

УДК 519.1 : 681.3

Ю.В. Паржин

Национальный технический университет «ХПИ», Харьков

О ФУНДАМЕНТАЛЬНОМ СВОЙСТВЕ НЕВЫРАЗИМОСТИ РЕПРЕЗЕНТАТИВНЫХ ФОРМАЛЬНЫХ СИСТЕМ

В работе впервые предложена и обоснована модель формальной системы интеллектуального типа (ФСИТ), содержащая системы презентации и репрезентации, определены функции данных систем, доказана необоснованность применения теоремы Геделя о неполноте формальной арифметики к ФСИТ, репрезентативные системы которых обладают неполнотой ввиду существования фундаментального свойства α - и β -невыразимости, определенного в работе.

Ключевые слова: формальная система интеллектуального типа, системы презентации и репрезентации.

Актуальность и цель исследований

Возможность создания искусственного интеллекта (ИИ) широко обсуждается и исследуется в различных сферах науки: философии, лингвистике, математике, нейрофизиологии, психологии, кибернетике, компьютерной технике и даже физике. Однако, до сих пор, достижения в области исследований ИИ имеют весьма скромные успехи, все еще ведутся споры о возможности и формах создания искусственного интеллекта. Большое количество исследователей считают, что одной из базовых ограничительных теорем, ставящих под сомнение принципиальную возможность создания ИИ, является известная теорема Геделя о неполноте формальной арифметики и индуцирование ее результатов на любые формальные системы – невозможность создания формальных систем в которых может быть непротиворечиво доказано неперечислимое множество истин [1 – 3].

Вопрос заключается в том, действительно ли результаты данной теоремы говорят о наличии непреодолимых препятствий на пути создания формальных систем интеллектуального типа (ФСИТ), адекватно моделирующих мыслительную деятельность человека, или они лишь формально подтверждают общепризнанную доктрину невозможности человеческого познания всего (всех истин) мироздания? Обсуждение данного вопроса находится в тесной связи (как и сама теорема Геделя) с основными положениями теории истинности Тарского, а также современных взглядов на проблему истинности Филда и Хинтички [4, 5]. Следует отметить, что невыразимость семантики для формальных языков (ФЯ) или невыразимость истины по Тарскому, с позиции универсионализма, связана, прежде всего, с отсутствием метаязыка более "сильного", чем любой семантически богатый ФЯ (к числу которых может быть отнесен и естественный язык (ЕЯ) человеческого общения (коммуникации)), в рамках которого можно было бы описать данный ФЯ.

Результаты исследований

Современный взгляд на проблему построения интеллектуальных систем (ИС) – фреймворк-подход, использование семантических сетей и продукционных моделей, во многом сформирован по воздействию идей М. Мински [6], рассматривавшего модель интеллектуальной деятельности человека как последовательную взаимосвязь процессов: "окружающий мир → процесс восприятия → процесс распознавания → процессы мышления на ЕЯ и воображения".

Однако, последние исследования в области нейрофизиологии, в частности работы Д. Хьюбела и Т. Визеля [7], результаты исследований В. Маунткэтла и Дж. Хокинса [8, 9] говорят о необходимости пересмотра как данной модели Мински, так и модели нейронного элемента Маккаллоха-Питтса, лежащей в основе современного взгляда на построение нейронных сетей [6].

Если ИИ может быть создан, то ИС его реализующая, или точнее ФСИТ, поскольку данная система должна быть алгоритмична, должна соответствовать следующей модели – рис. 1.

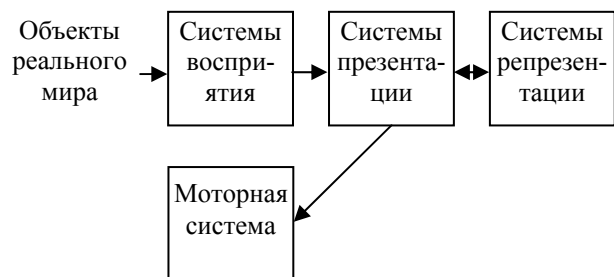


Рис. 1. Модель ФСИТ

Систем восприятия ФСИТ может существовать такое же количество, как и у человека (или более того), однако, для простоты исследований в дальнейшем будем рассматривать только одну наиболее информативную систему визуального (зрительного)

восприятия, так как по гипотезе Маунткастла все системы должны функционировать по одному и тому же алгоритму.

