

Как измерить понимание текста и количество текстовых вирусов

Н. Л. Крячков,
независимый эксперт по рискам
и нормализации текстовых документов,
Санкт-Петербург



ИННОВАЦИИ
В СОЦИАЛЬНОЙ,
КУЛЬТУРНОЙ
И ДРУГИХ
НЕТЕХНОЛОГИЧЕСКИХ
СФЕРАХ

Как известно, в толковых словарях одно слово может иметь от одного до нескольких вариантов толкования. Эти словари фиксируют исторически сложившийся результат накопления многообразия понятий. Толковые словари могут не содержать, например, узкопрофессиональных терминов, диалектных и т.п. слов, и поэтому от них нельзя требовать исчерпывающих сведений для понимания любых предметов, с которыми человек имеет дело. Если в исследуемой предметной области используются отличающиеся от имеющихся в толковых словарях варианты толкования слов, то эти варианты необходимо учитывать при оценке понимания текстовых документов.

Понимание предметов и их кодирование в слове зависело от сложившихся в той или иной эпохе обобщений (категорий), описывающих устройство мира. Например, «пространство» и «время» — известные с древних времен категории, которые могут конкретизироваться словами, созданными в более позднее время. Набор категорий — это своего рода ограничение для миропонимания, стабилизирующее цивилизацию как в смысле ее развития, так и в смысле ее разрушения. Люди, мыслящие в опре-

деленных категориях определенным образом, и развивают, и разрушают цивилизацию. Цивилизация менялась, если менялся набор категорий. Толковые словари в связи с этим представляют собой структурированное лишь по алфавиту неполное хранилище накопленных изменений в понятиях о совершаемых в природе, человеческом организме и во всех осуществляемых человеком действиях.

Описание словами действий представляет собой текст. Поэтому единицей текста целесообразно считать формулировку действия. Любое действие совершается в определенное время, в определенном месте (пространстве) ...

Категории «пространство» и «время» инвариантны, т.е. заключают в себе все основные признаки своих конкретных реализаций, выраженных словами. Слово может иметь несколько вариантов толкования. Это означает, что в некоей формулировке действия сведения о пространстве его совершения выражены, например, словом с двумя вариантами толкования, а сведения о времени — словом с тремя вариантами толкования (табл. 1).

Исходя из этого, мы имеем шесть вариантов толкования формулировки действия:

$$K_{\text{смд}} = \sum_{i=1}^6 X_i = 6$$

Таблица 1

Определение количества вариантов толкования формулировки действия

Инварианты				Варианты толкования формулировки действия (X_i)
Пространство		Время		
Слово	Варианты толкования слова	Слово	Варианты толкования слова	
xxxx		xxxxxx		
				X_1
				X_2
				X_3
				X_4
				X_5
				X_6
Количество вариантов толкования формулировки действия ($K_{\text{смд}}$)				6

ИННОВАЦИИ
В ГУМАНИТАРНОЙ СФЕРЕ

Понимание людьми друг друга с помощью слов означает тот факт, что автор и читатель текста придерживаются одного и того же варианта толкования формулировки действия из всего их возможного количества. Назовем это истинным пониманием (Π_u):

$$\Pi_u = \frac{1}{K_{\text{амд}}}$$

Все остальные варианты толкования формулировки действия образуют риск непонимания. Назовем это рисковым пониманием (Π_p):

$$\Pi_p = \frac{K_{\text{амд}} - 1}{K_{\text{амд}}}$$

Общее количество понимания (Π_o) есть сумма (Π_u) и (Π_p):

$$\Pi_o = \Pi_u + \Pi_p = \frac{1}{K_{\text{амд}}} + \frac{K_{\text{амд}} - 1}{K_{\text{амд}}} = 1$$

или, если выразить его в процентах, то оно составляет 100%.

Рассмотрим графическое представление характеристик $K_{\text{амд}}$, Π_u и Π_p (рис. 1, 2).

