

Семантические измерения в оценке качества ресурса

П.Ю. Шалимов

Аннотация – В статье рассматривается концепция понятия семантическая информация с позиции моделей количественной оценки. Представляется авторская информационная технология и модель оценки семантических атрибутов текстовых ресурсов. Вводится понятие семантическая информационная среда как инструмент семантических измерений. На основе использования моделей семантической информационной среды введена единица измерения семантической информации – SU (семантическая единица), как максимальное количество информации полной элементарной среды.

Описываются вопросы семантических измерений, вводятся прямые и производные показатели оценивания семантических атрибутов, инструменты и информационная технология оценивания ресурсов. Излагается принцип применения предложенной технологии на примере задачи по сравнению семантических параметров ресурсов.

Ключевые слова: Семантическая информация, модель количественной оценки, единица измерения семантической информации, семантическая информационная среда, семантические атрибуты ресурса.

I. Введение

Сведения, получаемые субъектом, на аксиоматическом уровне соответствуют семантической информации, вследствие сопоставлений новых сведений с имеющейся информацией [1,2]. В большинстве современных приложений не акцентируется внимание на различии понятий семантическая и синтаксическая информация. Синтаксическая информация не оперирует смысловым содержанием и в полной мере отождествляется с понятием данные. Единицы измерения синтаксической информации - бит, байт. Количество единиц определяется прочитанными пользователем знаками.

Семантическая информация, ассоциируемая со смыслом документа, возникает в процессе обмена

сообщениями, в котором принимает участие, как минимум, один субъект [3]. Любой текстовый источник является носителем семантической информации, что особенно значимо для учебной, справочной литературы.

Сложность и неоднозначность понятия «семантическая информация» определяет необходимость привлечения к изучению ряда научных дисциплин: философию, лингвистику, психологию, информатику [4]–[7].

При определении качества документов сетевых ресурсов учитывается поведение читателей, обращается внимание на количество прочитанных знаков до момента прекращения работы.

Рассматривается понятие «вовлеченность», которое коррелируется семантическим, уникальным содержанием ресурса [8]. Актуальным, при этом, является количество семантической информации, полученное пользователем до момента потери интереса.

Важным становится расширение методов и средств квалиметрии на категорию семантическая информация, применительно к источникам учебной, справочной направленности и документам сетевых ресурсов. Первым этапом оценки семантической информации должна стать разработка моделей семантических измерений, процедур проведения замеров, определение единиц семантической информации.

II. Специфика семантической информации

Под ресурсом понимается множество документов, объединенных по признаку единства: темы, владельца, источника. Документ — множество предложений. Для сравнительной оценки семантической информации (СИ) ресурсов использовалась модель и программное обеспечение, описанное в [9], где измерительным инструментом и носителем семантической информации служит модель семантической информационной среды (СИНФ).

В модели СИНФ используется положение о предложении как минимальной коммуникативной единице. Минимальная порция представления и передачи семантической информации — предложение, как кортеж лексем:

$$L = \{L_i \mid i \in I\}, \quad I = [2, \infty) \quad (1)$$

где L – размещаемое предложение, L_i - лексема.
Минимальное предложение:

Статья получена 29 апреля 2021

П.Ю. Шалимов, к.т.н., доцент, Брянский государственный технический университет (БГТУ), e-mail: shalimov.petr@gmail.com

$$L_{\min} = \{L_1, L_2\}, |L_{\min}|=2 \quad (2)$$

Семантическая информация обладает набором качественных признаков: субъективизм, недетерминизм, стохастичность.

Субъективизм — данные размещаются в экземпляре информационной среды S. Количество получаемой информации зависит от размера сообщения и параметров конкретного экземпляра принимающей среды S.

Недетерминизм — разная реакция на одинаковую внешнюю информацию:

$$F(C(t), S(t)) \neq F(C(t), S(t+1)), \quad (3)$$

где $C(t)$ – сообщение в момент t, $S(t)$ – среда в момент t, F – функция преобразования семантической информации.

Стохастичность — данные, размещаемые в экземпляре информационной среды, описываются законом распределения вероятностей. Случайное событие — предъявление информационной среде синтаксического предложения (1). Случайная переменная — появление на месте i в кортеже L лексемы Li. Поток случайных событий соответствует потоку предложений, предъявляемых экземпляру информационной среды. Аналогом временной оси потока событий служит порядковая переменная номера предложения. Исследование документа с информационной средой, в качестве измерительного инструмента, предполагает конкретизацию понятий: генеральная совокупность, выборка, уровень значимости.