Система презентации или презентативная система – система первичного (доязыкового) представления – отображения моделей (образов) реального мира. Одними из основных функций данной системы являются функции: запоминания (оперативного и долговременного), структурирования (создания структур первичного представления), ассоциативного поиска и активизации (выборки) информации по структурам первичного представления, сравнения активизированных структур (элементов структур) первичного представления, актуализации и установления взаимосвязи структур, распознавания - классификации и идентификации образов, первичного управления моторной системой, осуществления взаимосвязи элементов (структур) системы репрезентации, определе-

ния первичных качественных шкал и др.

Система репрезентации или репрезентативная система – система вторичного (языкового) представления – отображения образов реального мира. Для ИС естественно-языкового типа система репрезентации выполняет также коммуникативную функцию, функцию вторичного (сознательного) управления моторной системой, функцию построения семантических структур – модели функций сознания человека. Одними из возможных видов данных семантических структур могут быть вторичные (производные) абстрактные языки, например, язык математики и др.

Рассмотрим особенности построения систем презентации и репрезентации как формальных систем. Данные системы должны обладать наборами признаков (элементов и функций) формальных систем, представленными в табл. 1.

Таблица 1

Признаки системы репрезентации

Признаки системы презентации	Признаки системы репрезентации
– множество детектируемых элементарных входных сигналов – алфавит входных сигналов системы (языка системы) – А	– алфавит системы репрезентации – алфавит ФЯ (ЕЯ) – Е
– алгоритм обработки детектируемых сигналов – алгоритм (множество правил) формирования структур первичного представления	– алгоритм (множество правил) построения правильно построенных формул (ППФ) языка системы
– множество правильно построенных структур (ППС) первичного представления (правильно построенных формул (слов) языка первичного представления)	– множество ППФ (слов) языка системы
– множество ППС отображаемых (имеющих взаимосвязь) в ППФ системы репрезентации или актуально истинных ППС	– множество ППФ, являющихся образами ППС, или истинными формулами (словами) системы – аксиомами системы
– алгоритм (множество правил формирования) взаимосвязанных ППС – алгоритм вывода структур ППС (фраз)	– алгоритм (множество правил) вывода истинных структур (фраз или предложений) ППФ

Алфавит А системы презентации, в данном случае, формируется на основе детектирования (идентификации) сигналов, поступающих из системы визуального восприятия. В соответствии с результатами исследований [7], алфавит А представим в виде ориентационных функциональных единиц первичной зрительной коры мозга (V1 неокортекса). Для простоты, воспользуемся моделью "ледяных кубиков" Хьюбела и Визеля. Данная модель отображает процесс детектирования элементарных ориентированных отрезков (ЭОО), например, контура изображения, что соответствует образованию ориентационных колонок – специфической реакции нейронов стриарной коры (поле 17) головного мозга.

Определим, например, детектируемые направления ориентации так, как это показано на рис.2. Тогда алфавит А можно представить в виде последовательности из 16-ти матриц A^k , состоящих из детекторов a^k_{ij} , определяющих буквы алфавита – ЭОО (рис. 3).

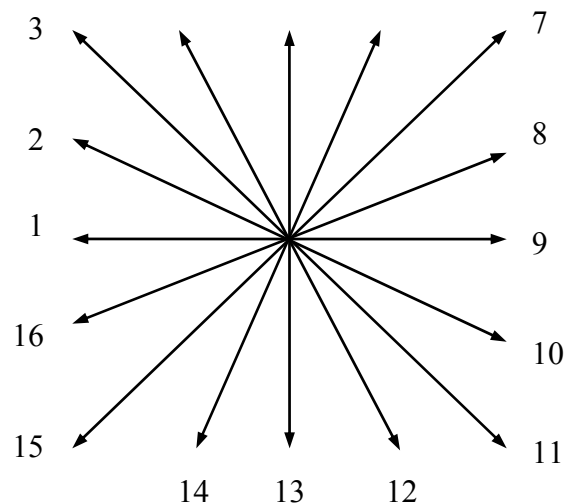


Рис. 2. Направления ориентации ЭОО

В свою очередь, *алфавит системы репрезентации Е* состоит, например, из букв русского алфавита или любого другого языка.