Истинное понимание формулировки действия (Π_u) тем больше (соответственно, рисковое понимание (Π_p) тем меньше), чем меньше вариантов толкования слов, обозначающих инварианты формулировки действия. При $K_{\text{амд}} = 1$ $\Pi_u = 1$ или 100%, а $\Pi_p = 0$.

Если в формулировке действия не находится слова, соответствующего необходимому инварианту, то это означает, что читатели могут применять для понимания все слова, удовлетворяющие инварианту со всеми их вариантами толкования. Значения $K_{\text{амд}}$ и Π_p при этом увеличиваются, а Π_u — уменьшается. Например, пусть в формулировке действия никак не обозначено пространство его совершения. Тогда читатель может домысливать пространство любым словом, его выражающим, и любым вариантом толкования этого слова, пользуясь доступным ему словарным запасом.

Категории «пространство» и «время» не единственные инварианты, определяющие формулировку действия. Необходимо определение полного перечня инвариантов и соответствующих им слов. Толкование слов необходимо критически переосмысливать. Структурированное таким образом сложившееся многообразие понятий может быть нормализовано, если инварианты будут выбраны относительно ограничений жизнедеятельности человека на планете Земля в целом (из-за глобального характера последствий этой самой жизнедеятельности).

Текст может состоять из формулировок нескольких действий. Интегрированную количественную оценку его понимания ($ИО$) представим как

$K_{\text{амд}}$	1	2	3	4	5	6	7	8	9	10
Π_u	1,0000	0,5000	0,3333	0,2500	0,2000	0,1667	0,1429	0,1250	0,1111	0,1000

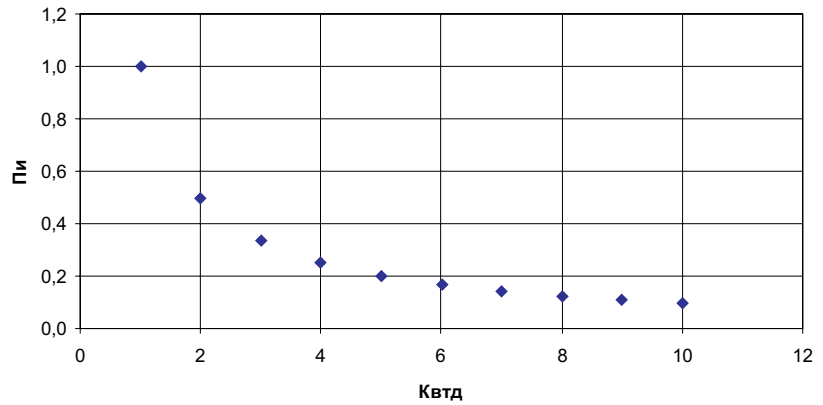


Рис. 1. Зависимость характеристики Π_u от $K_{\text{амд}}$.

$K_{\text{амд}}$	1	2	3	4	5	6	7	8	9	10
Π_p	0	0,5000	0,6667	0,7500	0,8000	0,8333	0,8571	0,8750	0,8889	0,9000

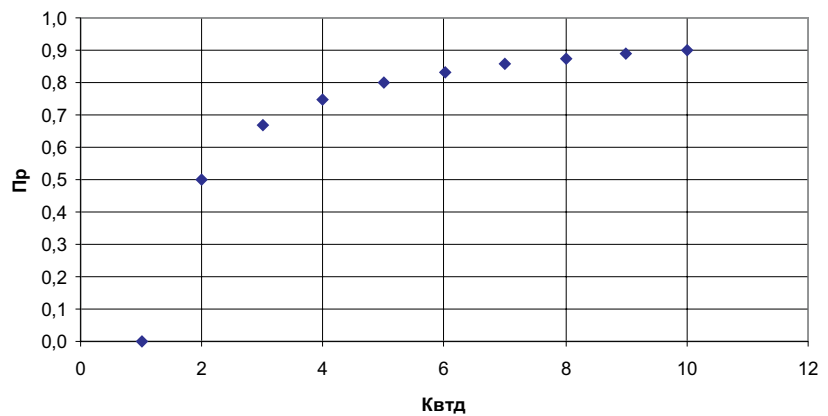


Рис. 2. Зависимость характеристики Π_p от $K_{\text{амд}}$.