III. Модели и инструменты

Модель СИНФ, используемая для количественных оценок текстового ресурса, считается семантическим измерительным инструментом. В процессе моделирования на входы СИНФ последовательно подаются предложения исходного текста, выход СИНФ – количество семантической информации.

Обобщенная математическая модель базируется на понятии «элемент имитации памяти» (ЭИП) и включает множество схем, различающихся уровнем коммутации элементов, характером учета внутреннего состояния: схема бипольных элементов; коммутируемые монополярный и бипольный слои; слой элементов имитации памяти (ЭИП) бипольного слоя; звено суммирования вкладов и определения итоговой оценки.

В настоящей статье используются описанные [9] модели бипольных СИНФ и программный комплекс ScuS, разработанные автором.

В модели СИНФ применяются функции активации, имеющие вид сигмоидальной двухпараметрической зависимости:

$$F(t) = \frac{1}{1 + \exp(-k(t - a))}, \quad (4)$$

где k, a – параметры функции активации, t- такт локальной активации ЭИП.

Параметры функции активации (4) влияют на размер линейного участка функции. Подбор параметров в каждом вычислительном эксперименте определяет линейность измерительного инструмента.

Значение функции активации определяет составляющую количества семантической информации экземпляра СИНФ. Сигмоидальные функции обладают важным для измерений свойством:

$$\lim_{t \rightarrow \infty} F(t) = 1 \quad (5)$$

Параметры используются в процессе имитации при моделировании качественных признаков семантической информации, имеющих отношение к линейности СИНФ как измерительного инструмента [10].

IV. Единицы семантических измерений

В основе организации метрологического обеспечения семантических измерений находятся модели и программное обеспечение СИНФ. В процессе взаимодействия СИНФ и ресурса происходит формирование экземпляра среды с приобретенной семантической информацией. Количественное значение на выходе экземпляра СИНФ считается оценкой семантической информации текстового ресурса.

Требует проработки ряд принципиальных вопросов:

- 1) создание единицы измерения семантической информации
- 2) обеспечение единства измерений
- 3) разработка информационных технологий обеспечения измерений

Полной схемой бипольных элементов имитации памяти считается СИНФ, получившая набор информационных сообщений в объеме генеральной совокупности. Генеральная совокупность для СИНФ, как понятие статистики, обладает рядом специфических особенностей, определяемых неоднозначным описанием предметной области множества текстовых материалов.

С позиций СИНФ как инструмента семантических измерений учитываются:

- 1) полнота СИНФ по количеству лексем (размеру схемы),
- 2) полнота по количеству семантической информации.

Понятие, имеющее большую прагматическую ценность чем «генеральная совокупность» - определение полной среды уровня α по количеству лексем.

Определение 1: Полной средой уровня α , размером схемы N, называется СИНФ по количеству лексем, такая, что предъявление нового

события L приведет к изменению размера схемы N с вероятностью $1 - \alpha$.

Определение 2: Полной средой по количеству семантической информации уровня α называется СИНФ, такая, что предъявление нового события L приведет к изменению количества семантической информации с вероятностью $1 - \alpha$.

Оценкой размера схемы N полной СИНФ по количеству лексем можно считать максимальное значение количества терминов предметных словарей, энциклопедий [11]–[13].

Определить максимальное значение количества информации полной СИНФ возможно, учитывая свойство (4) функции активации. Для СИНФ бипольных ЭИП схемой размера NxN количество семантической информации:

$$Kc = \sum_{i=1}^N \sum_{j=1}^N \sup(F(1)), \quad (6)$$

где i, j – измерения схемы СИНФ, N – размер схемы, l – номер события бипольного элемента.

Максимальное количество семантической информации полной СИНФ размера NxN по (6) с учетом (5) и $\sup(F(1)) = 1$:

$$Kc = N^2 \quad (7)$$

Элементарной средой называется СИНФ размера минимального предложения, для предложений (1) и (2) из двух лексем N=2.

