	нематериальная
сущность	Информация в качестве предмета (данные, документы)	Информация в качестве знания
процесс	Обработка информации (обработка данных)	Информация в качестве процесса (получать сведения)

На основании табл. 2 мы имеем две сущности: информация в качестве знания, которая относится к умственному состоянию человека и поэтому является нематериальной, и информация в качестве предмета (данные, документы), которая относится к ассимиляции и воплощению информации как знания в материальных формах и поэтому является материальной. Кроме этих двух сущностей мы имеем еще и два процесса: информация в качестве процесса, которая указывает на умственную обработку информации и поэтому является нематериальной, и обработка информации (Information processing), которая указывает на процесс управления информацией в качестве предмета и поэтому является материальной. По мнению Бакланда, предметом информатики является информация в качестве предмета (книги, документы, данные, существующие в компьютерных информационных системах в качестве физических сущностей и др.), а задача информационных систем – обработка информации.

Материальная сущность, которую можно непосредственно осматривать и пробовать, – это не знание (познание), а информация в качестве предмета. Информация в качестве знания нематериальна: ее невозможно осматривать или измерять. Знание и вера – это личностное, умственное и абстрактное. По этой причине для передачи знание должно превратиться в информацию в качестве предмета, т.е. должно войти в положение выражения и изложения и воплотиться в качестве знака или текста. Любое изложение, выражение и ассимиляция – это информация в качестве предмета.

По мнению Бакланда, информационные системы (в широком смысле) имеют дело с информацией в качестве предмета: библиотеки – с книгами и документами, компьютерные информационные системы – с данными (в форме физических битов и байтов), музеи – с воплощением мыслей и произведениями предшественников. Цель этих систем заключается либо в том, чтобы пользователи стали осведомленными (информация в качестве процесса), либо в том, чтобы знания человека изменились (информация в качестве знания). Однако средства, необходимые для этого, есть представленная информация (информация

в качестве предмета). То, с чем мы имеем дело в действительности, – это только системы, основанные на наглядных представлениях умственного познания. Обработка информации также указывает на управление (Handling), манипулирование (Manipulation) и производство образов и новых копий информации в качестве предмета [14, с. 352].

Идея информации в качестве предмета противоположна идее Норберта Винера, который говорил: информация – это информация, не материя и не энергия. Махлуп (Machlup) тоже отвергал идею информации в качестве предмета. Он давал определение информации только на основании коммуникации и говорил, что информация может иметь только два значения: 1) говорить что-либо (в общем смысле слова); 2) то, что говорится. Махлуп полагал, что всякое использование слова «информация», которое противоречит двум указанным значениям, является либо метафорой (или аллегорией), либо искажением значения слова «информация» [14, с. 352].

Фэрторн (Fairthorn) тоже отвергал то, что информация есть материя. Информация – это свойство, связанное со знанием получателя, и интерпретация знака, она не связана ни с отправителем, ни с всезнающим наблюдателем, ни с самим знаком [14, с. 352].

Бакланд, опровергая эти теории, указывает на ежедневное употребление слова «информация» и говорит, что с распространением компьютерной и сетевых технологий распространилось и понятие информации в качестве предмета: это значение слова «информация» используется, чтобы указать на информационные базы, документы, книги и др. Он полагает, что если информация в качестве предмета получает разные названия, то это область информационных систем, так как информационные системы, в конечном счете, могут иметь дело только с такой информацией.

4. Герменевтико-экзистенциальный взгляд на информацию

В распространенных исследованиях предметом обсуждения стала информация, отвлеченная от экзистенциальных правил жизни (Existential practices of living). Р. Капулло ставит под сомнение такие идеи. По его мнению, независимой сущности под названием «информация» не существует. Напротив, информация подвергается сильному влиянию экзистенциальных действий. Информация никоим образом не может быть бесполезной и нейтральной. Если Кант опровергал понятие «вещь в себе» (Thing itself), то Капулло опровергает понятие «информация в себе» (Information itself). Подчеркивая, что информация, по сути, содержит этические особенности, Капулло связывает ее с жизненным миром людей (Lebenswelt = life-world) и описывает в многослойных и многомерных контекстах (положениях) [15, с. 2].

Признание связи информации с жизненным миром – это опровержение редукционизма, который стремится превратить все в единственное отношение. В рамках редукционизма люди преобразовываются (даже разлагаются) в процессе окисления (биологизм), превращаются в систему «стимул-реакция» (психоло-

гизм) или в социальные элементы (социологизм). Подобным же образом и информация превращает все (конечно, включая людей) в единственное отношение, где даже наши души рассматриваются всего лишь как обработка информации. Информация приводит к идеям, сосредоточенным в информационной технологии (Information technology – centered ideas), которые стимулируют людей приобретать все больше и больше информационных (в основном компьютерных) навыков, чтобы быть превосходным homo informaticus* (информационно-технологичным человеком), в то время как те, у кого слабые информационные навыки, располагаются ниже. В конечном счете эти редукционизмы исключают и игнорируют наш жизненный мир, состоящий из человеческих значений. Герменевтический подход, включающий в себя рассмотрение жизненного мира, совершенно противоположен таким редукционизмам [15, с. 6].

По мнению Капурро, необходимо различать информацию в качестве предмета и информацию в качестве сигнала (Information-as-sign). Информация в качестве сигнала – это то, что содержит интерпретационные свойства. Интерпретационный взгляд направляет внимание от внешних свойств к спусковым механизмам (Release mechanisms). В герменевтическом взгляде, в отличие от позитивистского, содержатся факторы и объяснения, относящиеся к культуре, которые определяют информацию. Можно легко посчитать количество слов какого-либо документа и описать количество его информации на основе бита (позитивистский взгляд), но трудно сказать, для кого важен этот документ и какие важные вопросы, на которые этот документ может ответить, содержатся в нем (герменевтический взгляд). Информация, по сути, является герменевтической категорией.

Вопросы, связанные с интерпретацией, очень трудны, и в наше время, для которого характерен процесс солипсизма, эта трудность увеличивается, поскольку значение определяется в социальных и культурных руслах и нельзя исследовать его, используя индивидуалистический подход (Individualist approach) [16]. При рассмотрении герменевтико-экзистенциальной точки зрения на информацию Капурро вначале подвергает критическому анализу субъективистские предположения, господствующие над исследованиями, распространенными в сфере информационной науки. По мнению Капурро, в информационной науке преобладают три парадигмы, которые находятся под влиянием позитивизма и рационализма: парадигма представления (Representation paradigm), парадигма источника-канала-получателя (Source-channel-receiver paradigm) и платоническая парадигма (Platonistic paradigm). Все эти парадигмы рассматривают субъект познания во взаимодействии с информацией, которая, как предполагается, есть во внешнем мире [17].

* Капурро придумал термин homo informaticus по аналогии с термином homo faber (в значении «человек, изготавливающий орудия»). Если в процессе промышленной революции человек превратился в существо, изготавливающее орудия, сейчас человек считается существом, обрабатывающим информацию, и историю описывают с точки зрения развития информационных и коммуникационных средств. Это можно видеть в работах таких ученых, как Элвин Тоффлер, Маршалл Маклюэн.

Согласно парадигме представления, человек находится в субъекте, в котором он является исключительно наблюдателем внешней реальности. Приобретение знания состоит из составления понятий о вещах посредством их представления (изображения) в уме/ядре субъекта познания. Эти представления, так как они обрабатываются и кодируются в уме, можно передавать другим умам, сохранять и обрабатывать в машинах (компьютерах). С этой точки зрения люди – это биологические обработчики информации, а информация – кодированное изображение и представление реальности, которое люди могут использовать для специальных умственных целей. Основываясь на этом, можно сказать, что информационная наука связана с рассмотрением представлений, кодирования и разумным (рациональным) использованием информации [17].

Парадигма источника-канала-получателя говорит, что связь между источником и получателем – это и есть обмен информацией. Для того чтобы получатель мог понять значение сообщения, передаваемого отправителем, должен быть общий резерв символов. Однако обмен информацией можно рассматривать только с точки зрения структуры информации. В таком случае мы говорим о синтаксической информации (Syntactic information). В этой парадигме информационная наука в основном связана с влиянием информации на получателя. В то же время получатели являются не объектами влияния, а искателями информации, которые с ее помощью решают свои проблемы [17].

Платоническая парадигма признает «информацию по сути» (Information in itself) и считает ее одним из пространств человеческого знания, которое имеет материальное основание. Это можно назвать материалистическим платонизмом (Materialistic Platonism). В идеалистическом изложении этой парадигмы знание – это «объективный факт по сути» (Objective in itself), не зависящий ни от каких материальных оснований. Примером его может служить 3-й мир Поппера, т.е. мир информации, являющийся независимым как от разума (2-го мира), так и от объектива (1-го мира). Здесь информация занимает такое же положение, как законы логики и математики, которые не зависят от психических или биологических процессов, вмешивающихся в мышление. Конечно, в этой парадигме отношение между миром информации и миром субъекта знания остается проблематичным. В материалистическом изложении платонической парадигмы информационная наука рассматривает информацию с точки зрения ее объективности в материальном воплощении (например, в документах или электронных данных). В идеалистическом изложении информация рассматривается как объективная и в то же время нематериальная сущность.

По мнению Капурро, основанием этих трех парадигм в современной философии является различие между субъектом и объектом или различие между *res cogitans* и *res extensa*. Они описывают человеческую коммуникацию на основе отношения «субъект – объект»: объекты внешнего мира представляются в разуме и ядре субъекта, а процесс коммуникации со-

стоит из обмена информацией между субъектами, которыми эта информация связана с представлением объектов внешнего мира в их разуме.

Капулло считает человеческую коммуникацию познавательным взглядом на информацию. По его мнению, в экспериментах Кренфилда (Cranfield Experiments), которые были проведены под руководством Клевердона (Cleverdon) в 1960-х гг. и завершились рассмотрением масштабов выборки (recall) и точности (precision), мы становимся свидетелями превосходства механического взгляда, опирающегося на дуализм субъекта – объекта. Затем в 1970-х гг. с появлением теории ASK Белкина (теории нерегулярного положения познания) и познавательной точки зрения Ингверсена мы становимся свидетелями познавательного подхода. Здесь подчеркивается умственная структура знания человека. Этот факт стал причиной того, что некоторые исследователи путали познавательный подход с предпониманием (Vorverständnis) в философской герменевтике. Однако философская герменевтика имеет тенденцию отхода от дуализма субъекта – объекта, тогда как познавательная парадигма осталась в ограниченном пространстве этого дуализма [17]. С познавательной точки зрения информация считается сверхъестественной (super-physical) сущностью, которая существует только в умственных и познавательных пространствах.

5. От герменевтики – к ангелетике

Ангелетика – это название новой области научных исследований, которая изучает осведомление (изложение смысла), информирование (выбор из многочисленных возможностей сообщения) и понимание, которые являются тремя сферами коммуникации в человеческих обществах.

Почему так назвали? И каково ее происхождение? Слово angeletics образовано от греческого слова, которое означает «объявление» и «осведомление». Слово angel применяется для названия посланников Бога. Наука об ангелах (ангелогия) как тип ангелетики рассматривается в одном ряду с теологическими темами. Ангелетика – это парадигматическая наука в веке информации, которая относится к гуманитарным и социальным наукам и сильно связана с вопросами риторики. Цель ангелетики – изучение генерации, распределения, изложения и контроля сообщений в положении человека [18, 19].

Ангелетика может играть роль парадигмы в науке и обществе XXI в. Предметом ангелетики являются осведомление и сообщение. Ее проблематику образуют:

- источник, цель и содержание сообщения;
- структуры власти;
- техника и средства передачи сообщения;
- образы жизни;
- история сообщений и доставляющих сообщения;
- кодирование и декодирование;
- психологические, политические, экономические, эстетические, этические и религиозные аспекты сообщений [19].

Ангелетика – междисциплинарная область исследований, которая связана с изучением средств, семиотикой и герменевтикой. Капулло пишет: «Началом всякой интерпретации является передача сообщения. Гермес – посланник богов и толкователь этих посланий. Посланическая сущность коммуникации – это предмет анализа и исследования в ангелетике. Вместе с тем все процессы передачи сообщения основываются на герменевтическом положении, согласно которому отправитель и получатель имеют общую основу понимания. Другими словами, ангелетика, основываясь на том, что понимание, или, более общими словами, процесс выбора из двух систем, возможно, работает над дуализмом отправителя/получателя. Герменевтика, исходя из того, что объект процесса интерпретации (сообщение) успешно представлен получателю, который имеет значение объекта выбора, работает над различиями между предпониманием и интерпретацией. Семиотика изучает весь процесс, с помощью которого знак, т.е. то, что указывает и, как предполагается, выбрано интерпретатором, познается в качестве динамической и самоорганизующейся структуры» [18].

В ангелетике предпринимаются попытки описания структуры и явления передачи сообщений в историческом аспекте. Ангелетика должна обеспечить всестороннюю рамку, в которой обсуждаются механизмы генерации, совместного использования или передачи предпониманий или горизонтов значений. Для описания генерации или передачи сообщений необходимо рассмотреть горизонты значений или предпонимания с исторической точки зрения. Капулло следует этому направлению в рамках генеалогии информации [3, с. 185 – 187].

Явление информации связано с властью. Эта связь может быть вертикальной (сообщения устанавливаются) или горизонтальной (происходит обмен сообщениями). Генеалогия информации должна изучать механизмы власти в информационных процессах, обращая внимание на эти две структуры (вертикальную и горизонтальную). Генеалогический анализ показывает, что не существует совершенной вертикальности или горизонтальности, а также линейного или идеального развития от вертикальных структур к горизонтальным. В принципе, в генеалогии на историю не смотрят линейно и считается, что у каждого исторического периода свои специфические законы. Поэтому теорию свободного потока информации в прошлом и настоящем называют несбыточной мечтой [20].