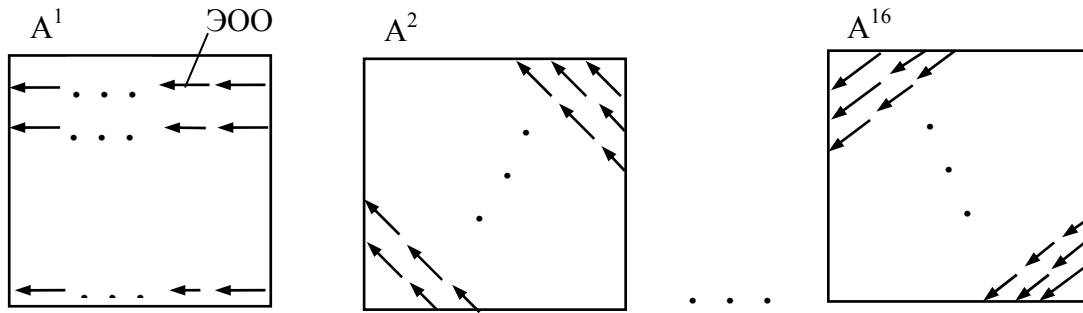


Рис. 3. Алфавит А системы презентации

В содержательном смысле, оба алфавита непосредственно не могут быть отображены друг в друга. Докажем следующую теорему.

Теорема 1. Для алфавитов А и Е презентативной и репрезентативной систем ФСИТ не существует функции отображения f , такой, что верно $E = f(A)$ или $A = f^{-1}(E)$.

Доказательство. Что представляют собой элементы алфавита А, выраженные с помощью элементов алфавита Е? Это выражения – слова, фразы или предложения ЕЯ, т.е. не только ППФ репрезентативной системы, но и истинные, имеющие содержательный смысл выражения (аксиомы), сформированные на основе семантического определения результатов процесса распознавания элементов структур объектов реального мира (объектов вне ФСИТ), осуществленного системой презентации.

Под *содержательным смыслом* следует понимать существование значений выражений (слов, фраз или предложений) системы репрезентации при выполнении репрезентативной и коммуникативной функций.

Следовательно: $f(A) = t$, где $t \in T$, T – множество аксиом репрезентативной системы (выражений ЕЯ). Очевидно, что и $f^{-1}(t) = A$. Но, $t = g(E)$, причем $f \neq g$, так как g является функцией, определяющей правила построения выражений ЕЯ, т.е. g определяет синтаксис ФЯ репрезентативной системы. В свою очередь, f будет определять семантику подмножества t ФЯ репрезентативной системы.

Таким образом, не существует семантики (лексической семантики) алфавита Е, так как буква алфавита Е, в качестве объекта реального мира, выражает только элементарную коммуникативную единицу (коммуникативный звук), т.е. саму себя, а не семантический элемент структуры объекта реального мира. Следовательно, f определяет репрезентативную функцию ФСИТ, а g – ее коммуникативную функцию и не существует $E = f(A)$ или $A = f^{-1}(E)$. Теорема доказана.

Результаты данной теоремы можно проиллюстрировать следующей диаграммой, представленной на рис. 4.

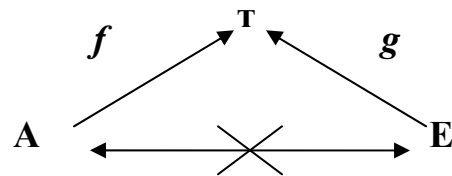


Рис. 4. Диаграмма отображений алфавитов А и Е

Следствие теоремы 1. Алфавит презентативной системы А семантически невыразим в алфавите Е репрезентативной системы ФСИТ.