Определение. За единицу семантической информации принимается максимальное количество информации, соответствующее полной элементарной СИНФ бипольных ЭИП для N=2. По (7) с учетом (1) и (2):

$$Kc=4 \quad (8)$$

Название единицы измерения семантической информации – SU (semantic unit). Значение Kc по (8) принимается в качестве нормирующего коэффициента Kn для перевода значений Kc произвольной СИНФ, при условии линейности работы инструмента:

$$Kc=Ks/Kn, \quad (9)$$

где Kc – количество семантической информации в единицах SU, Ks – количество семантической информации, полученное в результате расчета по модели СИНФ с учетом (6).

V. Семантические показатели

Тема – множество текстовых материалов для целевого аналитического исследования. Семантический раздел – блок текстового материала единой темы, подлежащего оцениванию. С позиций статистического анализа семантический раздел – аналог выборки, тема – генеральная совокупность. Семантические атрибуты раздела для описания точечных оценок на момент совершения события характеризуются основными и производными показателями.

Основной показатель – текущее количество семантической информации Kc(t) – характеризует накопленное экземпляром СИНФ количество СИ после получения сообщения в объеме t предложений. Измеряется в семантических единицах (SU). Текущее значение количества синтаксической информации Cc(t) – накопленное экземпляром СИНФ количество данных. Измеряется в байтах (B).

Производный показатель – эффективность семантической емкости синтаксической единицы Ec(t):

$$Ec(t) = Kc(t)/Cc(t), \quad (10)$$

измеряется в единицах отношения (SU/B). Учитывая низкие абсолютные значения в единицах отношения, чаще используются mSU/B (милиSU/B).

Семантический напор Dc(t) – скорость набора семантической информации:

$$Dc(t) = Kc(t)/t, \quad (11)$$

измеряется в единицах (SU/ON) (семантических единиц/предложение).

Основные и производные показатели предлагают точечную оценку семантической характеристики СИНФ. Количественные значения получаются после предъявления СИНФ предложений в объеме t и относятся к документу.

Производные показатели количественно выражают такие понятия как выразительность и лаконичность текста. Основные показатели характеризуют семантическую емкость текста и предназначены для прямого сопоставления понятий «семантическое» и «прагматическое» значение.

Для характеристик множества документов (выборки), после получения точечных значений, рассчитываются статистики выборки, например, среднего значения.

Среднее по выборкам набора данных количество семантической информации (SU):

$$\overline{Kc} = \frac{1}{w} \sum_{i=1}^w Kc_i(f) \quad (12)$$

Среднее значение количества синтаксической информации (B):

$$\overline{Cc} = \sum_{i=1}^w Cc_i(f) \quad (13)$$

Среднее значение семантической эффективности информации (SU/B):

$$\overline{Ec} = \sum_{i=1}^w Ec_i(f) \quad (14)$$

Среднее значение семантического напора (SU/ON):

$$\overline{Dc} = \sum_{i=1}^w Dc_i(f), \quad (15)$$

где W – количество выборок в наборе данных, f – момент учета конкурентных текущих значений точечных показателей (номер предложения).

Описанные выше семантические показатели могут выступать в роли критериев качества

документа с позиций эффективности использования семантической информации.

Комплекс задач при исследовании семантических отличий текстовых ресурсов

Цели исследования:

- 1) вычисление точечных оценок основных показателей $K_c(t)$, $C_c(t)$, $E_c(t)$, $D_c(t)$; семантического описания документа
- 2) определение семантических характеристик $\overline{K_c}$, $\overline{E_c}$, $\overline{D_c}$
- 3) анализируемых ресурсов;
- 4) выявление пар ресурсов, имеющих статистически значимые отличия;
- 5) интерпретация результатов с позиций смысловой ценности ресурса;
- 6) интерпретация результатов с позиций обоснованности финансовых затрат семантическому содержанию документа.

Входная информация сравнительных семантических исследований документа:

- 1) список анализируемых ресурсов;
- 2) выбранный документ ресурса, преобразованный к текстовому формату. (содержит набор предложений, используется на этапе вычисления точечных оценок $K_c(t)$, $C_c(t)$, $E_c(t)$, $D_c(t)$);
- 3) точечные оценки $K_c(t)$, $C_c(t)$, $E_c(t)$, $D_c(t)$ на этапе определения отличий ресурсов;
- 4) Момент учета текущих значений – предложение f .

Выходная информация: семантические характеристики $\overline{K_c}$, $\overline{E_c}$, $\overline{D_c}$; пары ресурсов, имеющих статистически значимые отличия; уровни статистической значимости отличий.