При рассмотрении вертикальных и горизонтальных структур передачи сообщений Капулло использует коммуникологию Вильяма Флуссера. Флуссер признает две отличающиеся друг от друга цели коммуникации: диалогическую цель, которая связана с созданием новой информации, и дискурсивную цель, которая связана с распределением информации. Средства, основанные на диалогической цели, имеют горизонтальную структуру, средства, основанные на дискурсивной цели, – вертикальную [18].

На этом основании Капурро рассматривает информацию в трех контекстах [20]:

- информация в контексте мифов, поэзии и открытия;
- информация в контексте философии, науки и технологии печати;
- информация в контексте электронной технологии.

6. Информация в контексте электронной технологии

Капурро полагает, что в конце современности, которая является веком развития электронных сетей, необыкновенным образом некоторые особенности речи передаются на письме, в том числе преимущество обращения, зрительное содержание и основание изложения.

С появлением всемирных электронных сетей, таких, как Интернет, мы становимся свидетелями двух явлений: с одной стороны, Интернет вращается вокруг CNN (Cellular Neural Network – клеточная нейронная сеть), т.е. обеспечивает возможность того, что получатель сообщения одновременно может быть и отправителем сообщения; с другой стороны, возникают новые виды олигополии информации (Information oligopoly) [18, 20].

В области электронных технологий идеальная структура без современной научной цензуры становится частью всемирной сети сообщений и передачи сообщений. Вместо легитимации знания посредством общественных органов (например, партии, университета) или идеологических рассказов о речи мы сталкиваемся с множествами распределенных сообщений в виде информационных баз. Но, с другой стороны, идея информации как общественной ценности (Social good), которая является характерной особенностью века Просвещения, уступает свое место идее информационной экономики. Информация производится уже не для общественной ценности и не на благо обществу, а для продажи. Мы становимся свидетелями превращения знания в товар (Commodification of knowledge), и в развитии информационной промышленности ключевое значение приобретает распределение рабочей силы [20].

II. ФИЛОСОФИЯ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ

Философию информационной технологии можно считать одной из ветвей философии технологии. Философское мышление о технологии, в общем, связано с пониманием сущности и значения создания и применения. Особенно мышление о том, что создано и применено до настоящего времени. Однако такие размышления возникают при столкновении между двумя большими традициями, одна из которых основана на инженерной точке зрения, а другая – на антропоцентрической.

С инженерной точки зрения технология является совершенно человеческим делом, и поэтому ее влияние на все аспекты человеческой жизни считается не только дозволенным, но и необходимым. Это точка зрения, требующая развития.

Сторонники антропоцентрической точки зрения считают технологию ограниченным аспектом человечества и поэтому всегда беспокоятся о том, чтобы этот ограниченный аспект не одержал верх над другими аспектами человеческой жизни. Это догматическая точка зрения.

Столкновение этих традиций происходит и в области информационных технологий (ИТ): одни празднуют приход информационных технологий, другие осмотрительно стараются ограничить их возможности и влияние.

Философия информационных технологий начинается с определения своей темы, т.е. собственно информационных технологий. Далее она пытается рассмотреть всю историю философии с точки зрения информационных технологий и на этом основании, во-первых, найти новые ответы на старые философские вопросы, используя понятие ИТ, во-вторых, решить понятийные вопросы и разногласия, возникшие относительно ИТ, и внутренние понятийные вопросы ИТ с помощью философских произведений.

Изучение истории философии с точки зрения ИТ устанавливает, что одновременно с приходом ИТ появилось новое культурное произведение. Были приложены многочисленные усилия для описания новых способов жизни в мире, который достигли ИТ, при этом основное внимание было сосредоточено на экономических и политических аспектах. Однако остаются и другие важные вопросы:

Что является реальным и очевидным после появления и распространения ИТ?

Какова глубинная структура ИТ-похожих явлений?

Будет необдуманно рассматривать ИТ только в качестве комплекта устройств обработки и коммуникации, полагая, что их структура и само понятие ИТ остаются реальностью незатронутой. Так, Бургман признает тройное преимущество информации:

- информация относительно реальности (естественные науки);
- информация для реальности (инженерное проектирование);
- информация в качестве реальности (расчетные представления).

Чем больше развиваются ИТ и расширяется область влияния информации в качестве реальности, тем больше отличаются человеческие понятия об основаниях, создающих реальность, от предшествующих. Рассмотрение этих проблем требует постановки метафизических вопросов об ИТ, среди которых тоже возникает столкновение инженерной и антропоцентрической точек зрения.

Инженерная точка зрения в ИТ имеет два основных аспекта:

- 1) физические существа в ИТ;
- 2) познавательные способности (cognitive) ИТ.

В 1948 г. Норберт Винер обозначил основные элементы, создающие реальность, следующим образом: материя, энергия и информация. С того времени стало возникать много вопросов относительно метафизического положения информации. Сам Винер начал рассмотрение этой проблемы с классификации ма-

шин на группы, которые были соотнесены с тремя основными элементами реальности:

а) машины, передающие материю (линия генерации);

б) машины, производящие энергию и передающие энергию (электростанция или двигатель);

в) машины, манипулирующие информацией (коммуникационные системы или компьютеры).

Феноменология столкновения человека с технологиями также показывает основное различие IT-оборудования и IT-машин. Кибернетическое и автоматическое свойства IT превратили эту технологию в подобие магистрали, которая ждет, чтобы по ней проходили люди. IT – это среда для деятельности человека (и поэтому она названа новым средством – передающей средой).

Познавательные способности IT вызывают другие онтологические вопросы, например:

Что такое искусственный интеллект?

Разработаны ли разные виды искусственного интеллекта (алгоритмический, эвристический)?

С другой стороны, если рассматривать область биотехнологических проблем, такие вопросы могут устранить различие между продукцией, созданной сверхсовременными технологиями, и живыми организмами. Стены, разделяющие их, находятся в состоянии разрушения. С точки зрения технологии информация существует повсюду: и в органическом мире, и в мире продукции. Исходя из этого, Ширрмахер говорит: IT – это своего рода искусственная природа, где мы свободны и можем жить беззаботно, игриво, на основе эстетических желаний, без власти предписанных законов, но для сохранения гармонии с ней мы должны принимать ограничения нового рода.

Антропоцентрическая точка зрения в IT имеет разные философские корни. Без сомнения, разделяя мир на «до IT» и «после IT», мы многое упрощаем, но такое сопоставление показывает некоторые аспекты реальности. Однако сторонники антропоцентрической точки зрения отдают предпочтение сопоставлениям такого рода, как сопоставление философий Гегеля и Сократа: критический позитивизм против заядлого спорщика. С гегелевской точки зрения с помощью IT мы приобрели силу народов разных миров, однако философский негативизм относительно нее закончится исключительно безответственностью перед этой силой. А с сократовской позиции инженерная точка зрения – это почтение к IT, которые нашли больше утверждения и приобрели мировую славу и популярность.

Таким образом, в условиях, при которых IT приобрели безраздельное господство, сторонники антропоцентрической точки зрения ищут ответы на вопросы, связанные с сущностью и значением IT. Такие вопросы обсуждаются в четырех категориях:

1) Этическая (нравственная) категория:

– Угрожает ли IT частной сфере человека?

– IT превратилась в среду обитания людей, однако эта среда практически не тестируется перед представлением. Не является ли распространение такого способа разработки программного обеспечения и

технологических продуктов распространением нравственной безответственности?

2) Политическая категория:

– Создан ли Интернет таким образом, чтобы обеспечивать социальную справедливость посредством предоставления равного доступа к нему?

– IT сидит на спине гигантской промышленности. Не ставит ли этот факт под вопрос справедливость и законность IT?

3) Психологическая категория.

Психологические вопросы возникают вместе с эпистемологическими:

– В условиях «информационной бомбардировки» способность людей понять ее уменьшится или будет развиваться вместе с ней? (Это при том, что информационное опасение является одним из основных парадоксов жизни в эпоху IT.)

4) Антропологическая категория.

Такие вопросы касаются социальных приспособлений и создают «новое толкование информации», например:

– Что означает «виртуальная жизнь»?

– Насколько будет развита «жизнь на экране»?

Информационная технология скрывает от нас реальность по очень запутанным и основательным причинам информационной бомбардировки. С антропоцентрической точки зрения IT может искривлять способ нашего бытия. Эту вероятность лучше обсуждать, опираясь на произведения Мартина Хайдеггера.

Рассматривая «вопрос о технологии», Хайдеггер стремится показать: технологию создает скорее раскрытие реальности, разглашение тайн и в то же время их скрывание, нежели технологические машины и оборудование. Современная технология не продвигается вместе с природой и согласно ее законам, вернее, она считает природу «источником» для своего манипулирования и вступает с ней в борьбу. IT доставляет языку те же неприятности, которые современная технология доставляет природе, т.е. превращает его в «источник» для своего манипулирования.

Информационная технология скрывает реальность настолько, насколько и разоблачает ее, и до тех пор, пока мы будем опираться на ее раскрытия, эти скрытия будут господствовать. Для объяснения этого утверждения нужно обратиться к герменевтике в понимании Хайдеггера (именно герменевтику, а не рациональность Хайдеггер считает отличительной чертой человечества).

В герменевтике Хайдеггер подчеркивает два основных момента:

а) Любое раскрытие (получение информации) – это не простой процесс, он включает в себя и процесс интерпретации; и сама интерпретация основывается на текстах, чувственном восприятии, мышлении и способах жизни, методе диалектики части и целого (герменевтическом круге). Часть можно понять только благодаря целому, и невозможно познать целое, не зная его части. Понимание – это процесс перемещения между частью и целым. Поэтому любая информация содержит свою интерпретацию и информационная технология является частью огромного мира, которую невозможно понять отдельно от целого.

б) Любое раскрытие – это новое скрытие. Информационная технология стала причиной того, что информация присутствует везде, однако это совершилось большой ценой: информация стала завесой над самым «существованием», основной сущностью реальности и уникальной связью человека с ней. Например, обратите внимание на виртуальную реальность, виртуальную жизнь или виртуальную личность.

Конечно, возражая Хайдеггеру, можно привести доказательства, что ИТ расширила определение реальности. Но Хайдеггер в ответ скажет, что для ИТ язык является источником своих манипуляций.

* * *

В заключение кратко остановимся на некоторых перспективных направлениях философии ИТ.

Во-первых, радикальная критика Хайдеггера, высказанная им по поводу технологии (и особенно ИТ), стала причиной того, что многие разработчики ИТ-систем стремятся пересмотреть свои разработки и создать мост между инженерной и антропоцентрической точкой зрения.

Во-вторых, Марк Постер (Mark Poster) убежден, что хайдеггеровская точка зрения относительно ИТ неверна, так как он критиковал науку и технологию, которая сама о себе не думает (позитивистская наука), в то время как ИТ всегда находится в положении пересмотра себя.

В-третьих, еще в 1980-е гг. двое исследователей – Виноград (Winograd) и Флорес (Flores) – заявили: исследования Хайдеггера содержат доказательство того, что ИТ не работают в соответствии с прогнозами ученых об их достижениях. В последующие десятилетия Коин (Coyne) помимо теорий Хайдеггера затронул теории других ученых, таких, как Жак Деррида (Jacques Derrida), и привел доказательства, что пришло время вместо многочисленных настаиваний на рациональных методах ИТ обратить также внимание на значение метафор в них. Такие аллегории, как Windows, Desktop, Icons или память (в качестве компьютера), ускорили развитие ИТ, но одновременно скрывали и скрывают в себе основную часть истины, и мы должны с помощью новых методов выявить эту истину.

СПИСОК ЛИТЕРАТУРЫ

Настоящий список содержит ряд библиографических описаний на языке фарси. Для удобства читателей в конце каждого из этих описаний приведен его перевод на русский язык. (Ред.)

1. Shannon C. E., Weaver W. The Mathematical Theory of Communication. – Urbana : The university of Illinois Press, 1949.
2. Chiu E. et al. Mathematical Theory of Claude Shannon: A Study of the Style and Content of His Work up to the Genesis of Information Theory [On-line]. – 2001. – URL: <http://mit.edu/6.933/www/Fall2001/Shannon1.pdf>
3. Khandan M. Falsafahe ittilaat. – Tehran : intesharate Chopor, 2009. [Хандон М. Информа-

ционная философия / Мухаммад Хандон. – Тегеран : Чопор, 2009.]

