

ТЕКСТОВЫЕ МОДЕЛИ И ИХ ИСПОЛЬЗОВАНИЕ

Аннотация. В работе рассматривается подход к моделированию социальных объектов на основе использования определенным способом составленных их текстовых описаний. Вводятся определения семантических параметров, которые используются для анализа соответствующих текстовых описаний. Такие текстовые описания совместно с правилами их преобразований представляют собой текстовые модели

Ключевые слова. Текстовые модели, семантические параметры, логические формулы, социальные объекты, семантические преобразования.

Введение.

Предметом исследований, в которых используются информационные средства, все чаще становятся сложно формализуемые объекты [1]. В связи с этим, многие исследователи используют слишком общие способы представления соответствующих объектов, что приводит к тому, что их формальное описание переходит в категорию качественных описаний. В данной работе, для решения задач моделирования таких объектов, к которым, в первую очередь, относятся социальные объекты (*CO*) [2], используется представление о текстовых информационных моделях (*TIM*).

Основные определения.

Определение 1. Текстовой информационной моделью (*TIM*) социального объекта называется текстовое описание такого объекта, в котором определены связи между компонентами объекта следующих типов: связь следования; неявная семантическая связь; связь текстовых посылок.

Текстовое представление *TIM* является структурированным текстом, в котором используются различные знаки разделения. Поскольку, по своей природе, текстовая форма является последовательной структурой, то одной из базовых координат является координата, которая определяет длину текстовой формы. Такая длина, чаще всего, измеряется количеством слов, которые в тексте расположены последовательно.

При формировании текстов, которые описывают *TIM*, используются следующие атрибуты:

- семантические словари S_C различного назначения,
- семантические параметры Ξ ,
- правила семантических преобразований Σ .

Семантический словарь состоит из однотипных элементов, каждый из которых записывается в виде:

$$x_i := \langle \alpha_{i1}, \dots, \alpha_{im} \rangle_1 \dots \langle \alpha_{i1}, \dots, \alpha_{im} \rangle_k, \quad (1)$$

процессе функционирования *TIM*, α_{ij} – слово естественного языка, которое используется для составления одной фразы интерпретационного описания; $\langle \alpha_{i1}, \dots, \alpha_{im} \rangle_m$ – отдельная фраза или отдельное предложение, которое используется в интерпретационном описании. В случаях, когда интерпретационные описания не будут приводиться в явной форме, соотношение (1) будем записывать в следующем виде: $x_i = j(x_i)$, где j – идентификатор интерпретационного описания. В этом случае словарь S_c можно представить в виде:

$$S_c = \begin{cases} [x_1 := \langle \alpha_{11}, \dots, \alpha_{1m} \rangle] \& \\ \& [x_2 := \langle \alpha_{21}, \dots, \alpha_{2n} \rangle] \& \\ \dots \dots \dots \dots \\ \& [x_n := \langle \alpha_{n1}, \dots, \alpha_{nk} \rangle] \end{cases} \quad (2)$$

Для описания $j(x_i)$ в S_c существуют некоторые ограничения. Первое из них состоит в том, что в словаре S_c , x_i не может содержать слов α_{ij} таких, которые представляют собой элемент словаря x_j , для которого имеет место соотношение $j > i$. Второе ограничение состоит в том, что $j(x_i)$ должны представлять собой нормализованные описания на естественном языке. Это означает, что в $j(x_i) := \langle \alpha_{i1}, \dots, \alpha_{im} \rangle$ не должны встречаться неоднозначности и избыточность. Третье ограничение состоит в том, что $j(x_i)$ содержит описания, которые определяют объект как таковой, но не содержат описаний, которые представляют допустимые функциональные преобразования соответствующих объектов.

Определение 2. Семантическое значение или семантическая значимость μ некоторого объекта x_i , который содержится в семантическом словаре S_c , определяется количеством слов, необходимых для описания интерпретационного расширения $j(x_i)$.

Формально, это определение можно записать следующим образом:

$$[j(x_i) := \langle \alpha_1, \dots, \alpha_m \rangle] \rightarrow [\mu(x_i) = \sum_{i=1}^m Sg(\alpha_i)], \quad (3)$$

где функция $Sg(n)$ описывается следующим образом

$$[(\alpha_i = 0) \rightarrow (Sg = 0)] \vee [(\alpha_i \neq 0) \rightarrow (Sg = 1)].$$

Определение 3. Семантическая актуальность $\lambda(x_i)$ слова x_i , которое описывается в семантическом словаре S_c , определяется числом, которое определяет сколько раз на протяжении одного эффективного цикла ΔT_i употреблялось данное слово в процессе функционирования *TIM*.