Так как не существует семантической взаимосвязи между А и Е, причем $|A| \gg |E|$, то и выразить (сопоставить, отобразить) любой элемент алфавита А каким-либо элементом (буквой) алфавита Е невозможно.

Таким образом, следствие говорит о семантической невыразимости объектов реального мира, воспринимаемых ФСИТ, в алфавите Е репрезентативной системы.

Этот очевидный факт, тем не менее, носит фундаментальный (всеобщий) характер и имеет существенное значение для построения излагаемой теории.

Данную невыразимость назовем α -невыразимостью репрезентативной системы ФСИТ.

Если А адекватно отображает элементы или свойства объектов реального мира в презентативной системе ФСИТ, то, следовательно, слова – множество формул Ф презентативной системы, являющиеся ППС, адекватно отображают объекты реального мира в целом. Очевидно, что Ф также могут быть отображены в Т. Таким образом, возникает противоречие: алфавит А не может быть отображен в алфавит Е, а ППС (Ф), построенные из элементов А, могут быть отображены в аксиомы репрезентативной системы Т, построенные из элементов алфавита Е.

Данное противоречие отражает суть общепhilosophической и математической проблемы теории истинности, а также проблемы теории соответствия, лежащей в основе теории абдукции Пирса и выво-

дов Эйнштейна относительно "логической формы истинного отображения конкретных фактов", получивших в литературе название "пропасти Эйнштейна" [10 – 12].

Таким образом, наличие α - невыразимости репрезентативной системы подтверждает гипотезу о необходимости существования в структуре ФСИТ системы презентации с описанными выше свойствами. Данная система презентации и должна "заполнить" "пропасть Эйнштейна", т.е. реализовать отображение (взаимосвязь) объектов, процессов,

явлений реального мира (вне ФСИТ) в репрезентативную систему ФСИТ и решить проблему истинности данного отображения. Кроме того, ФЯ системы презентации будет являться расширением ФЯ (ЕЯ) репрезентативной системы, что в совокупности создает функционально более выразительный и логически богатый язык ФСИТ, который, в некотором смысле, может быть рассмотрен как метаязык.

Проиллюстрируем все вышесказанное следующей диаграммой (рис. 5).

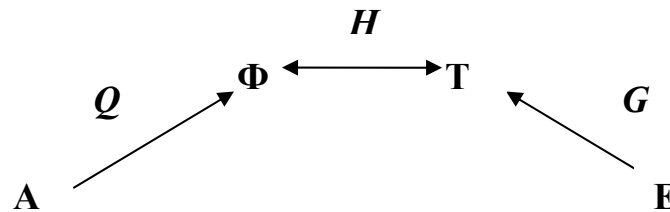


Рис. 5. Диаграмма отображений элементов систем презентации и репрезентации

На данной диаграмме представлено, что $T = H(\Phi)$ и $\Phi = H^{-1}(T)$, где H – биективное отображение. Очевидно, что для некоторых $t \in T$ и $\phi \in \Phi$, может не существовать отображения H , что свидетельствует о наличии основного алгоритма установления данного отображения как "установление извне" или установление в результате процедуры обучения с учителем. Исходя из данного положения, назовем $t = h(\phi)$, где $h \in H$, семантическим определителем ϕ . Множество Φ на котором установлено отображение H и является множеством актуально-истинных ППС.

В системе презентации ФСИТ, сюръективное отображение Q определяет алгоритм обработки детектируемых сигналов или множество правил формирования структур первичного представления $\Phi = Q(A)$.

В свою очередь, в системе репрезентации ФСИТ, отображение G определяет алгоритм (множество правил) построения ППФ и их структур, $g \in G$.