4. Hartley R. V. L. Transmission of information // Bell System Technical Journal. – 1928. – Vol. 7 (July). – P. 535 – 563.
5. Lubah Y. Nazariyee ittila' / tarjomahe Hasanali Ozarnoosh. – Mashhad : Danieshgahe Fidavsi, 2001. [Любе Я. Теория информации / пер. Хасанали Озарнуша. – Мешхед : Университет им. Фирдоуси, 2001.]
6. Petrovich N. N. Dar shammahe az ittilaat va irtibatat = Ittilaat va irtibatat: Aknoon va ayande / tarjomahe Abdolhosain Ozarang. – Tehran : Ketabdar, 1999. – Safhaye 25 – 28. [Петрович Н. Н. Немного об информации и коммуникации = Информация и коммуникация: настоящее и будущее / пер. Абдулхусайна Озаранга. – Тегеран : Китобдор, 1999. – С. 25 – 28.]
7. Sveiby K. E. What is information? [On-line]. – 1998. – URL: <http://www.sveiby.com/articles/information.html>
8. Rapoport A. What is information? // Introduction to Information Science / ed. T. Sarasevic. – New York : R. R. Bowker Company, 1970.
9. Severin W., Tankard J. Nazariyehae irtebatat / tarjomahe Aliriza Dehqan. – Tehran : intesharate Daneshgahe Tehran, 2005. [Северин У., Танкард Дж. Теории коммуникаций / пер. Алиризо Дехкана. – Тегеран : Тегеранский университет, 2005.]
10. Bar-Hillel Y., Carnap R. Semantic Information // Introduction to Information Science / ed. T. Sarasevic. – New York : R. R. Bowker Company, 1970. – P. 18 – 23.
11. Adams F. The Informational Turn in philosophy: minds and machines. – 2003. – Vol. 13 (November). – P. 471 – 501.
12. Bar-Hillel Y., Carnap R. An Outline of a Theory of Semantic Information // Bar-Hillel Y. Language and Information: Selected Essays on Their Theory and Application. – London : Addison-Wesley, 1964. – P. 221 – 274.
13. Dretske F. I. Knowledge and the flow of information. – Cambridge : MIT Press, 1981.
14. Buckland M. Information as thing // Journal of the American Society for Information science. – 1991. – Vol. 42. – P. 351 – 360.
15. Takenouchi T. Capurro's Hermeneutics Approach to Information Ethics: Ethics in the Information society and the Development of Angeletics // International Journal of Information Ethics. – 2004. – Vol. 1 (June). – P. 1 – 8.
16. Capurro R., Hjørland B. The concept of information [On-line]. – 2003. – URL: <http://www.capurro.de/infoconcept.html>
17. Capurro R. Foundation of information science: Review & perspectives [On-line]. – 1992. – URL: <http://www.capurro.de/tampere91.htm>
18. Capurro R. Angeletics: A Maessage theory [On-line]. – 2003. – URL: http://www.angeletic_zkm.html

19. Capurro R. What is Angeletics? [On-line]. – 2004. – URL: <http://www.capurro.de/angeletics.html>
20. Capurro R. On the genealogy of information [On-line]. – 1996. – URL: <http://www.capurro.de/cottinf.htm>

Материал поступил в редакцию 24.01.12.

Сведения об авторе

АЛОШТИ Худжатуллах Рашиди – член научной группы Исламского Университета Азад Коимшахра Исламской республики Иран (кафедра государственного управления), аспирант кафедры информационных и коммуникационных технологий Таджикского национального университета
E-mail: rsh_hojat@yahoo.com

Семантическое представление текста, описывающего механику силовых процессов, в универсальной онтологии

Область силовых процессов в универсальных онтологиях обычно описана значительно хуже, чем смежная с ней область пространственных отношений. В статье предпринята попытка показать возможный вариант семантического представления текста, описывающего ситуации силового взаимодействия, с помощью онтологии и лексикона, созданного на ее базе.

Ключевые слова: универсальная онтология, силовые процессы, Динамика Сил, автоматическая обработка текста, лексикон, семантическое представление текста

1. ОНТОЛОГИЯ СИЛОВЫХ ПРОЦЕССОВ

Онтологии, используемые в области искусственного интеллекта, представляют собой модель знаний о мире. В зависимости от задач, которые стоят перед разработчиками онтологий, эта модель может выглядеть по-разному. Онтологии могут описывать узкоспециализированные или достаточно широкие предметные области и быть более или менее формализованными. Менее формализованными, но имеющими значительно большую область покрытия являются универсальные онтологии, используемые для работы с текстами на естественном языке. Фрагмент такой онтологии мы и попытались разработать.

Обычно в ресурсах подобного типа достаточно подробно разработана область пространственных отношений, однако смежной с ней области силовых процессов уделяется значительно меньше внимания и концептуализация законов силовых взаимодействий отражена номинативно. Это происходит даже в тех случаях, когда создатели ресурса делают успешную попытку моделировать область силовых процессов не в научной, а в языковой картине мира (см. [1 – 4]). Наша задача состояла в том, чтобы представить область силовых взаимодействий разной природы не схематично (т. е. простым перечислением релевантных семантических признаков внутри концепта), а как единое пространство, объединенное общей семантической теорией. В качестве такой теории была выбрана теория Динамики Сил Л. Талми [5]. По нашему мнению, основное достоинство теории Талми в том, что она полностью охватывает предметную область и имеет достаточную выразительную силу, а также дает возможность эффективно учитывать импликации при анализе значения текста на естественном языке.

Следующим этапом нашей работы было создание лексикона, который позволяет учесть дополнительные параметры ситуации при выборе той или иной лексической единицы.

Таким образом, семантическое представление текста в предложенном нами формализме имеет двойной формат: с одной стороны, это сценарный контур с привлечением концептов онтологии, объединяющий примитивные концепты в единую динамическую структуру и представляющий закономерности протекания силовых процессов, с другой – более подробное описание, созданное на базе лексикона.

2. СТРУКТУРА ОНТОЛОГИЧЕСКОГО КОНЦЕПТА

Основное отличие разрабатываемого нами фрагмента онтологии от других доступных онтологий состоит во внутреннем устройстве концептов и характере отношений между ними. Как было показано нами в работе [6], языковая концептуализация силовых взаимодействий может быть представлена в терминах *тенденций* и *соотношений*, приводящих к определенному результату.

На основе предложенных Л. Талми концептуальных схем мы выделили 12 типов базовых ситуаций, которые отражают наивное представление о законах физической механики. Таким образом, в нашем распоряжении оказывается 12 онтологических концептов. К сожалению, описание каждого из них достаточно трудоемко, а поскольку для понимания устройства онтологии достаточно объяснить базовые принципы построения концептов, мы этим и ограничимся (для более подробного ознакомления со структурой концепта отсылаем читателя к работе [6]).

Каждый выделенный нами тип силового взаимодействия представлен в виде одной или нескольких

схем, структурными элементами которых являются (рис. 1):

- a) участники ситуации:
 - антагонист (Ant) – воздействующий,
 - агонист (Ago) – претерпевающий воздействие;
- b) внутренняя тенденция участников:
 - импульс к действию (toward action),
 - тенденция к покою (toward rest);
- c) соотношение сил:
 - более сильный (stronger entity) – менее сильный (weaker entity);
- d) результат взаимодействия:
 - действие (action) / движение (motion) – покой (rest)*.

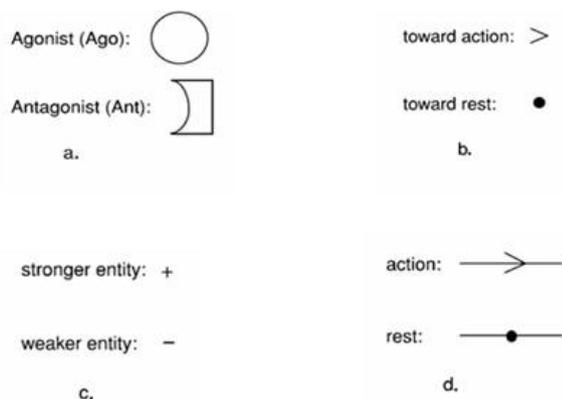


Рис. 1

К нотации Талми, представленной на рис. 1, мы добавили еще одно графическое обозначение – для процесса деформации (рис. 2):



Рис. 2

Представим предложение *Дверь не открывается* в его «силовой» интерпретации (другие возможные для него интерпретации в рамках данного исследования игнорируются) в виде сценарной схемы силовых процессов (рис. 3):

Дверь не открывается.

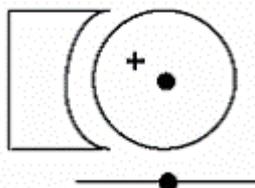


Рис. 3

Схема показывает, что на дверь воздействует некоторая сила, которой недостаточно, чтобы дверь открыть, поэтому результатом взаимодействия оказывается состояние покоя.

* Не всегда есть необходимость разграничивать понятия *движение (motion)* и *деформация (deform)*, поэтому состоянию *покой (rest)* в общем случае будем противопоставлять *изменение (action)*, под которым следует понимать дизъюнкцию (*движение^деформация*).

Однако не все концепты описываются одной схемой – некоторые имеют более сложную внутреннюю логическую структуру. Для примера рассмотрим глагол со значением импульсного воздействия *толкать*. Успешность воздействия не определена, а значит, и сценарное описание данного силового процесса не может быть однозначным. По этой причине мы будем описывать соответствующий ему концепт через дизъюнкцию, которая показывает, что движение (деформация) может не последовать, если силы воздействующего для этого недостаточно (рис. 4).

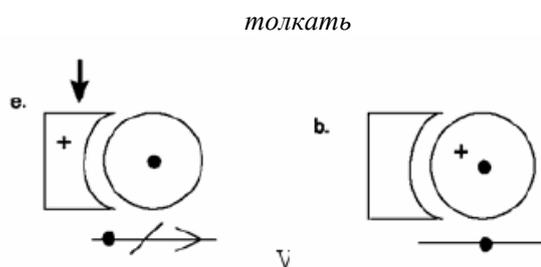


Рис. 4

О формальном неграфическом формате представления концепта мы скажем ниже.

3. ПРИНЦИП ПОСТРОЕНИЯ ЛЕКСИЧЕСКОЙ СТАТЬИ

В предыдущем разделе были описаны логическая структура концепта и базовые закономерности силовых переходов. Однако это не дает представления о конкретных признаках ситуаций силового взаимодействия, которые оказывают значительное влияние на употребление той или иной языковой единицы. Фиксация значений этих признаков происходит в лексиконе.

Начнем с рассмотрения основных ролей участников – Антагониста и Агониста.

Среди свойств Антагониста (воздействующего) выделяются два: *одушевленность / неодушевленность* и *лицо / не-лицо*. Важно сразу отметить, чем эти два параметра отличаются. *Одушевленность* во многом определяет тип Агониста (такие его характеристики, как, например, манипулируемость), а также распределение сил между Агонистом и Антагонистом. Эта характеристика описывает физическую сферу. *Лицо / не-лицо* – параметр, главным образом определяющий импликации, которые станут основой для дальнейшего семантического расширения. С ним неразрывно связано понятие цели действия (ср. глаголы *бросить* и *выбросить*). От того, является ли Антагонист лицом, могут зависеть и такие существенные вещи, как выбор предлога. Например, *тащить* что-либо можно *за* собой и *на* себе, только последний предлог предпочтительнее именно в том случае, когда Антагонист – лицо. В работе [7] также отмечается, что идея агенса напрямую связывается с идеей воли и цели. При этом акцентируется семантическая сопряженность с этой идеей значений граммы несовершенного вида и в качестве подтверждения этому наблюдению приводятся деструктивные глаголы

(рвать, разбивать, ломать и др.), которые не имеют актуально-длительного значения, когда человек выступает не в роли агенса, а в роли «невольной причины» разрушения. Ср. предложения (примеры взяты из [7]):

Его толкают, и он разбивает окно,

где деструктивное действие не намеренное и соответствующие глаголы употреблены нарративно, и

Посмотри-ка, они уже и витрины разбивают,

где агенс уже действует целенаправленно и реализуется актуально-длительное значение глагола несовершенного вида.

В общую концептуальную структуру также вводится описание физических и топологических свойств Антагониста.

Описание Агониста (претерпевающего воздействие) устроено похожим образом.

Остальные семантические признаки являются факультативными. Это значит, что их наличие в лексической статье в той или иной степени зависит от онтологического концепта, к которому относится слово. Например, не всегда необходимо исчерпывающее описание физических параметров Антагониста (таких, как размер и вес). Для онтологических концептов, которые фиксируют соотношение сил между Антагонистом и Агонистом, к которым относятся, например, глаголы *двигать* и *держать*, описание размера и веса обычно необязательно.

Вообще, в результате работы над лексиконом получилась достаточно объемная структура, состоящая из семантических признаков разного вида, но на практике заполняются не все из них. Набор значимых признаков определяется типом концепта. Выбор говорящим того или иного слова естественного языка для описания ситуации базируется на его представлении об этой ситуации. При этом определенные аспекты значения являются более важными, чем другие. В терминах Р. Лангакера, именно они профилируются, влияя на дальнейшую вербализацию [8]. Например, семантически важным может быть способ перемещения объекта (как у глаголов *нести* или *тащить*) или траектория движения участников ситуации (*раскачивать*), вектор приложения силы (*поднимать*) или форма объекта (*катить*).

Семантические признаки (или слоты), которые в той или иной степени влияют на выбор средства лексического выражения, мы назовем семантическими доминантами. Мы выделяем семантические доминанты с целью имитировать механизм распределения внимания в процессе порождения высказывания, поскольку вслед за Талми мы предполагаем, что именно он отвечает за схематизацию говорящим недискретной ситуации реального мира и определяет, какие аспекты референтной сцены учитываются в ее концептуальной репрезентации. За каждым концептом закреплены свои семантические доминанты.

Ниже мы перечислим семантические признаки, которые могут быть доминантными для одних концептов и факультативными для других.

Первый из них – форма. Форма Агониста играет решающую роль при выборе лексической единицы в ограниченном числе случаев. Самый яркий пример –

предикат *катить*, который в общем случае предполагает, что Агонист круглый / шарообразный.

Далее задается описание свойств материала, из которого сделан Агонист (в случае, если он неодушевленный). Этот раздел лексического описания важен для тех ситуаций, которые потенциально влекут за собой деформацию Агониста. В других случаях эти параметры также можно опустить.

Для ряда случаев (а именно для ситуаций с переменным динамическим контуром, где Антагонист препятствует движению Агониста) важно указывать, с какой скоростью Агонист движется. На данном этапе мы предлагаем три значения для этого признака: «высокая» (high), «средняя» (medium) и «низкая» (low).

От типов объектов перейдем к характеристике действия силы. Здесь мы имеем принципиальные различия в описании для концептов деформации и концептов движения.

Мы выделяем несколько направлений приложения силы: направления вверх-вниз, вправо-влево, а также четыре диагональных направления. Вектор приложения силы и наклон поверхности, по которой происходит движение, – взаимозависимые величины.

