

## АНАЛИЗ ИНТЕГРАЛЬНЫХ СЕМАНТИЧЕСКИХ ПАРАМЕТРОВ

При исследовании текстовых структур, которые описывают предметную область, основная информационная нагрузка приходится на семантику языка, который используется для построения описаний. Поэтому, семантические параметры, которые должны отображать такую семантику, должны быть достаточно развитыми [1]. Прежде чем рассматривать вопросы развитости семантических параметров, остановимся более подробно на таких параметрах, как семантическая непрерывность текста или его фрагмента и на параметре семантической открытости фразы, которые обозначаются символами  $\sigma(\psi) = f(\varphi_1, \dots, \varphi_n)$  и  $\gamma(\varphi_i)$ , соответственно. Семантическая непрерывность текста  $\sigma(\psi)$  определяется, как минимум, в рамках предложения  $\psi$  и может определяться в рамках отдельного фрагмента текста, который состоит из ряда предложений, что записывается в виде следующего соотношения:

$$\sigma(\Phi_i) = F\{\sigma_1(\psi_{i1}), \dots, \sigma_k(\psi_{ik})\} = F[f_1(\varphi_{i1}^1, \dots, \varphi_{in}^1), \dots, f_k(\varphi_{i1}^k, \dots, \varphi_{im}^k)], \quad (1)$$

где  $\Phi_i$  – фрагмент текстового описания, который состоит из  $k$  предложений  $\psi_i$ . Семантическая непрерывность фрагмента текста является интегральным семантическим параметром, который характеризует отдельный фрагмент текста в целом. Необходимость использования такого параметра определяется тем, что для исследования отдельных структур, функционирование которых в предметной области описывается соответствующим фрагментом текста, необходимы параметры, которые такую структуру характеризуют. Соответствующие параметры должны позволять решать задачи прогнозирования процесса поведения соответствующей структуры, которую в отличие от отдельного объекта, будем обозначать символом  $y$  и в этом случае, формальное описание такой структуры может быть представлено в виде  $y_i = f(x_{i1}, \dots, x_{im})$ , где  $f(\cdot)$  – функция, которая описывает взаимосвязь между отдельными объектами, которые составляют структуру  $y_i$  [2]. Каждая структура  $y_i$  функционирует заданным образом в  $W_i$ . В процессе такого функционирования, особенно, при взаимодействии  $y_i$  с другими структурами  $y_j$  или в результате воздействий информационных потоков, которые поступают из источников информации в  $W_i$  и могут оказывать управляющее воздействие на  $y_i$ , заданное функционирование  $y_i$  может измениться. Такие изменения не будут представлять собой единичные изменения, а в силу естественных особенностей структур  $y_i$  из  $W_i$  будут представлять собой определенный поведенческий процесс данных структур, который будет производиться в

зависимости от определенных этапов функционирования соответствующей структуры. Следует отметить, что этапы функционирования могут определяться не только по отношению к параметру времени функционирования, но и по отношению к любому другому параметру, в том числе и по отношению к изменениям семантических параметров.

Интегральные семантические параметры предназначены, в первую очередь, для описания процессов поведения структур  $y_i$ , а не для описания процессов их функционирования. С точки зрения текстового описания структур и описания их поведения, факт изменения в поведенческих процессах приводит к формированию новых описаний, которые отображают произошедшие на некотором этапе изменения, которые отобразились в структуре  $y_i$ . Соответствующие новые описания могут в своем большинстве представлять собой копии предшествующего описания с изменениями, которые обусловлены реализацией текущего этапа поведения. Таким образом, в рамках рассматриваемого подхода, семантические параметры должны описывать не только статические особенности структур или объектов, что имеет место в случае семантических параметров  $\mu$ ,  $\eta$ ,  $\lambda$ ,  $\gamma$  и  $\pi(\varphi_i)$ , но и характер поведения  $y_i$ , при воздействии на них информационных потоков  $F^I$ , которые реализуются в виде отдельных сеансов или некоторой совокупности сеансов связи. Таким образом, можно записать следующее соотношение для отображения соответствующего воздействия  $F^I$  на структуры  $y_i$ :