При сравнительных исследованиях должны быть выполнены основные требования корректности: эквивалентность по моделям и эквивалентность по данным.

Эквивалентность по моделям предполагает использование в имитационном моделировании одинаковых значений параметров k, r (2) функций активации. Эквивалентность по данным подразумевает использование в моделировании выборки, одинаковых по синтаксическому размеру или количеству событий.

Одинаковый синтаксический размер требует выделения выборочных данных по размеру фрагмента в байтах. Программное обеспечение и модель предполагают ввод целого количества предложений. Фактический синтаксический размер выборки будет различаться.

Использование выборки, одинаковых по количеству событий, будет более корректным для удовлетворения требования эквивалентности по данным.

В основу организации количественных семантических вычислений положен анализ прагматического значения исследований. Основные

группы прагматических значений и видов исследования: абсолютные и сравнительные.

Абсолютные исследования (АИ) ставят целью определение семантических атрибутов для последующего связывания с экономическими характеристиками: количество семантической информации ресурса в SU и стоимость в денежных единицах. При АИ выходным результатом будет значение $K_c(t)$ для всего оцениваемого документа (номер t соответствует последнему предложению). Результат указывается с параметрами моделирования (3). АИ проводятся в интересах собственника ресурса, используются при совершении финансовых операций.

Сравнительные исследования (СИ) проводятся с целью выяснения семантических преимуществ документа в группе ресурсов. Определяемые основные и производные атрибуты ассоциируются с таким прагматическим показателем как вовлеченность. При СИ в расчетах используется выборка, состоящая из текстовых фрагментов. Управляемый параметр – размер текстового фрагмента – используется одинаковый для всех экспериментов исследования. В расчетах определяются и сравниваются семантические показатели для фрагмента ресурса.

VI. Информационная технология определения семантических атрибутов

Цель исследования – определение семантических атрибутов текстовых ресурсов для выявления предпочтительности по критерию уникальности. Проводятся сравнительные исследования при определении семантических атрибутов в названный момент времени t . Множество ресурсов сравнительного анализа составляют кортеж исследования.

В качестве объектов сравнения (элементов кортежа исследования), выбраны текстовые ресурсы, имеющие сетевой интерфейс. Источниками данных в спецификации («гетероним» - url) выбраны:

- 1) «классика» - <https://ilibrary.ru/text/11/p.91/index.html>
- 2) «автор» - <https://www.kommersant.ru/authors/3>
- 3) «новости1» - <https://tass.ru/>
- 4) «новости2» - <https://rg.ru/>
- 5) «медиа» - <https://zen.yandex.ru/>

Генеральная совокупность – множество возможных элементов выборки конкретного источника данных. Цель обзора генеральной совокупности темы - оценка мощности множества документов Q , установление требуемой значимости статистических исследований. Опыт вычислительного эксперимента – представление СИНФ определенного элемента выборки.

В качестве оцениваемого текстового материала (элемент выборки), являющегося источником

данных, выступали как целые статьи, так и фрагменты из нескольких статей. Набор данных – множество элементов выборки, являющихся предметом оценивания с помощью специализированного программного обеспечения.

Семантический анализ текстовых ресурсов выполняется с использованием программного комплекса Scus. Программа Scus_t формирует схему информационной среды выборки конкретного набора данных. Для каждого из представленных в таблице наборов данных сформирован экземпляр СИНФ. В качестве исходных данных Scus_t используются все выборки для определения размера схемы, базы нормализованных лексем, результатов кодирования лексем.

Формирование схемы включает размещение лексем в узлах модели бипольных элементов. Функционирование экземпляра СИНФ в режиме имитации чтения набора данных выполняется в модуле Scus_m. Индивидуальные параметры экземпляра СИНФ для набора данных определяются значениями: размер схемы, параметры k, g, выходными значениями ЭИП для момента снятия значений.

Практическая реализация количественной оценки семантических атрибутов выполняется в последовательности процессов информационной технологии сравнительных исследований текстовых ресурсов (рис.1).