Как мы уже говорили, внутренний семантический потенциал предикатов силового взаимодействия может быть разным: распределение значимости между признаками, характеризующими ситуацию, неодинаково, значение одного и того же признака в одних случаях выводится из контекста, в других будет фиксированным. Так, на примере глаголов сопротивления хорошо видно, что в семантической структуре этих глаголов присутствует дополнительный признак «точка зрения» или внутренний фокус эмпатии. Возьмем пару глаголов *опираться* vs *подпирать*: практически идентичная ситуация дается сначала с точки зрения Агониста, потом – с точки зрения Антагониста.

Остальные признаки, описывающие природу приложения силы, универсальны. Интенсивность приложения силы определяется по шкале от 0 до 1. От того, какое значение получает данный параметр в том или ином случае, может зависеть выбор лексемы (ср. *стукнуть* – *шарахнуть*).

Следующий признак, описывающий характер приложения силы, бинарный и имеет значения «постоянно» (permanent) или «импульсно» (impulse) (ср. *тащить* vs *толкать*).

Выделен также параметр, с помощью которого кодируется тип контакта Агониста с поверхностью, по которой происходит движение. Претерпевающий воздействие силы объект может касаться поверхности частично, только одной стороной (single_side), как для глаголов *двигать*, *тащить*, либо полностью (comprehensive), как для глаголов *катить*, *перекачивать*. Если признак имеет последнее значение, то можно определить еще один признак – количество оборотов, совершенных Агонистом, значение которого представлено численно.

Наконец, последний параметр в этом разделе описывает опосредованность контакта между Антагонистом и Агонистом, указывая на наличие или отсутствие инструмента. Его значения: direct (непосредственный) и instr (опосредованный). При выборе второго значения мож-

но указать, какой именно инструмент использован при опосредованном контакте.

В последней части лексической статьи описано взаимодействие Агониста с окружающей средой, чаще всего – с поверхностью, по которой происходит или не происходит (из-за ее физических свойств) движение. Сюда же как внешний признак попадает траектория движения Агониста.

4. ФОРМАЛЬНОЕ ПРЕДСТАВЛЕНИЕ СТРУКТУРЫ ОНТОЛОГИЧЕСКОГО КОНЦЕПТА И ЛЕКСИЧЕСКОГО ОПИСАНИЯ

Структурные элементы схем формально представлены в виде слотов. Набор слотов фиксирован для каждого из концептов и будет выглядеть следующим образом.

- I. 1. Участники ситуации:
ANT (SEM (PHYSICAL_OBJECT))
AGO (SEM (PHYSICAL_OBJECT))
2. Внутренняя силовая тенденция:
(HAS_TENDENCY (SEM ({ACTION, REST, MOTION\DEFORM})))
- II. 3. Соотношение сил:
FORCE_BALANCE (SEM ({UNKNOWN, LETTING, STRONGER (SEM (ANT)), STRONGER (SEM (AGO))}))
- III. 4. Начальное состояние Агониста:
FD_INITIAL (SEM ({REST, MOTION\DEFORM}))
5. Результат взаимодействия:
FD_RESULT (SEM ({UNKNOWN, ACTION, MOTION\DEFORM, REST}))

Необходимо оговориться, что выбор фасетов – наследие терминологии, принятой в [1]. В фигурных скобках через запятую перечислены концепты, составляющие область значений слота. За первыми двумя слотами на данном этапе закреплено значение PHYSICAL_OBJECT (физический объект) – метафоры мы пока не рассматриваем.

По законам Динамики Сил Агонист и Антагонист должны стремиться либо к изменению (движение, деформация), либо к покою. Это отражено в слоте HAS_TENDENCY. При этом тенденции Антагониста и Агониста в рамках одной схемы не могут совпадать, поэтому достаточно задать только тенденцию Агониста, а значение слота HAS_TENDENCY для Антагониста будет определено как обратное. Для фиксации этой закономерности в онтологии используем слот OPPOSITE.

Для обозначения соотношения сил между основными участниками внутри концепта вводим четыре значения слота FORCE_BALANCE: STRONGER (SEM (ANT)), STRONGER (SEM (AGO)), LETTING, UNKNOWN. Значение LETTING соответствует снятию силового воздействия, которое Антагонист оказывает на Агониста, а значение UNKNOWN показывает, что в данном концепте соотношение сил не определено.

Значения слота FD_RESULT (результат взаимодействия) показывают конечное состояние Агониста (изменение или покой).

Осталось зафиксировать только динамический контур ситуации – постоянный или переменный (т. е. меняется ли начальное состояние Агониста или нет). Для этого целесообразно ввести еще один слот, который

будет отражать начальное состояние Агониста, а вместе с ним и системы объектов в целом. Такой слот назовем FD_INITIAL. Если его значение одновременно совпадает со значением HAS_TENDENCY для Агониста и не совпадает со значением слота FD_RESULT, мы имеем дело с ситуацией силового взаимодействия с **переменным** динамическим контуром.

Общая структура лексического описания

- [1] **Participant Type**
 - [1.1] **Antagonist**
[Ant_type ANIM(anim / inanim)
PERS (pers / non-pers)
VOLITIONAL (+, -) {намеренно/
ненамеренно}]
PHYSICAL_TYPE
 - [1.2] **Agonist**
[Ago_type
ANIM(anim / inanim)
PHYSICAL_TYPE
SHAPE(shape)
SIZE (0<>1)
WEIGHT (0<>1)
MATERIAL (x, (
ELASTICITY (0<>1))
SOLIDITY (0<>1))]
{В этом разделе в описании попадает ограниченное количество параметров: акцент может делаться как на материале (одной или нескольких характеристиках), так и на форме.}
- [2] **Focus**
[Focus_Value (Ant, Ago)]
- [3] **Force Application**
[Force_Application
VECTOR (towards_Ant, from_Ant,
down, up,
towards_Ant_down,
from_Ant_down,
towards_Ant_up,
from_Ant_up, towards_Ago, from_Ago)
INTENSITY (0<>1)
APPLICATION (permanent, impulse)
CONTACT_ANT (direct, instr)
INSTR_TYPE (artifact)
CONTACT_PATH (comprehensive, single-side)-{контакт с поверхностью: по всей поверхности агониста или точка соприкосновения с поверхностью фиксирована}
PROMPTNESS(n >= 0){число оборотов}]
- [4] **Environment**
 - [4.1] **Ago_trajectory**
[Ago_trajectory
TRAJECTORY (default {= vector value},
circled, wigwag{туда-сюда}, oscillation
{вращения}, stop)
TRAJECTORY_LENGTH (n >=0)]
VELOCITY (high, medium, low, n >=0)
{скорость движения Агониста}]
 - [4.2] **Environment**
[Path/Environment
ROUGHNESS (0<>1){шероховатость}
INSTABILITY (+, -) {нестабильность}
BIAS (towards_Ant_down (0<>1),
from_Ant_down (0<>1),
towards_Ant_up (0<>1),
from_Ant_up) (0<>1)]

5. ПРИМЕР АНАЛИЗА ТЕКСТА

Учитывая специфику предметной области, в качестве текстовых примеров нам удобно использовать задачи по физике для учащихся средней школы. Ниже с помощью онтологии будет проанализировано условие задачи: *«Два бильярдных шарика, находящиеся в покое, столкнули. После столкновения они откатились в разные стороны с одинаковыми скоростями. Что можно сказать о массах этих шариков?»*

В лексиконе содержатся следующие единицы, имеющие силовое значение: *находиться в покое, столкнуть, столкновение, откатиться*. Обнаружив их, мы можем построить сценарий силовых переходов, который графически будет выглядеть так (рис. 5).

Для этого примера характерно то, что предикаты с семантикой столкновения допускают множественную интерпретацию Агониста («два бильярдных шарика»), так как оба объекта находятся в фокусе внимания и обладают необходимой подвижностью, чтобы им была присвоена эта роль. Ситуация симметрично представлена со стороны каждого из шариков, поэтому для простоты мы объединяем их одним значком на схеме. Стоит также отметить, что в данном случае мы имеем дело с упругим ударом, а значит, в качестве Антагониста при деформации будет выступать не внешний объект, а сила упругости

В заключение приведем формальную семантическую репрезентацию текста с привлечением лексикона:

«Находиться в покое»

Participant Type

Agonist (=2) («два бильярдных шарика»)

```
[Ago_type
  ANIM(inanim)
  PHYSICAL_TYPE
  SHAPE(round)
  SIZE (0,3)
  WEIGHT (x)
  MATERIAL (x, (
    ELASTICITY (0,1))
  SOLIDITY (1))]
```

Focus

```
[Focus_Value (Ago)]
```

Force Application

```
[Force_Application
  VECTOR (-)
  INTENSITY (0)
  CONTACT_PATH (single-side)]
```

Environment

Ago_trajectory

```
[Ago_trajectory
  TRAJECTORY (stop)
  TRAJECTORY_LENGTH (n=0)
  VELOCITY (n=0)]
```

Environment

```
[Path/Environment
  ROUGHNESS (0) {шероховатость}
  BIAS (0)].
```

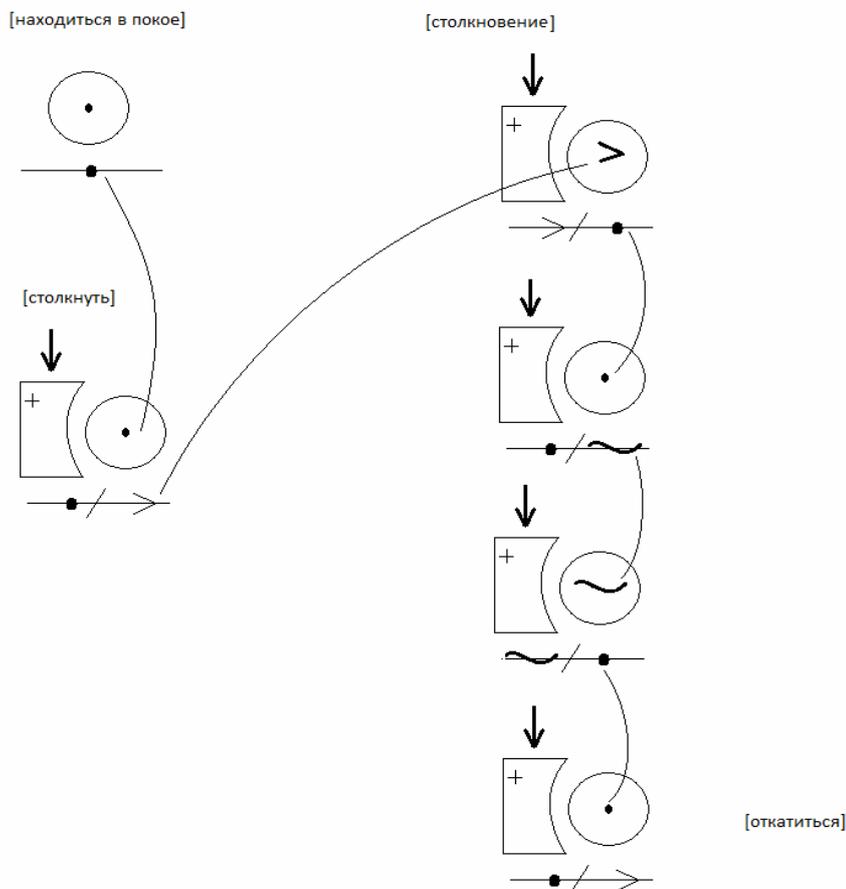


Рис. 5

«Столкнуть»

Participant Type

Antagonist ({использована неопределенно-личная форма глагола, Антагонист не определен})

[Ant_type ANIM(anim / inanim)
PERS (pers)
VOLITIONAL (+) {намеренно/
ненамеренно}]
PHYSICAL_TYPE

Agonist «два бильярдных шарика»

[Ago_type
ANIM(inanim)
PHYSICAL_TYPE
SHAPE(round)
SIZE (0,3)
WEIGHT (x)
MATERIAL (x, (
ELASTICITY (0,1))
SOLIDITY (1))]

Focus

[Focus_Value (Ago)]

Force Application

[Force_Application
VECTOR (towards_Ago)
INTENSITY (1)
APPLICATION (impulse)
CONTACT_ANT (direct)]

Environment

Ago_trajectory

[Ago_trajectory
TRAJECTORY (default {= vector
value})
TRAJECTORY_LENGTH (n)
VELOCITY (v){скорость движе-
ния Агониста}]

Environment

[Path/Environment
BIAS (horizontal)]

«Столкновение»

Antagonist

[Ant_type₁ BILLIARD BALL «биль-
ярдный шар»
Ant_type₂ INNER_ELASTIC_FORCE
{внутренняя сила упругости}]

Agonist₁ «два бильярдных шара» {харак-
теристики описаны выше}

Focus

[Focus_Value (Ago)]

Force Application

[Force_Application₁
VECTOR₁ (towards_Ago)
CONTACT_ANT₁ (direct)
Force_Application₂
VECTOR₂ (towards_Ago)
CONTACT_ANT₂ (direct)]

Ago_trajectory₂

[TRAJECTORY (stop)
TRAJECTORY_LENGTH (n)]

Environment

[Path/Environment
ROUGHNESS(0)]

«Откатиться»

Participant Type

Antagonist

[Ant_type INNER_ELASTIC_FORCE]
{внутренняя сила упругости}

Agonist «два бильярдных шара» {харак-
теристики описаны выше}

Focus

[Focus_Value (Ago)]

Force Application

[Force_Application
VECTOR (from_Ago)
APPLICATION (permanent)
CONTACT_ANT (direct)
CONTACT_PATH (comprehensive)
PROMPTNESS (n)]

[4] **Environment**

[4.1] **Ago_trajectory**

[Ago_trajectory
TRAJECTORY (default {= vec-
tor value})]

[4.2] **Environment**

[Path/Environment
ROUGHNESS(0<>0.3)
BIAS (horizontal)]

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Nirenburg S., Raskin V. Ontological semantics. – Cambridge : MIT Press, 2004.
2. Suggested Upper Merged Ontology (SUMO) [On-line]. – URL: <http://www.ontologyportal.org>
3. OpenCyc [On-line]. – URL: <http://www.opencyc.org>
4. WordNet. A lexical database for English [On-line]. – URL: <http://wordnet.princeton.edu>
5. Talmy L. Toward a cognitive semantics. Vol. 1. – Cambridge : MIT Press, 2000.
6. Марушкина А. С. «Наивная механика» в языке и онтологии // Сборник трудов международной конференции «Диалог». – М., 2006.
7. Русская языковая картина мира и системная лексикография / Апресян В. Ю., Апресян Ю. Д., Бабаева Е. Э., Богуславская О. Ю., Иомдин Б. Л., Крылова Т. В., Левонтина И. Б., Санников А. В., Урысон Е. В. ; отв. ред Ю. Д. Апресян. – М. : Языки славянской культуры, 2006.
8. Langacker R. A. Foundations of cognitive grammar. Vol. 1. – Stanford, 1987.
9. Падучева Е. В. Динамические модели в семантике лексики. – М. : Языки славянской культуры, 2004.
10. Кустова Г. И. Типы производных значений и механизмы языкового расширения. – М. : Языки славянской культуры, 2004.