Формальное описание этого параметра можно представить в виде:

$$\lambda(x_i)_i := \langle \alpha_{i1}, \dots, \alpha_{im} \rangle_1 | \dots | \langle \alpha_{i1}, \dots, \alpha_{im} \rangle_k \ll \xi_i \gg, \quad (4)$$

где компонента $\ll \xi_i \gg$ увеличивается на единицу каждый раз, когда x_i употребляется в новой фразе, которая входит в состав описания текстового

отображения. Тогда, $\lambda(x_i) = \xi_i$.

Определение 4. Величина семантической противоречивости между выбранной парой слов x_i и x_j в фразе φ_i определяется абсолютным значением разности величин семантической значимости этих слов, или $\eta(x_i, x_j) = |\mu(x_i) - \mu(x_j)|$.

Приведенное определение описывает локальную противоречивость в фразе φ_i , если она состоит из более чем двух слов. Обобщенная по всей фразе φ_i противоречивость, которая, по существу, является противоречивостью фразы, описывается нижеследующим определением.

Определение 5. Обобщенная или глобальная противоречивость $\eta(\varphi_i)$ равна средней величине локальных противоречивостей $\eta(x_i, x_j)$, которые определяются по всем парам слов, используемым в фразе φ_i и формально

описывается соотношением: $\eta(\varphi_i) = \left[\sum_{i=1}^{m-1} \eta(x_i, x_{i+1}) \right] / m$, где m – количество

слов в фразе φ_i .

Определение 6. Неопределенность $\omega(\varphi_i)$ фразы φ_i определяется величиной разброса локальных противоречивостей фразы η_i^G , которые определяются для различных комбинаций слов, при определении локальных противоречивостей η_i^L .

Формально, неопределенность фразы можно определить в соответствии с соотношением:

$$\omega_i(\varphi_i) = \left[\sum_j^{P(m-1)} \left[\sum_{i=1}^{m-1} \eta_{ji} / m - 1 \right] \right] / P(m-1), \quad (6)$$

где верхний индекс суммы определяется числом размещений пар слов x_i в фразе φ_i по два слова.

Определение 7. Двухвариантный конфликт γ^d определяется как абсолютное значение разности между отношениями семантической значимости к семантической актуальности одной пары из двух слов x_i и x_j .

Формально, это определение можно записать в виде:

$$\gamma_i^d = \left\{ \left[\beta_i \mu(x_i) / \lambda(x_i) \right] - \left[\beta_j \mu(x_j) / \lambda(x_j) \right] \right\}, \quad (7)$$

где β_i и β_j – коэффициенты приведения масштабов изменения величин μ и λ . Исходя из приведенного соотношения видно, что если $\lambda(x_i)$ и $\lambda(x_j)$ равны единице, то γ_i^d вырождается в определение локальной противоречивости.

Определение 8. Семантическая корректность $\pi(\varphi_i)$ представляет собой параметрическую функцию в пространстве (μ, λ) определенную на всей фразе φ_i и на всем предложении ψ_i , которая не допускает во фразе φ_i или предложении ψ_i возникновения семантической противоречивости η^L или

семантического конфликта γ^d .

Формально такая функция записывается в виде:

$$\forall(x_i \in \varphi_i) \neg \exists(x_i, x_j) [[\eta(x_i, x_j) \geq \eta_{\min}] \& [\gamma(x_i, x_j) \geq \gamma_{\min}]] \rightarrow [\pi(\varphi_i) = f(\mu, \lambda)]. \quad (8)$$

Поскольку $\pi(\varphi_i)$ определено на φ_i , то во всех точках функции $\pi(\varphi_i)$, которую аппроксимирует кривая линия в плоскости (μ, λ) , отсутствуют такие взаимозависимости между значениями μ_i и μ_j , что вычисленная по соотношению $\eta(x_i, x_j) = |\mu(x_i) - \mu(x_j)|$ величина η^L не попадет в диапазон не допустимых значений для η . Аналогично, ни одна точка на этой кривой для μ_i и ни одна точка для λ_i не приводят к тому, чтобы значение γ^d , вычисленное по соотношению (8) попало в диапазон не допустимых значений.

Использование представлений о ТИМ.