$$y_i^* = F_i^I(y_i, \delta\tau_i), \quad (2)$$

где  $y_i^*$  – модифицированная структура  $y_i$ ;  $\delta\tau_i$  – параметр, который определяет тип дискретности в преобразовании  $y_i \rightarrow y_i^*$ , которое произошло в результате реализации одного этапа проведения  $y_i$ ;  $F_i^I$  – функция, которая описывает способ реализации поведения  $y_i$  на этапе  $\delta\tau_i$  под воздействием сеанса информации  $P_i^I$ . Приведенное соотношение для  $y_i^*$  можно записать в более полной форме следующим образом:

$$y_i^* = F^I(P_i^I, y_i, \delta\tau_i). \quad (3)$$

Очевидно, что чрезвычайно важным является исследование возможных способов реализации  $F^I$  и предоставление их в явной форме. На данном этапе рассмотрения ограничимся семантическими параметрами, которые могли бы характеризовать результаты преобразований в структурах  $y_i$ , которые происходят в процессе поведения  $y_i$  и сами процессы поведения. Это позволит рассматривать процессы поведения  $y_i$  на более общем уровне, который не зависит от особенностей текстового описания той или иной структуры.

Чтобы проанализировать понятие семантической неразрывности, необходимо рассмотреть возможные причины семантической разрывности. В данном случае семантическая разрывность или неразрывность не определяется в виде некоторого факта, который может произойти или уже произошел. Семантическую разрывность будем обозначать как  $\sigma^{-1}$ , поскольку она будет нами рассматриваться как величина обратная семантической неразрывности или непрерывности. Как и некоторые другие семантические параметры, величина семантической прерывности определяется на некотором диапазоне значений. Это значит, что  $\sigma^{-1}$  или  $\sigma$  может иметь целый ряд значений, включая свою нижнюю и верхнюю границу. Прежде чем интерпретировать возможные величины значений семантической непрерывности, определим причины вследствие которых может произойти семантическая прерывность в фрагменте текста описывающего некоторую структуру [3]. Этими причинами могут быть:

- возникновение противоречивости  $\eta$ ;
- возникновение конфликта  $\gamma$ ;
- модификация фрагмента текста вследствие реализации одного или нескольких вариантов поведения  $y_i$ ;
- возникновение процессов противодействия  $y_i$  воздействию  $P^I$ , которые будем обозначать как  $-P^I$ ;
- процесс внутреннего развития  $y_i$ , который будем обозначать  $R$ .

Поскольку два семантических параметра  $\eta$  и  $\gamma$ , которые определяют возможность появления разрыва в  $\sigma(\Phi_i)$  изменяются в заданном интервале, который будем называть интервалом активности каждого из параметров, то есть необходимость согласовать между собой, с одной стороны, диапазоны изменений семантических параметров с диапазоном изменения параметра  $\sigma^{-1}(\Phi_i)$ , а с другой стороны, согласовать диапазон изменения  $\sigma^{-1}(\Phi_i)$  с процессом модификации  $y_i$ . Процесс модификации  $y_i$  является результатом реализации процессов поведения  $y_i$  и реализации процессов противодействия потоку  $P^I$ . Если установить, что величины значений параметра  $\sigma^{-1}$ , обусловленные  $\eta$  или  $\gamma$ , окажутся равными или пересекающимися величинами, которые обуславливаются процессами поведения  $y_i$ , что для удобства будем обозначать символами  $F^I(y_i)$  или процессами противодействия, которые будем обозначать  $-F^I(y_i)$ , то возможны следующие две ситуации.

Первая – значения величин  $\sigma^{-1}$ , зависящие от  $\eta$  и  $\gamma$ , будут обуславливать значения  $\sigma^{-1}$  в одной части диапазона значений, например,  $[\beta_1^\eta, \beta_2^\eta] \subset [\beta_{\min}^\sigma, \beta_{\max}^\sigma]$  и, аналогично, для  $\gamma - [\beta_1^\gamma, \beta_2^\gamma] \subset [\beta_{\min}^\sigma, \beta_{\max}^\sigma]$ , а

значения  $\sigma^{-1}$  обусловленные  $F^I(y_i)$  или  $-F^I(y_i)$ , будут вызывать изменения значений  $\sigma^{-1}$  в других частях диапазона. Например,  $[\beta_1^F, \beta_2^F] \subset [\beta_{\min}^\sigma, \beta_{\max}^\sigma]$ , а также  $[\beta_1^{-F}, \beta_2^{-F}] \subset [\beta_{\min}^\sigma, \beta_{\max}^\sigma]$  и будет иметь место:  $\{[\beta_1^\eta, \beta_2^\eta] \cap [\beta_1^\gamma, \beta_2^\gamma] \cap [\beta_1^F, \beta_2^F] \cap [\beta_1^{-F}, \beta_2^{-F}]\} = 0$ .