Материал поступил в редакцию 10.01.12.

Сведения об авторе

МАРУШКИНА Анастасия Сергеевна – старший преподаватель Национального исследовательского университета «Высшая школа экономики», Москва
E-mail: anastasia.marushkina@gmail.com

УДК 001.102 : 001.61

В. А. Яцко

Интерпретация закона Брэдфорда в терминах геометрической прогрессии

Предлагается интерпретация закона Брэдфорда в терминах геометрической прогрессии, выявляется константа, которая позволяет упростить применение закона. Описывается методика применения закона для анализа данных, относящихся к различным предметным областям.

Ключевые слова: закон Брэдфорда, геометрическая прогрессия, статистический анализ текста

Закон рассеяния информации, предложенный С. К. Брэдфордом в 1934 г., до сих пор привлекает внимание исследователей и не утратил своей актуальности.

Как известно, закон описывает зависимость между зонами распределения научных статей по журналам [1]. Изучая литературу по геофизике, а также по проблемам создания и применения смазочных материалов, С. К. Брэдфорд обнаружил, что в зависимости от соотношения количества научных статей (Р) и журналов (J) можно выделить три зоны, в которых одно и то же количество статей распределяется по разному количеству журналов (рис. 1).

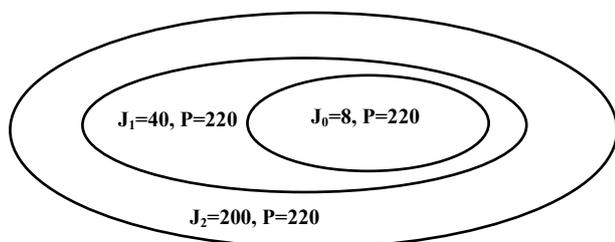


Рис. 1. Распределение научных статей по журналам в соответствии с законом Брэдфорда

Соотношение между зонами описывается константами, которые легко вычислить по следующим формулам:

$$N = J_2/J_1 = J_1/J_0 = 5$$

$$R_0 = P/J_0$$

$$R_1 = P/J_1$$

$$R_2 = P/J_2$$

$$N1 = R_0/R_1 = R_1/R_2 = 5$$

$$N = N1 = 5$$

В ряде работ неоднократно предпринимались попытки уточнить закон Брэдфорда, предложить его интерпретацию, проанализировать соотношение с другими законами, показать возможности его применения для анализа различных данных.

Наиболее известной интерпретацией закона является формула Б. Брукса [2]:

$$R(n) = k \log(n),$$

где n – ранг журнала, $R(n)$ – количество статей, опубликованных по данной теме в первых n журналах. В [3] справедливо указывается, что интерпретация Б. Брукса соотносится не с распределением журналов и статей по зонам, а с распределением статей по отдельным журналам, которым приписываются соответствующие ранги. Первый ранг получает журнал, содержащий наибольшее количество статей по данной теме. Соответственно, $R(1)$ равно количеству статей, содержащихся в журнале с первым рангом; $R(2)$ равно сумме количества статей, содержащихся в журналах с первым и вторым рангами и т. д.; k – константа, значение которой может варьироваться в зависимости от специфики предметной области. Нетрудно заметить сходство интерпретации Б. Брукса с законом Ципфа, в котором также описывается зависимость частотности от ранга. Не случайно интерпретацию, предложенную Б. Бруксом, называют распределением Брэдфорда – Ципфа [4].

Закон Брэдфорда применялся для анализа распределения научных статей по журналам в различных предметных областях, таких, как садоводство, биоэнергетика, генетика, а также для анализа цитирования в библиометрических исследованиях [5]. В целом сфера применения этого закона, по-видимому, не так обширна, как сфера применения закона Ципфа, который был разработан на лингвистическом материале, а затем стал применяться для анализа статистических данных, не имеющих ничего общего с лингвистикой, например, для анализа распределения населения по городам [6].

В настоящей статье предлагается интерпретация закона Брэдфорда в терминах геометрической прогрессии, выявляется константа, которая позволяет упростить применение закона, показывается возможность применения закона для анализа данных, относящихся к различным предметным областям.

* * *

Закон, сформулированный Брэдфордом, по сути дела, задает характерную для геометрической прогрессии числовую последовательность $b_{n+1} = b_n * q$, где b – одна из зон, содержащая n -е количество журналов, а q – знаменатель прогрессии (константа Брэдфорда). Соответственно, для интерпретации закона можно предложить формулу возрастающей геометрической прогрессии, используя J вместо b :

$$S_n = J_2(q_n - 1)/(q - 1),$$

где $n = 3$ – постоянная величина, равная количеству зон по закону Брэдфорда, q – знаменатель прогрессии, равный константе Брэдфорда, J_2 – вторая зона, характеризующаяся наибольшей степенью рассеяния информации, S_n – сумма членов геометрической прогрессии.

Если известны сумма, знаменатель и количество членов прогрессии, найти количественное значение зоны J_2 можно в результате преобразования формулы прогрессии в уравнение с одним неизвестным. Например, при $q = 3$

$$J_2 = S_n / ((q^n - 1)/(q - 1)) = S_n / ((3^3 - 1)/(3 - 1)) = S_n / (27 - 1) / 2 = S_n / 13.$$

Количественное значение остальных зон можно найти умножением на знаменатель прогрессии.

Как показал проведенный нами анализ, коэффициент, на который делится сумма при нахождении J_2 , можно найти и другим способом, выявив при этом дополнительную константу к закону Брэдфорда. Таким способом является построение числового ряда, в котором первое число равно 1, а остальные получают умножением на константу Брэдфорда (знаменатель прогрессии). Сумма трех членов числового ряда и будет составлять коэффициент, необходимый для нахождения J_2 . Такая методика позволяет также вычислить J_2 исходя из значения двух других членов прогрессии (табл. 1).

Таблица 1

Данные для определения зоны J_2

Значение q	Значение членов прогрессии	Сумма членов прогрессии	Варианты представления суммы
$q=2$	4	7	$(4*1)+3$
	2		$(2*3)+1$
	1		$1*7$
$q=3$	9	13	$(9*1)+4$
	3		$(3*4)+1$
	1		$1*13$
$q=4$	16	21	$(16*1)+5$
	4		$(4*5)+1$
	1		$1*21$

Как видно из таблицы, коэффициент, на который делится сумма для определения J_2 , можно найти исходя из значения не только первого, но и двух других членов числового ряда, применяя константу, равную $q+1$. Очевидно также, что таблицу можно продолжить для других значений q .

Предложенная интерпретация позволяет упростить вычисления и само применение закона Брэдфорда.

В статистическом анализе текста, как мы полагаем, зоне J_2 будет соответствовать диапазон с наибольшим количеством слов, имеющих наименьшую сумму частотностей. Соответственно, нулевая зона будет содержать наименьшее количество слов, имеющих наибольшую сумму частотностей. Сумму частотностей слов (S_n) в каком-либо тексте можно достаточно просто найти с помощью существующего лингвистического программного обеспечения, например, конкорданса AntConc [7].

Применение нашей методики можно продемонстрировать на примере статистического анализа корпуса, включающего произвольно выбранные тексты английских и американских писателей-классиков*, общим объемом 8 776 КБ, содержащего 24 527 уникальных токенов (рис. 2). Общая частотность токенов (т.е. величина S_n в нашей интерпретации), полученная с помощью программы AntConc, составила 1 612 296.

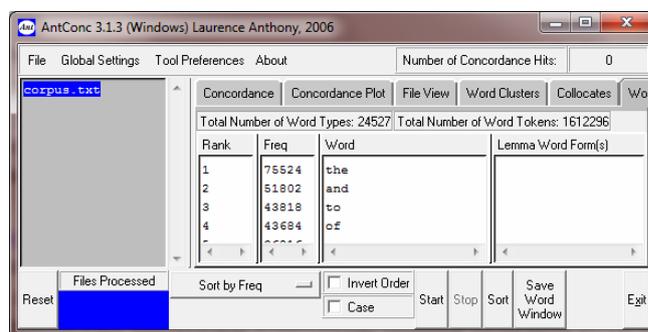


Рис. 2. Данные, полученные с помощью программы AntConc

Таким образом, были получены необходимые исходные данные, которые позволяют определить значение зоны J_2 . Как было показано ранее, при $q = 3$ получаем коэффициент 13, т.е. $J_2 = S_n / 13 = 1612296 / 13 = 124022,8$. Числовые значения остальных двух зон получаем последовательным умножением на знаменатель прогрессии:

$$J_1 = J_2 * 3 = 124022,8 * 3 = 372068,4$$

$$J_0 = J_1 * 3 = 372068,4 * 3 = 1116205,2.$$

Далее необходимо определить состав каждой из зон, установить, какие именно токены входят в каждую из них. Для этого копируем данные из AntConc в Excel (или другой табличный процессор) и выделением диапазона ячеек находим сумму числового ряда, наиболее близкую к сумме какой-либо зоны. Для нашего корпуса текстов получаем следующее распределение: 313 токенов в зоне J_0 , 3 769 токенов в зоне J_1 , 20 445 токенов в зоне J_2 .

Полученные результаты можно интерпретировать с помощью дискриминационной модели (discrimination model), разработанной Дж. Солтоном [8] на основе второго закона Ципфа [6]. В соответствии с этой моделью под дискриминирую-

* В корпус вошли следующие произведения: Jane Austen "Pride and Prejudice"; Theodore Dreiser "Sister Carrie"; John Galsworthy "The Man of Property", "In Chancery"; James Cabell "The Line of Love"; Somerset Maugham "Moon and sixpence", "The Jungle". Все тексты взяты с сайта Project Gutenberg, см.: <http://www.gutenberg.org>

щей силой понимается способность терминов уникально идентифицировать данный текст или группу текстов. Наименьшей дискриминирующей силой обладают наиболее высокочастотные и наиболее низкочастотные термины, а наибольшей дискриминирующей силой – термины со средней частотностью. Удаление или игнорирование высокочастотных или низкочастотных терминов позволяет существенно уменьшить размер индекса информационно-поисковой системы, повисить ее быстродействие, а также точность поиска.

Экстраполируя эти выводы на полученные нами результаты, можно утверждать, что в зону J_0 входят термины, встречающиеся с одинаково высокой частотностью в различных текстах. К ним относятся артикли, предлоги, союзы, местоимения и другие служебные слова, которые в информационном поиске рассматриваются как стоп-слова. В зону J_2 входят наиболее низкочастотные слова, к которым могут относиться слова с опечатками, ошибками, неологизмы и вообще редко используемые термины. Наконец, зона J_1 включает термины, которые встречаются в разных текстах, однако их распределение по частотности уникально для данного текста.

В наших рассуждениях недостаточно очевидным является применение коэффициента $q = 3$. В работах, посвященных закону Брэдфорда, отсутствуют однозначные критерии для определения значения этого коэффициента и обычно утверждается, что этот коэффициент зависит от специфики поиска [3, с. 44]. Мы полагаем, что можно применить несколько методов верификации предложенного значения данного коэффициента:

1. Соотнесение терминов в зоне J_0 со списками стоп-слов. Совпадение терминов из этой зоны и терминов из списков стоп-слов указывает на адекватность применяемого значения коэффициента. В нашем корпусе совпадение терминов из зоны J_0 со списком стоп-слов [9] составило 78%, что свидетельствует о достаточной адекватности применяемого коэффициента. Вместе с тем, применение коэффициента $q = 5$ приводит к расширению зоны до 860 терминов и размыванию плотности стоп-слов до 47%, что также свидетельствует в пользу применяемого критерия.

2. Тестирование зоны J_2 на наличие терминов с опечатками, ошибками, а также неологизмов. Наличие данных терминов указывает на адекватность применения коэффициента.

3. Тестирование зоны J_1 на дискриминирующую силу и уникальность идентификации текста. Построение распределения терминов зоны и сопоставление этого распределения с распределением аналогичной зоны другого текста. Разница распределений терминов в зонах двух текстов будет указывать на дискриминирующую силу коэффициента.

В некоторых случаях количественный состав зоны J_2 бывает известен, что позволяет точно определить значение коэффициента q . Так, после прошедших 4 декабря 2011 г. выборов в Государственную Думу Президент Д. А. Медведев заявил: «Я считаю, что нужно обратить внимание на те регионы, где наши люди отказали в серьезном доверии "Единой России"» [10]. В результате руководители ряда регионов

были вызваны в Администрацию Президента для обсуждения результатов выборов. По разным данным количество вызванных глав регионов составляло от 15 до 17 человек [11], что и составляет количество соответствующих регионов в зоне J_2 . Результаты голосования за партию «Единая Россия» по регионам, представленные в табл. 2, позволяют достаточно просто определить значение коэффициента q : при $S_n = 3920,44$ и $q = 2$ числовое значение зоны $J_2 = S_n / 7 \approx 560,06$, что и составляет пороговое значение для определения количественного состава зоны J_2 . Такое число получается в результате суммирования процентов голосов, отданных за партию «Единая Россия» в последних 17 регионах, указанных в таблице, начиная с Новгородской области.