Приведенные выше определения позволяют сформировать представление о способе описания CO с помощью $ТИМ$, который представляет собой группу людей, объединяемых определенными связями, свойствами, целями или другими особенностями, которые достаточно сложно формализовать [3]. Основным возмущающим или управляющим фактором является информационный поток P_i , который представляется в виде текстового описания. Различные CO представляют собой структурированные текстовые описания различного содержания [4]. Поэтому, можно говорить о взаимодействии P_i с y_i , где y_i это CO . Совокупность y_i составляет предметную область W_s , совокупность P_i составляет предметную область W_i . Для осуществления преобразований текстовых описаний y_i в результате воздействия W_i используется система семантических преобразований Σ , которая составляет вместе с описанием интерпретации этих правил предметную область W_p [5]. Группа людей, которые связаны между собой одним или несколькими параметрами внутреннего характера, а также параметрами, которые описывают или характеризуют возможность или необходимость взаимодействия с компонентами внешней среды, тоже будет обозначаться символом y^i , но, при этом, будут описываться соотношением: $y^i = \varphi_i[r_1(y_k^i), \dots, r_m(y_m^i), v_1, \dots, v_k]$, где $r_i(y_j^i)$ - описание соотношения y_j^i с компонентной из W_s , которая представляет собой y_j^i . Любой социальный объект функционирует в пределах некоторого стационарного режима. Воздействие информационного потока на такой объект рассматривается в аспекте возможности вывода его из устойчивого или стационарного режима функционирования. Следует подчеркнуть, что в рамках данной работы воздействие информационного потока на социальные объекты не является физическим взаимодействием, а носит чисто информационный характер. Это

определяет способ воздействия информационного потока, который отображается в виде изменения описания социального объекта, происходящее из-за синтеза информации из информационного потока с информационным описанием социального объекта. В этом случае могут иметь место следующие ситуации:

- когда в результате объединения информационных компонент с информационным описанием y_i формируется результирующее описание, которое приобретает новые значения семантических параметров, например, увеличивается величина семантической противоречивости и т.д.;

- ситуация, в которой семантические параметры не приводят к значимым изменениям в результате воздействия на них информационного потока.

Взаимодействие информационных компонент информационного потока W_i и описания социального объекта из W_s , который представляется в нормализованном виде, может реализовываться с помощью различных схем, определяющих процедуры синтеза. Нормализация представляет собой ограничение способов описания y_i , ограничение правил их преобразования по отношению к возможностям грамматических правил естественного языка, а также ограничение количества элементов x_i , которые могут использоваться для описания объектов y_i .

Рассмотрим на качественном уровне возможные процедуры синтеза W_i и W_s . Для этого введем следующие определения, связанные с сопоставлением описаний информационных потоков и W_s .

Определение 9. Описание Q_i представленное в порции информационного пакета $\omega_i^s \in W_S$ сопоставимо с описанием y_i , если количество элементов x_i , каждый из которых соответствует слову атрибуту предметной области, в ω_i не меньше чем на $\delta(y_i)$, количества слов соответствующих атрибутам W_s в описании y_i .

Формально это определение можно записать в форме:

$$u(y_i, \omega_i) = \delta(y_i) - \left[\sum_j^m (x_{ij} \in y_i) - \sum_{j=1}^k (x_{ij}^* \in \omega_i) \right],$$

где $u(y_i, \omega_i)$ - величина сопоставимости ω_i с y_i в W_s .

Процесс синтеза фрагмента $\omega_i \in W_i$ с y_i может происходить различными способами. Принимая во внимание, что синтезируются два сопоставимые текстовые описания, можно выделить следующие методы синтеза: конкатенация целых текстовых описаний; размещение предложений ψ_i из ω_i в среде описания социальных объектов; модификация $\psi_i \in y_i$ фразами из ω_i , которые будем обозначать символом $\varphi_i(\omega_i)$; модификация $(\psi_i \vee \varphi_i) \in y_i$ компонентами x_i , которые выбираются из семантического

словаря S_c , на основе процедур вывода, которые иницируются W_I . Алгоритм задачи синтеза формируется на основе анализа величин семантических параметров y_i и ω_i и интерпретации этих величин. Поскольку социальный объект представляет собой некоторую совокупность моделей, которые описываются в текстовой форме, то поступающая к ним информация действует на соответствующих представителях y_i в зависимости от значений семантических параметров, которые, по существу, характеризуют возможную эффективность воздействия этой информации на соответствующий социальный объект.

Определение 10. Предсказуемость реакции y_i на ω_i или воздействия ω_i на y_i определяется семантическими параметрами ω_i , которые формируются в процессе моделирования взаимодействия $W_I \rightarrow W_S$.

Наиболее сложным методом синтеза ω_i с y_i является метод, который основывается на использовании семантических правил вывода. При этом, появляется возможность, в результате реализации такого вывода, использовать новое слово, которое в тексте *ТИМ* не использовалось, но имеется в семантических словарях S_c предметной области W_s . Такие правила определяются на основе анализа семантических параметров с использованием логических связей между отдельными элементами, которые являются переменными соответствующих правил. В связи со специфическим характером таких переменных, в состав таких правил, кроме классических правил вывода [6], вводятся расширения правил, рассмотрение которых требует принятия ряда формальных ограничений и соответствующих логических схем.