Вторая – различные поддиапазоны интервала  $[\beta_{\min}^\sigma, \beta_{\max}^\sigma]$  в различных комбинациях пересекаются между собой, что записывается в виде  $\{[\beta_1^\eta, \beta_2^\eta] \cap [\beta_1^\gamma, \beta_2^\gamma] \cap [\beta_1^F, \beta_2^F] \cap [\beta_1^{-F}, \beta_2^{-F}]\} \neq 0$ .

Во втором случае, имеет место ситуация, когда различные причины обуславливают одну и ту же ситуацию в  $\sigma^{-1}$ . В этом случае необходимо рассмотреть ситуацию, можно ли утверждать что причины изменения  $\sigma^{-1}$ , которые привели к одинаковым значениям величины  $\sigma^{-1}$  как-то связаны между собой. Аналогичные рассуждения можно провести и по отношению к причине приводящей к изменению  $\sigma^{-1}$ , которая обозначается выше символом  $R$ .

Исходя из изложенного, можно сформулировать определение параметра  $\sigma^{-1}$  аналогично определению параметра  $\pi(\varphi_i)$ , но при этом, необходимо ввести дополнительные условия.

**Условие 1.** Диапазон значений для  $\eta$  необходимо разделить на части – критические и некритические, что запишется в виде соотношения  $[\eta] = [\eta^k] \cup [\eta^{-k}]$ , где  $[\eta^k]$  – диапазон критических значений  $\eta$ , а  $[\eta^{-k}]$  – диапазон некритических значений  $\eta$ .

**Условие 2.** Диапазон значений для  $\gamma$  необходимо разделить на часть критических значений и часть некритических значений, что формально записывается в виде соотношения  $[\gamma] = [\gamma^{-d}] \cup [\gamma^d]$ , где  $[\gamma^d]$  – поддиапазон критических значений конфликтов  $\gamma$ , и  $[\gamma^{-d}]$  – поддиапазон некритических значений конфликтов  $\gamma$ .

**Условие 3.** Диапазон значений  $Z_i$ , принимаемых при воздействии  $P^I$  на  $y_i$  должен делиться на поддиапазон допустимых значений и поддиапазон критических значений  $Z_i^k = P^I(y)$ .

**Условие 4.** Диапазон значений  $Z_i$ , при влиянии противодействующих факторов  $-P^I$  должен делиться на поддиапазон допустимых значений и поддиапазон критических значений  $Z_i^k = -P^I(y_i)$ .

**Условие 5.** Диапазон значений  $Z_i$ , при воздействии внутренних факторов развития  $R$  должен делиться на поддиапазон допустимых значений и поддиапазон критических значений  $Z_i^k = R(y_i)$ .

В этом случае можно сформулировать следующее определение  $\sigma$  для  $\Phi_i$ .

**Определение 1.** Семантическая непрерывность  $\sigma$  представляет собой параметрическую функцию, которая определена в пространстве  $(\eta, \gamma, z_i)$  таким образом, что ее значения не попадают в поддиапазоны критических значений соответствующих координат.