сумму соответствующих оценок по каждому действию текста:

$$ИО_{\Pi_u} = \sum_{H_{\delta}=1}^m \Pi_u H_{\delta}$$

$$ИО_{\Pi_p} = \sum_{H_{\delta}=1}^m \Pi_p H_{\delta}$$

$$ИО_{\Pi_o} = \sum_{H_{\delta}=1}^m \Pi_o H_{\delta},$$

где:

H_{δ} — номера действий в тексте, которых может быть от 1 до m .

Рассчитаем интегрированные количественные оценки понимания текста, состоящего, например, из четырех действий с возможными значениями Π_u , Π_p по каждому действию (рис. 3, 4).

Итак, $ИО_{\Pi_u} = 2,2$, $ИО_{\Pi_p} = 1,8$, $ИО_{\Pi_o} = 4$. Если $ИО_{\Pi_o} = 4$ принять за 100%, то $ИО_{\Pi_u} = 2,2$ составляет 55%, а $ИО_{\Pi_p} = 1,8$ составляет 45%.

При $ИО_{\Pi_p} > 0$ в тексте имеются действия, сформулированные неоднозначно. В нашем примере это действия №№ 1, 3, 4 (рис. 3, 4). Это может означать, что автор текста мог предпо-

лагать один вариант понимания своего текста и, следовательно, один вариант поведения согласно этому тексту, а читатель (исполнитель) текста может (в этом и состоит риск) осуществить иной вариант поведения. Например, по действию № 3 иных вариантов 80% от общего их количества.

Определим каждый вариант толкования формулировки действия из количества $K_{\text{амд}} - 1$ как ТЕКСТОВОЙ ВИРУС ($K_{\text{мс}}$):

$$K_{\text{мс}} = K_{\text{амд}} - 1,$$

приводящий к риску непонимания и «неожиданным» вариантам поведения. Тогда количество текстовых вирусов ($K_{\text{мс}}$) по каждому действию и по тексту в целом ($ИО_{K_{\text{мс}}}$) возможно рассчитать, используя, например, таблицу 2, составленную для вышеназванного примера текста, состоящего из четырех действий.

Исходя из этого, в тексте из четырех действий мы имеем шесть текстовых вирусов:

$$ИО_{K_{\text{мс}}} = \sum_{H_{\delta}=1}^4 K_{\text{мс}} H_{\delta} = 6$$

H_o	1	2	3	4	IO
Π_u	0,5	1,0	0,2	0,5	2,2
Π_p	0,5	0	0,8	0,5	1,8
Π_o	1	1	1	1	4

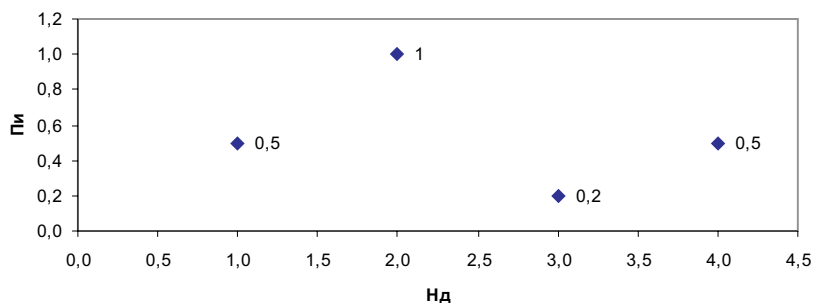


Рис. 3. Возможные значения Π_u в тексте, состоящем из четырех действий.