Семантическая непрерывность текста $\sigma(\psi)$ определяется в рамках предложения ψ или отдельного фрагмента текста, который состоит из ряда предложений, что записывается в виде следующего соотношения:

$$\sigma(\Phi_i) = F\{\sigma_1(\psi_{i1}), \dots, \sigma_k(\psi_{ik})\} = F[f_1(\varphi_{i1}^1, \dots, \varphi_{in}^1), \dots, f_k(\varphi_{i1}^k, \dots, \varphi_{im}^k)], \quad (9)$$

где Φ_i – фрагмент текстового описания, который состоит из k предложений ψ_i . Семантические параметры η и γ , которые определяют возможность появления разрыва в $\sigma(\Phi_i)$ изменяются в заданном интервале, который будем называть интервалом активности каждого из параметров, поэтому необходимо согласовать между собой, с одной стороны, диапазоны изменений семантических параметров с диапазоном изменения параметра $\sigma(\Phi)$, а с другой стороны, согласовать диапазон изменения $\sigma(\Phi)$ с процессом модификации y_i . Введем переменную z_i , характеризующую семантические параметры, которые не являются активными, при воздействии P_i на y_i . В этом случае, можно ввести следующее определение.

Определение 11. Семантическая непрерывность σ представляет собой параметрическую функцию, которая определена в пространстве (η, γ, z_i) таким

образом, что ее значения не попадают в под-диапазоны критических значений соответствующих координат.

Формально, это определение можно записать в виде:

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

Определение 12. Семантическая открытость χ_i фрагмента Φ_i определяет меру доступности соответствующего фрагмента Φ_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 этого семантического параметра можно представить следующим образом:

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

где α – коэффициент согласования размерностей, $F(\cdot)$ – функция которая описывает итоговое влияние параметров $\sigma(\Phi_i)$ и $\pi(\varphi_i)$ на величину $\chi(\Phi_i)$, а m – количество фраз в фрагменте Φ_i .

Выводы.

В рамках данной работы вводится ряд ограничений и условий, которые принимаются как исходные условия для рассматриваемых CO . К таким условиям можно отнести следующие:

Условие 1. Процесс функционирования объекта социального типа является устойчивым при отсутствии P_i и параметры CO изменяются в процессе такого функционирования в рамках ограничений, сформированных для каждого отдельного CO .

Условие 2. Модель TIM не отображает процессов устойчивого функционирования объектов социального типа и предназначена только для описания процессов функционирования CO , которые обуславливаются внешними факторами и приводят к выводу из состояния равновесия соответствующий социальный объект.

Условие 3. Изменения в CO в результате воздействия на него внешних факторов выражается в изменениях текстового описания соответствующего CO в рамках TIM .

Условие 4. Носителем внешнего воздействия на CO является соответствующее текстовое описание P_i , которое составлено на языке, используемом для текстового описания TIM .

Отдельный элемент социального объекта, как и CO в целом, является инерциальным. Мера этой инерционности влияет на возможную эффективность внешнего воздействия. В связи с этим, возникает проблема определения величины соответствующей инерционности некоторого CO .

Определение 13. Величина инерционности CO определяется величиной изменения семантических параметров, которая приводит к изменениям в моделированном объекте.

Это определение означает, что эффективность внешнего воздействия на моделируемый объект определяется величиной изменения значения семантических параметров элементов $TIM \delta P_i^S$, который определяется для отдельных CO на основе экспериментальных исследований.

1. *Нечипоренко В.И.* Структурный анализ систем. М.: Советское радио, 1977. -216 с.
2. *Ядов В.А.* Стратегия социологического исследования. М.: Добросвет, 1998. -325 с.
3. Психология масс. Хрестоматия. Самара: Издательский Дом «БАХРАХ-М», 2006. - 592 с.
4. *Гийом Г.* Принципы теоретической лингвистики. М.: Ком-Книга, 2006. – 256 с.
5. *Сеиш А.* Очерк логической структуры предложений. М.: Ком-Книга, 2006. – 115 с.
6. *Акритас А.* Основы компьютерной алгебры с приложениями. М.: Мир, 1994. – 544 с.

Поступила 21.02.2011г.

УДК 683.06

Є.Д.Бабинець

ЗАГАЛЬНА ОРГАНІЗАЦІЯ ВИКОРИСТАННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АВТОМАТИЗОВАНОГО ПРЕКТУВАННЯ ОБРАЗУ КНИЖКОВОГО ВИДАННЯ

Інформаційна технологія, по своєму визначенню, представляє собою певну сукупність інформаційних засобів, з допомогою яких можна реалізувати необхідні технологічні процеси для створення певних продуктів [1,2]. Це означає, що відповідна технологія представляє собою повну систему окремих компонент, до яких відносяться:

- алгоритми аналізу та перетворення даних,
- бази даних, вміщують дані про предметну область, в якій