Формально, это определение можно записать в виде логической функции [4]:

$$[\forall(x_i, x_j) \rightarrow \exists(x_i, x_j) [(\eta(x_i, x_j) \geq \eta_{kr}) \& (\gamma(x_i, x_j) \geq \gamma_{kr})] \& [\forall Z_i \rightarrow \exists(Z_i = Z_i^k)]] \\ \rightarrow [\sigma(\Phi_i) = f(\eta, \gamma, z_i)].$$

Как и в случае определения  $\pi(\Phi_i)$ , это соотношение позволяет только проверять принадлежность точек некоторой кривой в параметрическом пространстве функции  $\sigma(\Phi_i)$ . При определении значения  $z_i$  на данном этапе рассмотрения, примем, что значение некоторого семантического параметра  $z_i$ , который характеризует результат воздействия на  $y_i$  потока  $P^I$ , процессов противодействия  $\neg P^I$  и внутренних процессов  $R$ , определяется в различных областях значения этого параметра, в зависимости от вида воздействия на  $y_i$ . Кроме того, принято, что соответствующий параметр является одним для отображения влияния  $P^I$ ,  $\neg P^I$  и  $R$  на структуру  $y_i$ . Последнее допущение может быть изменено таким образом, что таких семантических параметров может оказаться больше. Например, для каждого вида воздействия на структуру  $y_i$  может использоваться отдельный параметр, который характеризует соответствующее воздействие. Это будет рассматриваться более детально при рассмотрении процессов изменений структур  $y_i$  под воздействием соответствующих факторов.

В общем случае, можно следующим образом описать суть семантического параметра  $\sigma(\Phi_i)$ . Параметр семантической непрерывности фрагмента текста, который описывает соответствующую структуру  $y_i$  из предметной области  $W$ , характеризует отсутствие в пределах данного описания семантических разрывов. В этом случае, семантический разрыв определяется критическими значениями параметров противоречивости, конфликтности и обобщенного семантического параметра  $z_i$ , который характеризует изменения, происходящие в описании структуры, в результате воздействия на  $y_i$  факторов  $P^I$ ,  $\neg P^I$  и  $R$ . Такая ситуация является непротиворечивой несмотря на то, что интуитивно структура  $y_i$  в  $W$  и текстовое описание структуры  $y_i$  физически различные. Но следует помнить, что фактор  $P^I$ , который порождается источником передачи сообщения,

представляет собой текст, который кроме воздействия на структуру  $y_i$ , может связываться с текстовым описанием соответствующей структуры. В результате синтеза двух таких текстовых образов может быть сформировано новое текстовое описание структуры  $y_i$ , в котором отображается  $y_i^*$ , которая соответствует структуре, изменившейся под воздействием фактора  $P^I$ .

В случае фактора  $R$ , в рамках текстового представления структуры  $y_i$  должны описываться компоненты, которые определяют условия возникновения и существования фактора  $R$  в среде структур  $y_i$ , что обеспечивает возможность моделирования его воздействия на  $y_i$  в рамках текстовых описаний.

Рассмотрим параметр семантической открытости  $\chi$ , предложения  $\psi$  или фрагмента текстового описания структуры  $y_i$ , который формально будем обозначать символом  $\chi(\Phi_i)$ . Как и в случае параметра  $\sigma(\Phi_i)$ , параметр  $\chi(\Phi_i)$  является интегральным параметром фрагмента текста  $\Phi_i$ . На содержательном уровне этот параметр характеризует степень податливости структуры на внешние воздействия, независимо от целей такого воздействия. Обоснованность использования такого параметра состоит в том, что объектами и структурами информационного воздействия являются социальные объекты и структуры. При этом такие социальные структуры не представляют собой какие-либо социальные объединения, которые существуют физически, а представляют собой определенным образом классифицированные группы людей, которые физически объединяются некоторыми признаками их локализации. Например, территорией, на которой они проживают, местом и характером производственной или какой-либо другой деятельности и т.д. Более детальный анализ природы структур  $y_i$  связан с непосредственным рассмотрением предметной области интерпретации  $W_S$ , структуры которой являются объектами воздействия информационных потоков  $P^I$ , которые описываются в предметной области интерпретации  $W_I$ . Исходя из приведенного описания параметра  $\chi(\Phi_i)$  можно утверждать, что он характеризует не произвольный фрагмент текста  $\Phi_i$ , а такой фрагмент, который описывает отдельную структуру  $y_i$  в  $W_S$ . Таким образом, параметр семантической открытости  $\chi(\Phi_i)$  является параметром, который средствами семантики описывает свойство структуры  $y_i$ . Если принять, что структура  $y_i$  как некоторый виртуальный физический объект из предметной области  $W_S$ , может описываться на разных уровнях абстрактного представления, то наименее абстрактным описанием является описание в виде определенных соотношений между параметрами, которые характеризуют соответствующие структуры  $y_i$  и объекты  $x_i$ . Наименее абстрактное описание таких структур представляет собой описание в виде текста на естественном языке. Последнее описание является достаточно сложным для обеспечения числовых исследований над соответствующими