Тогда, $T = G(E)$ выражает коммуникативную функцию репрезентативной системы ФСИТ, а $T = H(Q(A))$ или $T = F(A)$, где $f \in F$ – ее репрезентативную функцию. Таким образом, отображение H , в некотором смысле, можно считать "формализацией содержательного понятия истинны" [10]

Предположим, что существует непосредственно (в данный момент времени t_0) наблюдаемый системой восприятия объект реального мира $r_z \in R$, где $z \in Z$, Z – множество объектов реального мира, доступных человеку для восприятия в процессе всей жизни. Можно также предположить, что Z перечислимо (конечно). Однако в данной интерпретации r_z определяется как классовый объект, который также

может иметь конечное множество реальных примеров P , каждый из которых, в свою очередь, может наблюдаться в конечном множестве состояний S . Поэтому, будем интерпретировать r_z как непосредственно наблюдаемое состояние примера классового объекта $z \in K$, $K \subset Z$, где K – множество классовых объектов или множество классов. Следует также уточнить, что, как правило, в момент времени t_0 наблюдается не один объект, а множество объектов, находящихся в определенной ситуативной взаимосвязи, или ситуации SI_j , причем, множество SI для конечных систем (с конечным жизненным циклом или конечным числом воспринимаемых образов) также будет конечным. Данный простой вывод ведет к следствию о конечности множества T и видимой необоснованности "геделевского аргумента" против возможности создания ФСИТ [2], хотя, пока что, не доказывает конечности всех истин репрезентативной системы T ($T \in T$), для которых, на данном этапе исследований, не выполняется условие $T = F(A)$.

Конечно, если множествам K, P, S, SI , а значит и T придать смысл актуальной или даже потенциальной бесконечности для гипотетических идеальных ФСИТ, то очевидно, что подобных непротиворечивых формальных систем в соответствии с указанным "аргументом" построить нельзя.

Введем r_z в диаграмму отображений (рис. 6), где D – функция декомпозиционного отображения (детектирования), выполняемая системой восприятия ФСИТ.

Безусловно, что абсолютный прообраз – объект реального мира обладает определенной степенью априорной невыразимости ввиду функциональных ограничений системы восприятия ФСИТ.

Однако, для простоты изложения, предположим, что r_z – объект реального мира, воспринимаемый ФСИТ в виде двуградационного контурного изображения с унарной толщиной контурных отрезков и системой ориентации, соответствующей направлениям, представленным на рис.2.

Тогда, алфавит A будет полностью соответствовать системе, которая представлена на рис.3., и $\{a_{ij}^k\} = D(r_z)$.

При этом множество одновременно возбужденных элементов a_{ij}^k в момент времени t_0 можно выразить суммой

$$\sum_k \sum_{ij} a_{ij}^k(r_z). \quad (1)$$

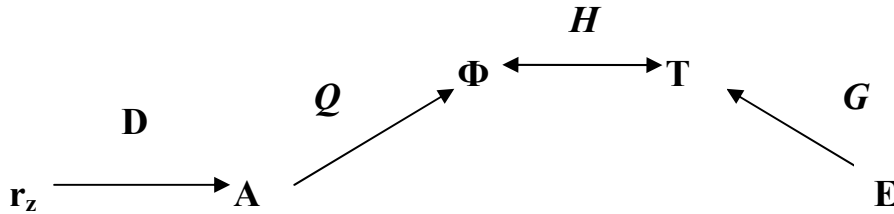


Рис. 6. Диаграмма отображений элементов систем в ФСИТ

Определим, что объект r_z является распознанным (классифицированным) ФСИТ, если существует:

1. $\phi_z = g(\{a_{ij}^k\})$;
2. $\tau_z = h(\phi_z)$ и $\phi_z = h^{-1}(\tau_z)$.

Определение. Назовем $\phi_z = h^{-1}(\tau_z)$ редуцированным образом или $red\phi_z$.

Существование редуцированных образов подтверждается тем, что данные представления (презентации) могут быть отображены моторной системой человека, например, в виде наиболее примитивных графических изображений, которые назовем $M(red\phi^1)$, где $red\phi^1$ – редуцированные образы 1-го уровня редукции, $red\phi^1 \subseteq \phi$.