Таблица 2

**Результаты голосования за партию
«Единая Россия» на выборах в Государственную
Думу 4 декабря 2011 г. по регионам
(по данным сайта <http://vybory2011.yandex.ru>)**

№	Регион	Количество голосов, отданных за партию «Единая Россия», %	Зона
1.	Чеченская Республика	99,48	J ₀
2.	Республика Дагестан	91,84	
3.	Республика Мордовия	91,75	
4.	Республика Карачаево-Черкесия	89,82	
5.	Республика Тыва	85,29	
6.	Республика Кабардино-Балкария	81,31	
7.	Республика Татарстан	77,85	
8.	Тюменская область	74,13	
9.	Ямало-Ненецкий АО	71,68	
10.	Республика Ингушетия	70,8	
11.	Республика Башкирия	70,5	
12.	Чукотский АО	70,32	
13.	Республика Северная Осетия-Алания	68,4	
14.	Тамбовская область	66,66	
15.	Республика Калмыкия	66,1	
16.	Саратовская область	65,13	
17.	Кемеровская область	64,53	
18.	Республика Адыгея	61,09	

№	Регион	Количество голо- сов, отданных за партию «Единая Россия», %	Зона	
19.	Республика Коми	58,79		
20.	Краснодарский край	56,65		
21.	Пензенская область	56,19		
22.	Астраханская область	54,6		
23.	Республика Алтай	53,33		
24.	Республика Марий Эл	52,54		
25.	Брянская область	50,12		
26.	Ростовская область	50,22		
27.	Еврейская АО	49,9		
28.	Воронежская область	49,54		
29.	Республика Саха (Якутия)	49,16		
30.	Республика Бурятия	49,02		
31.	Ставропольский край	48,95		
32.	Москва	46,48		
33.	Челябинская область	45,98		
34.	Курская область	45,72		
35.	Камчатский край	45,24		
36.	Республика Удмуртия	45,09		J ₁
37.	Нижегородская область	45,05		
38.	Курганская область	44,41		
39.	Ульяновская область	43,56		
40.	Амурская область	43,53		
41.	Республика Чувашия	43,46		
42.	Забайкальский край	43,28		
43.	Сахалинская область	41,91		
44.	Липецкая область	41,17		
45.	Магаданская область	41,04		
46.	Ханты- Мансийский АО	41,02		
47.	Калужская область	40,46		
48.	Республика Хакасия	40,43		
49.	Тульская область	40,12		

№	Регион	Количество голо- сов, отданных за партию «Единая Россия», %	Зона	
50.	Рязанская область	39,82		
51.	Омская область	39,61		
52.	Орловская область	38,97		
53.	Тверская область	38,44		
54.	Псковская область	38,88		
55.	Хабаровский край	38,11		
56.	Самарская область	38,1		
57.	Томская область	37,39		
58.	Калининградская область	37,07		
59.	Алтайский край	37,21		
60.	Волгоградская область	36,77		
61.	Красноярский край	36,67		
62.	Пермский край	36,33		
63.	Смоленская область	36,23		
64.	Новгородская область	36,13		J ₂
65.	Ненецкий АО	36,04		
66.	Иркутская область	34,93		
67.	Кировская область	34,9		
68.	Оренбургская область	34,89		
69.	Костромская область	34,19		
70.	Новосибирская область	33,84		
71.	Ленинградская область	33,77		
72.	Санкт-Петербург	32,47		
73.	Вологодская область	33,49		
74.	Приморский край	33,47		
75.	Московская область	32,71		
76.	Свердловская область	32,66		
77.	Республика Карелия	32,26		
78.	Архангельская область	31,81		
79.	Владимирская область	30,6		
80.	Ярославская область	29,04		

Итак, в соответствии с нашей интерпретацией, закон Брэдфорда описывает зависимость между зонами

распределения информации по источникам, которая описывается константой и может быть представлена в виде монотонно возрастающей или убывающей последовательности. Следовательно, как мы полагаем, данный закон можно интерпретировать в терминах геометрической прогрессии, где константа Брэдфорда будет являться знаменателем, а количество членов прогрессии будет равно количеству зон. Такая интерпретация позволяет применить для вычислений по закону Брэдфорда традиционную формулу геометрической прогрессии и упростить вычисление количественного значения трех зон, выделяемых в соответствии с законом.

Предлагаемая нами методика включает следующие процедуры:

1. Вычисление суммы информационного числового ряда (S_n), например, суммы частотностей терминов в тексте или суммы данных о голосовании на выборах.

2. Определение с помощью данных таблицы 1 числового значения зоны J_2 . Данное значение зависит от значения коэффициента Брэдфорда q (знаменателя прогрессии), которое определяется с помощью эмпирической верификации.

3. Определение числового значения зон J_1 и J_0 , которое выполняется последовательным умножением на коэффициент q .

4. Определение количественного состава источников в каждой из зон (например, количества терминов, слов, регионов), которое выполняется путем вычисления диапазона, дающего сумму числового ряда, соответствующую той или иной зоне.

5. Интерпретация результатов. Результаты могут быть использованы для сокращения индекса информационно-поисковых систем за счет учета только терминов, занимающих позиции в центральной зоне (зоне J_1), а также с целью автоматической классификации и категоризации текстов, поскольку распределения терминов по частотности в каждой из зон могут уникально характеризовать данную категорию, жанр или группу текстов.

Предложенная интерпретация позволяет, как мы полагаем, расширить сферу применения закона Брэдфорда.

СПИСОК ЛИТЕРАТУРЫ

1. Hjørland B., Nicolaisen J. Bradford's law of scattering: ambiguities in the concept of «subject» //

5th Conference on concepts of library and information science. – 2005. – P. 96 – 106. – (Lecture notes in computer science ; № 3507); см. также: <http://comminfo.rutgers.edu/~muresan/IR/Docs/Articles/colisHjorland2005.pdf>

2. Brookes B. C. The derivation and application of the Bradford-Zipf distribution // Journal of Documentation. – 1968. – Vol. 24, Issue 4. – P. 247 – 265.
3. Drott M. C. Bradford's law: theory, empiricism and the gaps between // Library trends. – 1981. – Vol. 30, №1. – P. 41 – 52.
4. Hjørland B. Scattering [On-line]. – 2007. – URL: <http://www.iva.dk/bh/core%20concepts%20in%20lis/articles%20a-z/scattering.htm>
5. Sudhier K. G. Application of bradford's law of scattering to the physics literature: a study of doctoral theses citations at the Indian Institute of Science // Journal of library & information technology. – 2010. – Vol. 30, №2. – P. 3 – 14.
6. Li W. Zipf's Law Everywhere // Glottometrics. – 2002. – № 5. – P. 14 – 21.
7. См.: <http://www.antlab.sci.waseda.ac.jp/software.html>
8. Salton G., Yang C. S., Yu C. T. A theory of term importance in automatic text analysis // Journal of the American Society for Information Science. – 1975. – Vol. 26, Issue 1. – P. 33 – 44.
9. Stop Word List 1 [On-line]. – URL: <http://www.lextek.com/manuals/onix/stopwords1.html>
10. Цит. по: <http://www.rb.ru/topstory/politics/2011/12/08/102654.html>
11. Губернаторов вызвали в Кремль [Электрон. ресурс]. – URL: <http://www.postsovet.ru/blog/russia/292101.html>

Материал поступил в редакцию 31.01.12.

Сведения об авторе

ЯЦКО Вячеслав Александрович – доктор филологических наук, профессор, профессор Хакасского государственного университета им. Н. Ф. Катанова, г. Абакан

E-mail: viatcheslav-yatsko@rambler.ru

УДК 81'322 : [811.18 + 811.432.853.2 + 811.351.32 + 811.221.18]

Т. А. Архангельский

Электронные корпуса албанского, калмыцкого, лезгинского и осетинского языков

Представлены четыре электронных корпуса, созданные в 2011 г. в рамках Программы фундаментальных исследований РАН «Корпусная лингвистика»: албанский, калмыцкий, лезгинский и осетинский. Дается описание интерфейса и функциональности этих корпусов, освещаются технические вопросы, которые пришлось решать при их создании, обсуждаются перспективы их развития. Особое внимание уделяется вопросам составления грамматических словарей и автоматической грамматической разметки корпусов.

Ключевые слова: корпус, корпусная лингвистика, морфологический парсер, албанский язык, лезгинский язык, калмыцкий язык, осетинский язык

1. ВВЕДЕНИЕ

Лингвисту при изучении какого-либо языкового объекта или явления – лексем, грамматической категории, конструкции и т. д. – часто бывает необходимо получить большое количество примеров этого явления в исследуемом языке. Традиционный способ получения языковых примеров – опрос информантов – имеет ряд очевидных недостатков, например, невозможность моментально получить ответ от информанта в любое время, необходимость опроса большого количества информантов для получения информации о языке в целом, а не об идиолекте одного человека, и, возможно, самое главное – неспонтанность получаемых данных, от-рефлексируемый вариант языка, данные не столько о реальном использовании языка, сколько о языке в представлении информанта. Для решения этих проблем в последние два десятилетия лингвистами активно создаются специальные электронные инструменты – корпуса языков.

Корпусом языка называется собрание текстов на этом языке, в котором текстам или их фрагментам (абзацам, предложениям, словоформам или даже морфемам) приписана дополнительная лингвистически релевантная информация (аннотация) и которое снабжено поисковым механизмом, позволяющим производить поиск по этой информации. Аннотация может включать в себя любую информацию, в зависимости от задач, стоящих перед создателями корпуса и исследователями. При этом корпус, в отличие от электронной библиотеки, предназначен не для просмотра полных текстов имеющихся в нём произведений (хотя в некоторых корпусах такая возможность есть). Основная задача корпуса – предоставить исследователю языка возможность быстро получать реальные языковые примеры по заданному запросу и

выяснять относительную частоту появления в текстах языковых объектов, соответствующих этому запросу. Содержание поисковых запросов, которые позволяет делать корпус, зависит от того, какая дополнительная информация в него внесена.

Среди информации, которой могут снабжаться тексты корпуса, следует в первую очередь выделить метатекстовую информацию и пословную разметку. Метатекстовая информация – это информация, которая приписывается тексту в целом. Она может включать в себя имя автора, время создания текста, размер текста и т. п. (например, ср. описание метатекстовой разметки Национального корпуса русского языка в [1]). Пословная разметка, т. е. дополнительная информация, указываемая при каждой словоформе текста, содержит прежде всего грамматическую информацию, но также может включать перевод на другой язык, информацию о тех или иных семантических признаках словоформы и др. Грамматическая информация может включать в себя лемму (начальную форму), часть речи, словоклассифицирующие и словоизменяемые грамматические признаки.

В то время как для многих крупных языков Европы корпуса существуют и давно используются при проведении лингвистических исследований (например, английские British National Corpus <<http://www.natcorp.ox.ac.uk/>> и Corpus of Contemporary American English <<http://corpus.byu.edu/coca/>>, Национальный корпус русского языка <<http://www.ruscorpora.ru/>>, Чешский национальный корпус <<http://ucnk.ff.cuni.cz/>>, Венгерский национальный корпус <http://corpus.nytud.hu/mnsz/index_eng.html>), для большинства менее распространённых языков, в том числе для большинства языков России, корпусов не существует. В 2011 г. в рамках программы

фундаментальных исследований Президиума РАН «Корпусная лингвистика» были созданы, среди прочего, корпуса албанского, калмыцкого, лезгинского и осетинского (иронский диалект) языков*. Эти корпуса имеют схожую функциональность, поскольку созданы на одной платформе и с использованием одних и тех же принципов и одной и той же компьютерной системы грамматической разметки. В настоящее время корпуса размещены на сайтах ossetic-studies.org (осетинский) и web-corpora.net (остальные); все корпуса доступны по ссылкам со страницы <http://web-corpora.net/>.

В разделе 2 настоящей статьи будет дана общая характеристика созданных корпусов с описанием их функциональности. В разделе 3 будет кратко описан процесс создания корпуса и особо рассмотрена его важная часть – создание системы морфологического анализа. В последнем разделе будут изложены перспективы дальнейшей работы над этими корпусами.

2. ОБЩАЯ ХАРАКТЕРИСТИКА КОРПУСОВ

Все упомянутые корпуса включают письменные тексты на нормативном варианте соответствующего языка. Приблизительные объёмы корпусов таковы: албанский – 750 тыс. словоупотреблений, калмыцкий и лезгинский – по 800 тыс., осетинский – 5 млн. На данный момент эти корпуса уступают в объёме корпусам крупных европейских языков (например, объём Национального корпуса русского языка в настоящий момент составляет почти 200 млн словоупотреблений) и добавление новых текстов является одним из приоритетов их развития. Однако стоит учесть, что корпуса письменных текстов на малых и средних языках не могут достичь таких показателей, как, например, корпус русского языка, по той причине, что такого количества текстов на этих языках просто не существует. В корпуса включены созданные в XX – XXI вв. художественные произведения, мемуары, очерки и некоторые периодические издания (последние – в осетинский и калмыцкий).