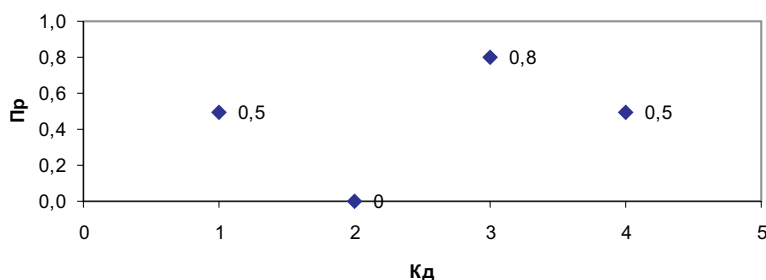


Рис. 4. Возможные значения Π_p в тексте, состоящем из четырех действий.

Таблица 2

Определение количества текстовых вирусов

Номера действий	Π_u	Π_p	K_{md} (см. рис. 1, 2)	K_{ms}
1	0,5	0,5	2	1
2	1	0	1	0
3	0,2	0,8	5	4
4	0,5	0,5	2	1
IO_{Kms}				6

Предложенные характеристики понимания текстовых документов могли бы быть основой для:

- стандартов понимания текстовых документов;
- диагностики управления, осуществляемого согласно текстовым документам;
- дальнейшего исследования текстовых вирусов и поведения людей — носителей текстовых вирусов;
- нормализации текстовых документов.

Нормализация текста на 100% ($IO_{\Pi_u} = 1$) означала бы:

- переформулирование действий словами с однозначным толкованием в пределах инвариантов, определяющих действие, либо

- установление единственного толкования многозначных слов (в пределах инвариантов, определяющих каждое действие) ссылкой на соответствующее толкование из сопровождающего текст словаря или ссылкой на определение этого единственного толкования в предыдущих формулировках действий этого же текста или текста, являющегося основанием для данного текста.

Нормализованный (в рамках инвариантов, удовлетворяющих ограничениям жизнедеятельности человека на планете Земля) текст позволил бы осуществлять программирование на естественных языках для человеческого мозга («биологического компьютера»), в т.ч. и самопрограммирование.

Самопрограммируемый индивидуум мог бы осуществлять относительно независимую от иерархических структур жизнедеятельность. Он был бы способен самостоятельно и упреждающе анализировать свои и не свои действия, а не распределять эту работу по разным специалистам, уровням иерархии или ожидать совершения событий, чтобы описывающие их данные подвергнуть, например, статистической компьютерной обработке и экстраполировать результаты на будущие события, на практике часто отличающиеся по своему содержанию от событий предшествующих.

Человеческий мозг («биологический компьютер») не нуждается в электробатареях или электросетях, его не надо разбирать и наращивать вычислительные мощности. Инвариантное программное обеспечение на естественных языках требует лишь постоянного применения для решения ежедневных задач, а результатом будет продление здоровой жизнедеятельности. Человек, таким образом, перестает быть исключительно частью искусственной иерархической социальной организации, а становится частью природы и может взаимодействовать с индивидуумами на равных, исключая отношения подчинения (насилия), он может усмотреть ложь (текстовой вирус, созданный человеком) и соответствующим образом отреагировать.

Необходимо отметить, что программирование на естественных языках может использоваться и как информационное оружие, направленное на изменение поведения людей не в пользу продления их жизнедеятельности. Вызвать ожидаемое неправильное поведение возможно, зная, какими текстовыми вирусами заражено их сознание, или увеличивая неоднозначность используемых людьми текстов. Причем неоднозначность должна обеспечиваться текстовыми вирусами с противоположными по смыслу вариантами толкования формулировок действий. Защитой (иммунитетом) от такого оружия мог бы стать метод, начала которого представлены Вашему вниманию. Знания, таким образом, становится возможным переосмысливать на предмет их годности к использованию для жизнедеятельности, чтобы не оказываться в ситуации, когда вместо того, чтобы обрабатывать и использовать знания, знания (а точнее, их авторы или интерпретаторы) использовали бы Вас.