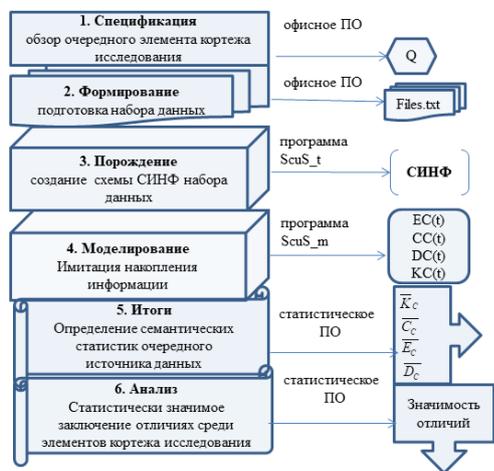


Рис. 1. Процессы информационной технологии оценки семантических атрибутов

В ходе сравнительных имитационных исследований были сформированы экземпляры СИНФ: «СИНФ классика», «СИНФ автор», «СИНФ новости1», «СИНФ новости2», «СИНФ медиа».

VII. Результаты и обсуждение

В таблице показаны маркерные значения семантических показателей ресурса, с разбросом от

минимального $\overline{Kc} = 20$ (SU) до максимального значения $\overline{Kc} = 150$ (SU).

Количественные показатели \overline{Kc} , \overline{Ec} , \overline{Dc} в таблице относятся к набору данных и являются характеристиками физического источника данных (ресурса). Проверка значимости отличий семантических атрибутов \overline{Kc} , \overline{Ec} , \overline{Dc} проводилась в группах «классика» - «новости1», «классика-медиа», «новости1» - «медиа» по критерию Манна-Уитни. Различия статистически значимы для уровня $p < 0,05$.

На основе анализа количественных значений семантических показателей предлагается следующая градация уровней по критериям количества семантической информации \overline{Kc} и семантического напора \overline{Dc} ;

- 1) главный ($\overline{Kc} > 100$ SU, $\overline{Dc} > 3$ SU/ON);
- 2) общий ($50 > \overline{Kc} > 100$ SU, $1,5 > \overline{Dc} < 3$ SU/ON);
- 3) простой ($20 > \overline{Kc} > 50$ SU, $0,5 > \overline{Dc} < 1,5$ SU/ON).

Таблица. Результаты исследований

Набор данных	\overline{Kc} (SU)	\overline{Cc} (B)	\overline{Ec} (mSU/B)	\overline{Dc} (SU/ON)	W
классика	150,1	3900	37,5	3,9	28
автор	130,5	3810	34,1	3,25	34
новости1	63,6	2860	22,3	1,6	41
новости2	73,1	2900	25,1	1,8	59
медиа	22,3	1539	14,5	0,55	54

Данные таблицы и градация имеют иллюстративное значение. После проведения обширных статистических исследований и анализа с методами кластеризации результаты уточняются, используются для интерпретации и описания.

VIII. Заключение

Современные методы определения качества учебной литературы учитывают синтаксическую информацию и применяют порядковые оценочные шкалы [14]. Объем семантической информации документа не учитывается и не принимается во внимание при определении качества документа.

Включение в состав выходных данных учебной литературы количественных значений $Kc(t)$, \overline{Ec} , \overline{Dc} как критериев качества, позволит пользователю получить априорные сведения по семантическому облику документа. Показатель $Kc(t)$ оценивает объем семантической информации издания. Значения $Kc(t)$ определяются для значения t соответствующего номеру последнего предложения документа. Критерий \overline{Ec} характеризует эффективность использования автором синтаксических конструкций для выражения смысла. Критерий \overline{Dc} оценивает лаконичность и выразительность документа.

Семантические измерения в сетевых ресурсах ориентированы на сопоставление семантических показателей качества документа и моделей поведения пользователя. Анализируются время нахождения на ресурсе, момент прекращения чтения, возвращения. Значения $Kc(t)$, \overline{Ec} , \overline{Dc} соответствуют определенному моменту $t=f$. Где f – момент чтения, связанный с потерей интереса пользователя. Полагается, что к этому моменту желательно предложить существенную порцию семантической информации.

Модель, основанная на бипольных элементах имитации памяти и сигмоидальной функции оценивания, использованная в работе, является основной в линейке схем СИИФ. Не учитывается ряд свойств, присущих основному субъекту, оперирующему семантической информацией – человеку: забывание, смена тематических приоритетов, активность в поиске информации. Уточнение модели может проходить в направлении, заданном в [15]–[21].

Управление семантическим содержанием контента посредством управления количественными атрибутами должно относиться к задаче управления качеством продукта.