Все корпуса снабжены базовой метатекстовой и подробной грамматической разметкой. Метатекстовая информация включает в себя название текста, имя автора или название периодического издания, год написания (или временной промежуток, если точная дата неизвестна или написание текста заняло более одного года) и жанр. Грамматическая разметка включает в себя начальную форму слова и информацию обо всех грамматических категориях, выраженных в данной словоформе морфологически. Грамматическая разметка производилась программными средствами (см. ниже) и имеется более чем у 70% словоформ в каждом корпусе. При подобной автоматической разметке одна и та же словоформа может получить несколько разных грамматических разборов (например, албанское *nga* может быть предлогом, союзом или одной из двух форм глагола *ngas*). Снятие омонимии, т. е. выбор для каждой такой слово-

формы в каждом её контексте в корпусе одного из предложенных разборов, является довольно сложной и времязатратной задачей и в данных корпусах не проводилось, поэтому любая подобная словоформа во всех контекстах будет иметь все присвоенные ей разборы (ср. подкорпус со снятой омонимией в Национальном корпусе русского языка).

Все корпуса используют одну и ту же платформу (поисковую систему и веб-интерфейс), которая изначально была создана компанией «Corpus Technologies» для Восточноармянского национального корпуса <http://www.eanc.net/> [2]. Эта платформа обладает мощным поисковым функционалом, позволяющим задавать сложные контекстные запросы, использовать метаинформацию для формирования пользовательских подкорпусов (т. е. подмножеств текстов, в которых осуществляется поиск, например: только периодические издания или только тексты до 1980 г.) и настраивать вид получаемых результатов по многим параметрам.

Основным инструментом поиска является запрос по словоформе, лемме и грамматической информации. При выборе вкладки «Словоформа» будут найдены все вхождения введённой словоформы, т. е. все контексты, где встречается введённая в поисковое поле последовательность символов. При выборе вкладки «Лемма» будут найдены все вхождения всех словоформ, которые были размечены как формы лексем, лемма которой совпадает с введённой последовательностью символов. Таким образом, в отличие от поиска по словоформе, при поиске по лемме могут быть найдены только те словоформы, которым был приписан грамматический разбор. При задании искомой словоформы или леммы можно использовать знак «*», означающий любое количество любых символов, знак дизъюнкции «|», знак конъюнкции «&» и знак отрицания «~». Например, по запросу «*mi*&~miḡe*» в албанском корпусе будут найдены все вхождения всех словоформ, начинающихся на *mi-*, но не совпадающих со словоформой «*miḡe*» («добро»), а по запросу «**tḡy|*tḡl*» в осетинском корпусе будут найдены все словоформы, оканчивающиеся на *-tḡy* или *-tḡl*. В случае использования нескольких логических операций их порядок можно задавать при помощи скобок. Все специальные знаки, включая звёздочку, можно использовать в запросе неограниченное число раз, однако высокая сложность запроса может привести к существенному увеличению времени поиска, которое в случае запроса обычной сложности составляет несколько секунд.

Кроме поиска по лемме и словоформе возможен поиск по переводу словоформы на другой язык, что удобно для исследователей, не владеющих исследуемым языком на достаточном уровне. В албанском корпусе размеченные словоформы снабжены переводом на английский язык, в калмыцком и лезгинском – на русский, а в осетинском – как на английский, так и на русский. При поиске по переводу можно использовать те же знаки, что и при поиске по лемме/словоформе.

В поле «Грамматика и части речи» можно задать запрос на грамматические признаки словоформы. Функциональность поиска по грамматическим кате-

* В разработке албанского корпуса принимали участие М. С. Морозова, М. В. Домосилецкая, А. Ю. Русаков и Е. Д. Бернацкая, калмыцкого – А. Э. Ванькаева, лезгинского – Д. С. Ганенков, осетинского – О. И. Беляев и А. П. Выдрин.

гориям в целом аналогична функциональности Национального корпуса русского языка. В данное поле помещаются краткие обозначения грамматических категорий, которые можно как ввести с клавиатуры, так и выбрать из списка. В поле грамматических категорий, как и в поле ввода леммы/словоформы, можно использовать знаки трёх логических операций, комбинируя наборы грамматических показателей искомым словоформ. Запрос на грамматические признаки можно комбинировать с запросом на лемму или словоформу. Так, если в калмыцком корпусе ввести «*д» в поле «Лемма» и «N,ins» в поле «Грамматика и части речи», будут найдены все вхождения существительных в творительном падеже, начальная форма которых заканчивается на -д.

Существует также ряд дополнительных условий, которые можно наложить на искомую словоформу или её ближайший контекст, раскрыв группу полей под заголовком «Дополнительно». С помощью этих полей можно потребовать наличия какого-либо знака пунктуации слева и/или справа от словоформы или отсутствия какой бы то ни было пунктуации, ограничить позицию словоформы в предложении (в начале, в конце, в середине), регистр словоформы (с прописной буквы, все прописные, все строчные, как введено), а также разрешить или запретить находить словоформы, имеющие несколько омонимичных разборов.

Поиск можно проводить одновременно по нескольким словоформам, определённым образом расположенным друг по отношению к другу в контексте, иными словами, можно искать словосочетания и конструкции. Для этого, задав нужное количество запросов на словоформы (по умолчанию в панели создания запроса включены поля для двух словоформ, но их количество можно увеличивать), необходимо указать расстояния между вхождениями этих словоформ в тексте. Расстояния можно указывать в виде диапазона, а также использовать отрицательные значения. Например, если в лезгинском корпусе указать в грамматике первой словоформы «N», в грамматике второй – «A» и установить расстояние между ними от -1 до 1, будут найдены все сочетания «прилагательное + существительное» и «существительное + прилагательное». Вместо указания точного числового диапазона в словах можно потребовать, чтобы искомые словоформы находились в одном предложении, в том числе на любом расстоянии, или в одном документе (с указанием диапазона расстояний в предложениях).

Выдача результатов возможна в нескольких представлениях: полным, кратком, KWIC (Key Word In Context) и глоссированном. Основной особенностью представления KWIC является то, что ключевая словоформа каждого найденного предложения отцентрована. В глоссированном формате под каждой словоформой текста подписаны её начальная форма и грамматические характеристики (при этом разбиение словоформы на морфемы не производится). Имеется возможность настроить количество результатов на одной странице выдачи и порядок сортировки примеров (по умолчанию примеры выдаются в случайном порядке). В лезгинском, осетинском и калмыц-

ком корпусах можно выбрать алфавит выдачи: кириллица или транслитерация.

При выводе результатов поиска показываются сами найденные примеры, их общее количество и количество документов, в которых они были найдены. При каждом примере указывается название текста, автор или название периодического издания и год создания текста. У каждой словоформы, имеющей хотя бы один грамматический разбор, во всплывающем окне указывается лемма, грамматические признаки, разделённые на словоклассифицирующие и словоизменяемые, и перевод. Каждый пример содержит одно предложение текста; контекст можно расширять. Максимальный размер расширенного контекста составляет 7 предложений (3 предложения слева от найденного, само найденное предложение и 3 предложения справа). Контексты большего размера не предоставляются по юридическим причинам: многие тексты корпуса являются объектом авторского права, и размещение больших фрагментов текста могло бы быть признано его нарушением.

Документы, по которым производится поиск, можно ограничить с помощью опции «Задать подкорпус». В частности, можно ограничить подкорпус текстами определённых жанров, авторов, лет или просто выбрать тексты из списка.

Интерфейсы всех корпусов выполнены на русском и английском языках, а интерфейс албанского корпуса переведён также на албанский язык.

3. СИСТЕМА МОРФОЛОГИЧЕСКОГО АНАЛИЗА

Создание корпуса включает несколько основных этапов. На первом этапе необходимо собрать тексты, включаемые в корпус. Источниками текстов могут быть Интернет, издательства, отсканированные книги и расшифровки устной речи. В рассматриваемых корпусах использовались первые два источника. На втором этапе необходимо внести в собранные тексты грамматическую информацию. В случае корпусов небольшого размера (десятки тысяч слов) это можно сделать вручную. Однако для более крупных корпусов этот вариант практически неприемлем, поскольку требует очень больших ресурсов. Поэтому для разметки текстов необходимо разработать систему автоматического морфологического анализа. Третьим этапом является создание или адаптация существующей поисковой платформы и размещение корпуса в Интернете.

Разработка системы морфологического анализа является одним из центральных этапов создания корпуса. Задача создания такой системы для данного языка состоит из нескольких частей. Первым шагом является описание морфологической системы языка, т. е. описание имеющихся в нём частей речи, грамматических значений и способов их выражения, описание классов лексем с одинаковым словоизменением (например, склонений существительных) и т. п. Для рассматриваемых языков эта задача была выполнена экспертами – разработчиками соответствующих корпусов. Вторым шагом является создание грамматического словаря, включающего в себя словоизменяемые таблицы и перечень лексем с указанием их

словоизменительного типа, основы/основ и всей остальной информации, необходимой для построения полной парадигмы. Примером такого словаря является созданный ещё в 1970-е гг. и многократно переизданный «Грамматический словарь русского языка» [3]. Однако для рассматриваемых языков подобных словарей не существовало. Третьим шагом является создание парсера – программы, способной размечать текст, используя данные грамматического словаря.

В ходе работы над проектами их участниками были созданы электронные грамматические словари рассматриваемых языков (на основе существующих грамматик, переводных словарей и других материалов). На данный момент словари носят пилотный характер и содержат от 4 до 15 тыс. лексем.

В качестве парсера во всех корпусах использовалась разработанная автором настоящей статьи система UniParser. Эта система одинаково подходит как для флективных языков, какими являются албанский и, частично, осетинский, так и для агглютинативных, какими являются лезгинский и в особенности калмыцкий. Система UniParser принимает на вход грамматический словарь в специальном формате, основанном на YAML, и использует его для разметки текстов. Кроме того, она может по запросу порождать полные парадигмы имеющихся в словаре лексем (листинг), что удобно для проверки правильности заполнения словаря и присваивания словоизменительных классов.

Грамматический словарь, предназначенный для использования в этом парсере, состоит из нескольких файлов, самыми важными из которых являются список словоизменительных типов, список лексем и список продуктивных дериваций. Список словоизменительных типов предназначен для описания изменяющейся части слов, в то время как список лексем содержит основу или основы каждой лексемы со ссылкой на её парадигму и другую лексическую информацию.

Каждый словоизменительный тип имеет название и представляет собой перечисление флексий (морфем или комбинаций морфем), покрывающее всю парадигму. При каждой флексии указывается набор выражаемых ею грамматических значений. Если лексемы соответствующего словоизменительного типа имеют несколько основ, то при флексии указывается номер основы, с которой она употребляется (нумерация начинается с нуля). Ниже приведён фрагмент описания одной из именных парадигм албанского языка, в котором перечислено несколько флексий множественного числа (неопределённый номинатив: пустая флексия, неопределённый генитив: *-ve*, определённый номинатив: *-t*, определённый генитив: *-ve*), каждая из которых употребляется с первой основой лексемы:

```
-paradigm: 5a
-flex: <0>.
  gramm: indef,pl,nom
-flex: <0>.ve
  gramm: indef,pl,gen
-flex: <0>.t
  gramm: def,pl,nom
-flex: <0>.ve
  gramm: def,pl,gen
```

Хотя в рассматриваемых языках подавляющее большинство грамматических значений выражается аффиксами справа от основы или изменениями в основе, формат описания словоизменения подходит и для языков, в которых грамматические значения выражаются аффиксами слева от основы, полиаффиксами (составными аффиксами, имеющими несколько несмежных частей, в том числе располагающихся с обеих сторон от основы), трансфиксами и т. п. В случае флективных языков с небольшими парадигмами (албанский) в каждом типе перечисляются все флексии. Для облегчения работы с агглютинативными языками, такими, как калмыцкий, где глагольная парадигма состоит из сотен форм, а на границах аффиксов практически не происходит контактных изменений, существует возможность задать парадигму комбинацией нескольких элементарных подпарадигм. Так, парадигму калмыцкого глагола можно представить в виде комбинации последовательно идущих подпарадигм, одна из которых включает в себя только показатели каузатива, другая – только аспектуальные показатели, третья – только личностно-числовые показатели и т. п. (своего рода аналог «грамматики позиций»).

Каждая запись, содержащая информацию о лексеме, содержит лемму, основу или несколько основ, название словоизменительного типа, перевод и любую другую лексическую информацию. Лексема может иметь несколько ссылок на словоизменительные типы (например, в албанском грамматическом словаре парадигмы множественного и единственного числа были описаны как разные парадигмы и, соответственно, у каждого существительного отдельно указывался его словоизменительный тип в единственном и во множественном числе). Основ также может быть несколько; основы разделяются знаком «|». Как указывалось выше, в случае, если лексема имеет несколько основ, дополнительно распределённых в парадигме, при каждом аффиксе соответствующего словоизменительного типа должен быть указан номер основы, с которой он употребляется. Ниже приведено описание лексемы *vajzë* «девочка» из албанского словаря:

```
-lexeme
lex: vajzë
stem: vajza.|vajzë.|vajz.
gramm: N,f,anim
paradigm: 3a
paradigm: 5a
transl_en: girl
```

Эта лексема имеет три основы. В описании также указаны грамматическая информация (существительное, женский род, одушевлённое), словоизменительный тип (парадигма 3a содержит флексии единственного числа, парадигма 5a, часть которой приведена выше, — множественного), начальная форма и английский перевод. Каждая основа лексемы представляет собой слово, в котором точками обозначаются места присоединения аффиксов, в то время как в аффиксах точка используется для обозначения частей основы (т. е. своего рода «точек сцепления» основы и показателя). При построении сло-

воформы парсер соединяет нужные основу и аффикс, вставляя на место точек в основе нужные части аффикса и наоборот. Например, определённый номинатив множественного числа данной лексемы будет образован сложением первой основы лексемы с соответствующим аффиксом: *vajza. + t → vajzat*. Формат предусматривает также специальные средства для описания вариативности основ и аффиксов.