структурами. Поэтому введение семантических параметров, которые определенным образом могут быть оценены числовыми величинами, позволяет исходя из текстовых описаний структур  $y_i$  с тем или иным приближением, переходить на более высокий уровень абстракции. Построение описаний  $y_i$  на высоком уровне абстракции позволяет находить более общие решения задач, возникающих в предметных областях  $W_S$  и  $W_A$ .

Понятие семантической открытости  $\Phi_i$  тесно связано с понятиями эффективности воздействия информационного потока  $P^I$  на соответствующую структуру  $y_i$ . Более детально это понятие рассмотрим при анализе описаний информационных потоков. Для более адекватного определения семантического параметра  $\chi(\Phi_i)$ , рассмотрим его связь с другими семантическими параметрами и определим чем, с точки зрения интерпретации  $\chi(\Phi_i)$ , может определяться изменение значения этого параметра. Понятно, что интегральный параметр по своей природе, может зависеть от самых разных факторов и особенно, если объектами исследования являются объекты социального типа, которые сами по себе достаточно сложно поддаются формализации. Поэтому на данном этапе анализа ограничимся некоторым минимальным набором факторов от которых может зависеть значение величины  $\gamma(\Phi_i)$ , а в дальнейшем, по мере необходимости, количество таких факторов будем увеличивать.

Рассмотрим зависимость параметра  $\chi(\Phi_i)$  от параметра  $\sigma(\Phi_i)$ . Непрерывность  $\sigma(\Phi_i)$  означает, что с точки зрения критических ситуаций, которые могут существовать в  $y_i$  в виде противоречивости или конфликтности, соответствующая структура  $y_i$  при высоком уровне непрерывности является безопасной. Это, в свою очередь, означает, что возможность внутренней модификации соответствующего описания  $y_i$  является небольшой, поскольку она может привести к увеличению значения  $\eta$  или  $\gamma$ . Под внутренней модификацией  $y_i$  будем понимать такое воздействие  $P^I$  на  $y_i$ , при котором изменение в текстовом описании реализуется внутри фрагмента  $\Phi_i$ . Таким образом, можно говорить, что чем выше семантическая непрерывность  $\sigma(\Phi_i)$  тем ниже уровень семантической открытости  $\chi(\Phi_i)$  по отношению к внутренним типам воздействия  $P^I$  на  $y_i$ . Под внешней модификацией  $y_i$ , которая представлена в виде  $\Phi_i$  будем понимать такую модификацию соответствующего фрагмента, которая состоит в дополнении текстового описания  $y_i$  путем добавления к имеющимся частям фрагмента  $\Phi_i$  нового фрагмента, размер которого не меньше отдельной фразы. Формально, такие модификации можно описать в виде следующих соотношений. Внутренняя модификация может быть представлена в виде:

$$F_V(P_i^I, \Phi_i) = \omega_{i1} * \omega_{i2} * \dots * \omega_{ij} * \omega_i(P_{i1}^I, \dots, P_{ik}^I) * \omega_{ij+1} * \dots * \omega_{in}, \quad (4)$$

где  $\Phi_i = \omega_{i1} * \omega_{i2} * \dots * \omega_{ij} * \omega_{ij+1} * \dots * \omega_{in}$ , а поток  $P^I$  поступивший в  $W_S$  представляет собой  $P_i^I = \omega_i(P_{i1}^I, \dots, P_{ik}^I)$ , а  $F_V$  - функция внутренней модификации, которая в случае модификации способом расширения выбирает координаты в  $\Phi_i$  для размещения модифицирующего расширения. Внутренняя модификация может состоять не только в расширении  $\Phi_i$ , но и в замене фрагментов  $\omega_{ij}$  из  $\Phi_i$  на фрагменты  $\omega_i(P_{i1}^I, \dots, P_{ik}^I)$ . Внешняя модификация  $\Phi_i$  формально запишется в виде:

$$F_W(P_i^I, \Phi_i) = \omega_{i1} * \omega_{i2} * \dots * \omega_{in} * \omega_i(P_{i1}^I, \dots, P_{ik}^I), \quad (5)$$

где  $F_W$  - функция внешней модификации описания  $y_i$ .