Ориентированность на аудиторию, в данном случае, будет означать предложение контента с адекватными значениями семантических атрибутов, что видно из приведенной таблицы и предложенной градации. Требуется дополнительное исследование вопроса оптимальности семантических показателей качества текста, темы, ресурса и адекватности семантических атрибутов запросам аудитории и экономическим возможностям владельца ресурса.

БИБЛИОГРАФИЯ

- [1] Гуревич, И.М. Информация как универсальная неоднородность / И.М. Гуревич // Информ. Технологии, 2010. - №4. - С. 66-74
- [2] Кобозева И.М. Лингвистическая семантика. / И.М. Кобозева // Изд. 7, испр. и доп. URSS. 2021. 360 с.
- [3] Р. С. Гиляревский, “Информатика как наука об информации”, Системы и средства информ., 2006, «Научно-методологические проблемы информатики», 59–87
- [4] Ханжин А.Г., Кожокару А. А. Ревизия понятия информации // Научно-техническая информация. Сер. 2. – 2008. – № 6. – С. 1-9.
- [5] Чернавская О.Д. Динамическая теория информации как базис естественно-конструктивистского подхода к моделированию мышления// Компьютерные исследования и моделирование. 2017 Т. 9 № 3 С. 433–447
- [6] Шрейдер, Ю.А. Семантика и категоризация / Ю.А. Шрейдер // М.: Наука, 1991. - 168 с
- [7] Фридланд А. Я. О сущности информации: два подхода // Информационные технологии. – 2008. – № 5. – С. 75-85
- [8] Любимцева М.А. Факторы, влияющие на время чтения и дочитываемости материалов онлайн-медиа // Вестн. Моск. Ун-та. Сер. 10. Журналистика 2020. № 2. С. 53-77
- [9] Шалимов П.Ю. Модели количественной шкалы оценки семантической информации // Автоматизация и моделирование в проектировании и управлении № 2(08) 2020. С. 24-32
- [10] ГОСТ Р ИСО 11095-2007. Статистические методы. Линейная калибровка с использованием образцов сравнения
- [11] Биологический энциклопедический словарь. Режим доступа: <https://rus-bio-dict.slovaronline.com/> (дата обращения: 20.03.2021).
- [12] Большой энциклопедический словарь. Режим доступа: <https://rus-big-enc-dict.slovaronline.com/> (дата обращения: 20.03.2021).
- [13] Киселев, Ю. А. Современное состояние электронных тезаурусов русского языка: качество, полнота и доступность / Ю.А. Киселев, С.М. Поршнев, М.Ю. Мухин // Программная инженерия, 2015. - № 6. - С. 34–40.
- [14] Михеева С. А. Система формализованных критериев оценки школьного учебника // Вопросы образования. 2015. № 4. С. 147–183
- [15] Шполянская И.Ю., Середкина Т.А. Семантические технологии в системе поддержки онлайн-обучения // Прикладная информатика. 2020. Т. 15 №5. С. 52–61.
- [16] Романова Е.В., Обеспечение качества данных в государственных информационных системах // Прикладная информатика. 2017. Т. 12 №5. С. 15–23.
- [17] М. Г. Крейнс, Е. М. Крейнс, “Матричные модели текстов. Интерпретация моделей и экспериментальная верификация”//Матем. моделирование, 32:7 (2020), 24–46
- [18] А.П. Михайлов, Н.А. Маревцева. Модели информационной борьбы // Математическое моделирование, 2011. т.23, № 10, с.19–32.
- [19] Л. Е. Мистров, В. П. Морозов Терминологический поиск в метрологических информационных системах// Метрология . – 2017. – № 1. – С. 33-40
- [20] Sardinha T. B. Segmenting corpora of texts // DELTA. 2002. Vol. 18. P. 273–286.
- [21] Fatma Deniz, Anwar O. Nunez-Elizalde, Alexander G. Huth and Jack L. Gallant. The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. Journal of Neuroscience 25 September 2019, 39 (39) 7722-7736; DOI: <https://doi.org/10.1523/JNEUROSCI.0675-19.2019>

П.Ю. Шалимов
К.т.н., доцент, доцент кафедры «Информатика и программное обеспечение»
e-mail: shalimov.petr@gmail.com
Брянский государственный технический университет,
241035, Россия, г. Брянск, бул. 50 лет Октября, д. 7,

Semantic dimensions in assessing the quality of a resource

P.Y. Shalimov

Abstract - The article discusses the concept of semantic information from the standpoint of quantitative assessment models. The author's information technology and a model for evaluating the semantic attributes of text resources are presented. The concept of semantic information environment is introduced as a tool for semantic measurements. Based on the use of models of the semantic information environment, a unit of measurement of semantic information - SU (semantic unit) has been introduced, as the maximum amount of information of a complete elementary environment.