Одной из важных особенностей системы UniParser является возможность описания продуктивных деривационных схем. Деривационная схема представляет собой правило, по которому из имеющейся в словаре лексемы можно автоматически получить другую лексему. В деривационной схеме описываются изменения, вносимые в основу, лемму и грамматическую информацию исходной лексемы, и даётся ссылка на словоизменительный тип деривированной лексемы. При этом лексические признаки деривированной лексемы могут как замещать лексические признаки исходной лексемы, так и дополнять их, а лемма деривированной лексемы может быть образована по заданным правилам от её основы или совпадать с леммой исходной лексемы.

Перечисленные особенности позволяют использовать деривационные схемы для двух типов задач. В первом случае с их помощью автоматически порождаются и заносятся в грамматический словарь новые лексемы. Например, в осетинском языке имеется продуктивный способ образования прилагательного от существительного с помощью суффикса *-он*. После занесения соответствующей информации в список деривационных схем с указанием, что подобная схема может применяться ко всем существительным, словарь пополняется прилагательными на *-он*, образованными от каждого имеющегося в словаре существительного, что позволяет не заносить такие прилагательные в словарь в виде отдельных записей и тем самым существенно экономить усилия по составлению грамматического словника. Ниже приведён другой, несколько более сложный пример такого рода из осетинского словаря – образование перфективной формы глагола с помощью префикса *ных-*:

```
-deriv-type: V-nyx
lex: <0>ных[.]ын
stem: ных[.].
regex-stem: x[^ъ].*
gramm: +pv,pv-ny
```

Точкой в квадратных скобках обозначается основа деривируемого глагола, точка без скобок имеет то же значение, что и в описании лексемы, – место присоединения словоизменительных аффиксов. В этой деривационной схеме указано, что начальная форма деривированной лексемы образуется присоединением префикса *ных-* и инфинитивного суффикса *-ын* к первой основе исходной (деривируемой) лексемы; основа или несколько основ деривированной лексемы образуются присоединением префикса *ных-* к основам исходной лексемы. К грамматическим параметрам исходной лексемы добавляются значения *pv* («глагол с провербом») и *pv-ny* («глагол, содержащий один из алломорфов проверба *ны-*»), что необходимо для обеспечения

возможности поиска в корпусе глаголов с провербами. В поле *regex-stem* с помощью языка регулярных выражений задано условие, ограничивающее область действия деривации только теми лексемами, основа которых начинается на *x-*, но не на *хъ-*. В ходе обработки грамматического словаря парсер, применяя это деривационное правило, например, к глаголу *хауын* «падать; иметь отношение», автоматически добавляет в словарь его перфективный коррелят *ныххауын*.

Во втором случае деривационные схемы можно использовать для описания наборов форм, которые традиционно интерпретируются как принадлежащие к одной парадигме. Пример такого типа в русском языке – причастные формы глаголов. Хотя традиционная грамматика считает причастие формой глагола, с точки зрения экономичности описания удобнее представлять его в виде регулярного отглагольного прилагательного, задав соответствующую деривационную схему. Такое представление позволяет избежать описания всех причастных форм в каждой глагольной парадигме. При этом деривированные формы можно считать принадлежащими парадигме исходной лексемы, для этого достаточно указать в деривационной схеме, что лемма деривированной лексемы совпадает с леммой исходной (таким образом, при поиске по этой лемме будут найдены все формы: и исходные, и деривированные). Такой тип дериваций был использован, например, в калмыцком корпусе для описания субстантивного словоизменения причастий.

Получив на вход словарь и грамматику языка, описанные в указанных файлах, парсер строит базу данных, на основе которой в дальнейшем можно производить морфологическую разметку текстов. При разметке парсер принимает на вход текстовый файл, а результатом работы является файл на XML-языке, приближенном к используемому в Национальном корпусе русского языка. Ниже приведён фрагмент размеченного албанского текста:

```
<w><ana
lex="hedh"
gr="V, 3, sg, aor, ind, act"
transl_en="throw"></ana>hodhi</w>
```

```
<w><ana
lex="sy"
gr="S, m, inanim, def, pl, acc"
transl_en="eye"></ana><ana
lex="sÿ"
gr="S, m, inanim, def, pl, nom"
transl_en="eye"></ana>sytë</w>
```

Внутри тэга *<w>* заключена одна словоформа и грамматическая информация о ней. Последняя содержится в атрибутах тэга *<ana>*, например: *lex* – начальная форма, *gr* – перечисленные через запятую грамматические пометы, *transl_en* – перевод на английский язык. Если парсер присвоил лексеме несколько альтернативных разборов, словоформа будет содержать несколько тэгов *<ana>*, как во второй словоформе из примера.

4. ПЕРСПЕКТИВЫ РАЗВИТИЯ

Одним из главных направлений дальнейшей работы над корпусами является увеличение текстовой базы и расширение грамматических словарей.

При сборе новых текстов важен не только их объём, но и тип. Для того чтобы корпус был репрезентативным, т. е. действительно мог представлять язык во всех его вариантах, в нём должны быть представлены тексты разных стилей и жанров: художественные и нехудожественные; переводные и непереводные; поэзия и проза; пресса, официальные документы, тексты из интернет-форумов и т. д., – причем желательно, чтобы ни один из типов текстов не имел существенного количественного преобладания над остальными. Ясно, что не для всех перечисленных языков существуют тексты всех этих жанров, но стремиться разнообразить состав корпуса, несомненно, следует.

Расширение грамматических словарей позволит повысить качество грамматической разметки. В настоящее время грамматические словари для перечисленных языков позволяют разбирать не менее 70% встречающихся словоупотреблений. Если для достижения этого результата хватает словарей, имеющих не более 15 тыс. лексем, то для дальнейшего повышения качества грамматической разметки это количество необходимо будет существенно увеличить. Поскольку в целом лексемы заносятся в словарь в порядке убывания их частотности, каждая следующая внесённая лексема будет иметь всё меньшее влияние на общее количество разобранных словоформ в корпусе. Таким образом, каждый следующий процент прироста количества разобранных словоформ требует всё большего количества дополнительных лексем в словаре. Для пополнения словаря можно пользоваться автоматически порождаемым частотным списком неразобранных словоформ, внося в словарь соответствующие лексемы в порядке частоты их появления в корпусе.

Другим направлением дальнейшей работы является расширение функциональности корпуса. В частности, планируется добавить в число вариантов отображения результатов поиска глоссированную выдачу (в настоящее время есть вариант «псевдо-

глоссированной» выдачи – текст с грамматическим подстрочником, но без разбиения на морфемы). Формат описания грамматики уже сейчас позволяет парсеру выделять в словоформах аффиксы, остаётся добавить в интерфейс возможность отображать эту информацию.

Ещё одним возможным расширением функциональности является размещение на страницах корпусов статистической информации, такой, как частотные списки словоформ и лексем (ср. соответствующую страницу на сайте Восточноармянского национального корпуса <<http://www.eanc.net/>>).

Наконец, в рамках продолжающейся корпусной программы Президиума РАН кроме расширения имеющихся четырёх корпусов планируется создание на той же платформе новых корпусов. На ближайший год запланировано создание крупного корпуса новогреческого языка и ряда корпусов средних языков России и мира.

СПИСОК ЛИТЕРАТУРЫ

1. Савчук С. О. Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // Национальный корпус русского языка: 2003 – 2005. Результаты и перспективы. – М. : Индрик, 2005. – С. 62 – 88.
2. Даниэль М. А., Левонян Д. В., Плунгян В. А., Поляков А. Е., Рубаков С. В., Хуршудян В. Г. Восточноармянский национальный корпус // Армянский гуманитарный вестник. – М.; Ереван : Зангак-97, 2009. – Вып. 2/3-II. – С. 9 – 33.
3. Зализняк А. А. Грамматический словарь русского языка: Словоизменение. – 4-е изд., испр. и доп. – М. : Русские словари, 2007.

Материал поступил в редакцию 24.01.12.

Сведения об авторе

АРХАНГЕЛЬСКИЙ Тимофей Александрович – преподаватель факультета филологии Национального исследовательского университета «Высшая школа экономики», Москва
E-mail: tarkhangelskiy@hse.ru

**Учреждение Российской академии наук
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ВИНИТИ РАН)**

предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через *систему депонирования*.

«Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных мероприятий – конференций, симпозиумов, съездов, семинаров) узкоспециального профиля, разрешенных в установленном порядке к открытому опубликованию, которые нецелесообразно издавать полиграфическим способом печати, а также работ широкого профиля, срочная информация о которых необходима для утверждения их приоритета. Депонирование предусматривает прием, учет, регистрацию, хранение научных работ и обязательное размещение информации о них в специальных информационных изданиях».

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2003).

Результатом депонирования является публикация информации о депонированных научных работах в информационных изданиях ВИНИТИ РАН – реферативном журнале и библиографическом указателе «Депонированные научные работы».

В соответствии с «Положением о порядке присуждения ученых степеней», утв. Постановлением Правительства Российской Федерации № 74 от 30.01.2002 г., депонированные научные работы признаны публикациями, учитываемыми при защите кандидатских и докторских диссертаций.

Подать научную работу на депонирование можно, обратившись в Отдел депонирования ВИНИТИ РАН по адресу:

**125190, Москва, ул. Усиевича, 20.
ВИНИТИ РАН, Отдел депонирования научных работ.
Тел.: 8 (499) 155-43-28, Факс: 8 (499) 943-00-60.
E-mail: dep@viniti.ru**

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНИТИ РАН:
<http://www.viniti.ru>

БАЗА ДАННЫХ ВИНИТИ

ВИНИТИ РАН – один из крупнейших генераторов базы данных по естественным, точным и техническим наукам. Ежегодно с 1981 г. документальная База данных пополняется на ~1 млн. документов по всем разделам РЖ. На основе Базы данных (БД) ВИНТИ предлагает следующие виды продуктов и услуг:

- Диалоговый поиск в БД по сети Internet (WWW-сервер - <http://www.viniti.ru>);
- Политематические или проблемно-ориентированные БД по всем разделам естественных, технических и точных наук на CD-ROM в информационно-поисковой системе "Сокол";
- Любые фрагменты БД ВИНТИ в различных коммуникативных форматах;
- Тематический поиск в БД ВИНТИ по разовым и постоянным запросам силами специалистов ВИНТИ по заказам пользователей.

База данных ВИНТИ в сети ИНТЕРНЕТ

Сервер ВИНТИ обеспечивает доступ к документальной Базе данных ВИНТИ, дает возможность поиска информации в текущих и ретроспективных массивах документов. Web-страницы ВИНТИ организованы в соответствии с мировыми стандартами и содержат полную информацию о ВИНТИ, его информационных продуктах, видах информационных услуг.

База данных на CD-ROM

Предоставляются любые политематические или проблемно-ориентированные фрагменты БД ВИНТИ в поисковой системе "Сокол", обеспечивающей:

- Быстрый поиск "сразу" и "езде" Поиск можно вести в годовом или ретроспективном массиве (за несколько лет сразу) в одном или нескольких тематических фрагментах – "езде".
- Поиск по словам и любым словосочетаниям из заглавия, реферата или из ключевых слов.
- Использование года, языка, рубрик, шифров отдельных выпусков БД для уточнения поиска.
- Поиск по словарю, выполняющему функции многоаспектного указателя, в том числе авторского, предметного, источников, индексов МПК, номеров патентных документов и депонированных рукописей и т.д.
- Запоминание запросов и редактирование их.
- Вывод результатов поиска на экран, в файл и на печать.

За подробной информацией обращаться по адресу:

125190, Россия, Москва, ул. Усиевича, 20, АБНД ВИНТИ

Телефоны: 8 (499) 155-45-01, 155-45-02, 155-45-25, 155-46-20

Факс: 8 (499) 155-45-25

E-mail: csbd@viniti.ru, davydova@viniti.ru

УВАЖАЕМЫЕ ЧИТАТЕЛИ!

ЦЕНТР НАУЧНО-ИНФОРМАЦИОННОГО ОБСЛУЖИВАНИЯ ВИНИТИ РАН

ПРЕДОСТАВЛЯЕТ КОПИИ ПЕРВОИСТОЧНИКОВ

ВИНИТИ осуществляет обслуживание копиями первоисточников, хранящихся в фонде научно-технической литературы ВИНИТИ, в фондах других библиотек, а также в доступных ВИНИТИ электронных ресурсах.

Фонд научно-технической литературы ВИНИТИ включает более 2 млн. изданий по точным, естественным и техническим наукам, в том числе:

- отечественные и иностранные периодические и продолжающиеся издания – с 1987 г.;
- отечественные книги – с 1987 г.;
- иностранные книги – с 1991 г.;
- рукописи, депонированные в ВИНИТИ – с 1962 г.

Заказы на бумажные или электронные копии первоисточников принимает Центр научно-информационного обслуживания (ЦНИО) ВИНИТИ. ЦНИО ВИНИТИ обслуживает коллективных (организации и учреждения) и индивидуальных пользователей.

Формы обслуживания:

- абонементная (договора, на основе предоплаты);
- разовые заказы на основе гарантийных писем (с предоплатой заказа по счету);
- индивидуальная форма обслуживания в читальном зале ЦНИО ВИНИТИ.

На сайте ВИНИТИ (<http://www.viniti.ru>) представлен полный Электронный каталог научно-технической литературы (<http://catalog.viniti.ru>), зарегистрированной в ВИНИТИ с 1994 г. Доступ для просмотра и поиска по Каталогу свободный. Постоянные абоненты ЦНИО ВИНИТИ, имеющие логин и пароль для работы с Каталогом, могут делать заказ копий непосредственно через Каталог.

Услуги по изготовлению копий первоисточников из фондов других библиотек предоставляются только постоянным абонентам. Место хранения первоисточников указывается в Электронном каталоге.

За подробной информацией обращаться по адресу:
125190, Россия, Москва, ул. Усиевича, 20, ВИНИТИ РАН. ЦНИО

Телефоны: 8 (499)155-42-43, 155-42-09, 152-54-59

Факс: 8 (499) 943-00-60

E-mail: cnio@viniti.ru; **URL:** <http://www.viniti.ru>