Семантическая открытость фрагмента  $\Phi_i$  связана с семантической корректностью  $\pi(\varphi)$  следующим образом. Поскольку семантическая корректность  $\pi(\varphi_i)$  определяется в виде кривой, которая в пространстве  $(\eta, \gamma)$  описывает меру отдаленности точек этой кривой от точек соответствующих критическим значениям  $\eta$  и  $\gamma$ , то это означает, что внутренняя модификация  $F_V(P_i^I, \Phi_i)$  приведет к уменьшению такой отдаленности или даже к возникновению критических значений  $\eta$  или  $\gamma$  в модифицируемом фрагменте  $\Phi_i$ . Логично принять, что исходные описания  $y_i$  в виде  $\Phi_i$  обладают максимальной мерой  $\pi(\varphi_i)$  и  $\sigma(\Phi_i)$ . Рассмотрим следующее определение параметра семантической открытости  $\Phi_i \leftrightarrow y_i$ , которая обозначается  $\gamma(\Phi_i)$ .

**Определение 2.** Семантическая открытость  $\gamma_i$  фрагмента  $\Phi_i$  определяет меру доступности соответствующего фрагмента  $\Phi_i$ , для его модификации потоком  $P_i^I$ . Для предметной области  $W^* = \{W_S \cup W_I\}$  этот параметр является чрезвычайно важным, так как он определяет в значительной мере возможность успешного воздействия информационного потока  $P_i^I$  на объекты воздействия, которые составляют  $W_S$ . Очевидно, что в случае использования параметра  $\chi(\Phi_i)$  необходимо учитывать такие факторы как противодействие  $y_i$  информационному потоку  $P_i^I$  и фактор внутреннего развития  $R$ . Формальное описание семантического параметра можно представить следующим образом:

$$\chi(\Phi_i) = \alpha / F[\sigma(\Phi_i), (\sum_{i=1}^m \pi_i(\varphi_i))], \quad (6)$$

где  $\alpha$  – коэффициент согласования размерностей,  $F(\cdot)$  – функция которая описывает итоговое влияние параметров  $\sigma(\Phi_i)$  и  $\pi(\varphi_i)$  на величину  $\gamma(\Phi_i)$ , а  $m$  – количество фраз в фрагменте  $\Phi_i$ . В простейшем случае функция  $F(\cdot)$  может означать суммирование значений параметров  $\sigma(\Phi_i)$  и  $\pi(\varphi_i)$ . Приведенное соотношение (6) может быть использовано для количественной оценки величины  $\chi(\Phi_i)$  благодаря тому, что численные значения параметров  $\sigma(\Phi_i)$  и  $\pi(\varphi_i)$  определяются на основе использования элементарных семантических параметров  $\mu$  и  $\lambda$ , которые, в свою очередь, определены численно.

1. *Падучева Е.В.* Понятие презумпции в лингвистике/Семиотика и информатика №8, 1977.
2. *Мендельсон Э.* Введение в математическую логику. М.: Наука, 1971, - 320 с.
3. *Гладкий А.В.* Формальные грамматики и языки. М.: Наука, 1973.
4. *Клини С.* Математическая логика. М.: Мир, 1973.

*Поступила 2.02.2009р.*

УДК 004.832.3

Ю. Р. Валькман, д.т.н., зав. отд., А. Ю. Рыхальский, м. н. с.,  
отдел распределенных интеллектуальных систем Международного научно-учебного центра информационных технологий и систем НАН и МОН Украины

## КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ ИНДУКТИВНОГО ФОРМИРОВАНИЯ ЗНАНИЙ

**Введение.** Объектом исследования являются процессы порождения новых знаний в вычислительной среде. Предмет исследования – компьютерные технологии индуктивного формирования знаний. Цель исследования – анализ и разработка обобщенной модели индуктивного формирования знаний в исследованиях сложных систем. Ожидаемые результаты – разработка интеллектуальной компьютерной технологии, поддерживающей синтез и анализ новых знаний об исследуемых процессах и