Issues of semantic measurements are described, direct and derived indicators for evaluating semantic attributes, tools and information technology for conducting resource estimation operations are introduced. The principle of application of the proposed technology is stated on the example of the problem of comparing the semantic parameters of resources.

Key words: Semantic information, quantitative assessment model, unit of measurement of semantic information, semantic information environment, semantic attributes of a resource.

REFERENCES

- [1] Gurevich, I.M. Information as a universal heterogeneity / I.M. Gurevich // Inform. Technologies, 2010. - No. 4. - S. 66-74
- [2] Kobozeva I.M. Linguistic semantics. / THEM. Kobozeva // Ed. 7, rev. and add. URSS. 2021.360 p.
- [3] R.S. Gilyarevsky, "Informatics as a science of information", Systems and means of inform., 2006, "Scientific and methodological problems of informatics", 59–87
- [4] Khanzhin A.G., Cojocar A.A. Revision of the concept of information // Scientific and technical information. Ser. 2. - 2008. - No. 6. - P. 1-9.
- [5] Chernavskaya O.D. Dynamic information theory as the basis of the natural constructivist approach to the modeling of thinking // Computer research and modeling. 2017 T. 9 No. 3 P. 433–447
- [6] Shreider, Yu.A. Semantics and categorization / Yu.A. Schreider // M.: Nauka, 1991. -- 168 p.
- [7] Friedland A. Ya. On the essence of information: two approaches // Information technologies. - 2008. - No. 5. - P. 75-85
- [8] Lyubimtseva M.A. Factors Influencing the Time of Reading and Reading of Online Media Materials // Vestn. Moscow University. Ser. 10. Journalism 2020. No. 2. P. 53-77
- [9] Shalimov P.Yu. Models of a quantitative scale for evaluating semantic information // Automation and modeling in design and management № 2 (08) 2020. P. 24-32
- [10] GOST R ISO 11095-2007. Statistical methods. Linear Calibration Using Reference Samples
- [11] Biological encyclopedic dictionary. Access mode: <https://rus-bio-dict.slovaronline.com/> (date of access: 20.03.2021).
- [12] Big encyclopedic dictionary. Access mode: <https://rus-big-enc-dict.slovaronline.com/> (date of access: 20.03.2021).
- [13] Kiselev, Yu. A. The current state of electronic thesauri of the Russian language: quality, completeness and availability / Yu.A. Kiselev, S.M. Porshnev, M. Yu. Mukhin // Software Engineering, 2015. - No. 6. - P. 34–40.
- [14] Mikheeva SA The system of formalized criteria for evaluating a school textbook // Education Issues. 2015. No. 4. P. 147–183
- [15] Shpolyanskaya I.Yu., Seredkina T.A. Semantic Technologies in the Online Learning Support System // Applied Informatics. 2020.Vol. 15 No. 5. S. 52–61.
- [16] Romanova EV, Ensuring the quality of data in state information systems // Applied Informatics. 2017.Vol. 12 No. 5. S. 15–23.
- [17] MG Kreines, EM Kreines, "Matrix models of texts. Interpretation of models and experimental verification" // Matem. modeling, 32: 7 (2020), 24–46
- [18] A.P. Mikhailov, N.A. Marevtseva. Models of information warfare // Mathematical modeling, 2011. v.23, no. 10, pp.19–32.
- [19] LE Mistrov, VP Morozov Terminological search in metrological information systems // Metrology. - 2017. - No. 1. - P. 33-40
- [20] Sardinha T. B. Segmenting corpora of texts // DELTA. 2002. Vol. 18. P. 273-286.
- [21] Fatma Deniz, Anwar O. Nunez-Elizalde, Alexander G. Huth and Jack L. Gallant. The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. Journal of Neuroscience September 25, 2019, 39 (39) 7722-7736; DOI: <https://doi.org/10.1523/JNEUROSCI.0675-19.2019>