Очевидно, что при изменении состояния объекта или примера объекта определенного класса z , τ_z остается неизменным (сюръекция R в T).

Тогда,

$$E = \frac{\sum_n \sum_{k,ij} a_{ij}^k(r_z)}{\sum_{k,ij} a_{ij}^k(M(red\phi_z^1))}, \quad (2)$$

где E – средняя редуционная энтропия образов объекта r_z ; n – мощность множества z . Если в знаменателе (2) не учитывать n , то получим абсолютную редуционную энтропию.

Редуционная энтропия показывает наличие множества элементов $D(r_z)$, "потерянных" в результате их отображения $q \in Q$ в ϕ_z , и, следовательно, невыразимых в τ_z . Данную невыразимость назовем β -невыразимостью репрезентативной системы ФСИТ. Указанный эффект "потери" информации в репрезентативной системе на самом деле является результатом применения алгоритма структурного обобщения с сохранением иерархической структуры данных, поэтому говорить о невыразимости "поте-

рянной" информации можно лишь применительно к репрезентативной системе. Кроме того, β -невыразимость свидетельствует о *неполноте репрезентативной системы ФСИТ* относительно репрезентативной системы, а значит и относительно классового множества (множества, элементы которого принадлежат одному классу объектов) объектов реального мира, воспринимаемых ФСИТ. Действительно, τ_z не зависит от мощности множества примеров или наблюдаемых состояний объектов класса z и, следовательно, детектируемое изменение данных примеров или состояний в репрезентативной системе не повлечет изменений в определении ϕ_z .

Приведенные выше рассуждения ведут к следующему утверждению.

Утверждение 1. Любая репрезентативная система ФСИТ обладает неполнотой относительно множества объектов восприятия Z .

Данное утверждение говорит о том, что любая ФС, алфавит языка которой обладает α -невыразимостью, обладает неполнотой, т.е. не все объекты реального мира детектируемые системой восприятия ФСИТ в виде множества $D(r_z)$, а значит и актуально-истинные в репрезентативной системе, имеют биективное отображение в репрезентативной системе. То есть, в виду β -невыразимости их истинность в репрезентативной системе не может быть доказана – не существует дедуктики [3] или алгоритма, выполняющего функцию f .

Таким образом, предположение о конечности множества T и утверждение 1 говорят о необоснованности применения теоремы Геделя о неполноте к ФСИТ, так как нарушаются условия данной теоремы.

Очевидно, что существуют и другие виды невыразимости репрезентативной системы ФСИТ, например, такие как: ситуационная невыразимость,

невыразимость действий (процессов), пространственно-временная и причинно-следственная невыразимость, чувственная и метафизическая невыразимость.

Возможно, что именно невыразимость и связанная с ней неполнота репрезентативных систем лежит в основе их противоречивости, выражающейся в существовании логических парадоксов. Возможно также, что устранение данной противоречивости может быть осуществлено в результате расширения ФЯ (ЕЯ) репрезентативной системы за счет ФЯ системы презентации.

Одним из косвенных результатов вывода о необходимости существования презентативной и репрезентативной систем ФСИТ, является подтверждение возможности использования теста Тьюринга для определения интеллектуальности ФСИТ, а вернее ее репрезентативной системы, и опровержение критики данного теста Сирлом в эксперименте "китайская комната" ввиду того, что прохождение репрезентативной системой данного теста невозможно при отсутствии презентативной системы.

Выводы

В данной работе впервые предложена и обоснована модель ФСИТ, содержащая системы презентации и репрезентации, определены функции данных систем, доказана необоснованность применения теоремы Геделя о неполноте формальной арифметики к ФСИТ, репрезентативные системы которых обладают фундаментальным (изначальным) свойством неполноты ввиду их α - и β -невыразимости.

Список литературы

1. Пенроуз Р. Тени разума: пер. с англ. / Р. Пенроуз. – 1993. – 252 с. – [Электронный ресурс]. – Режим доступа к документу: <http://hotmix.narod.ru>.
2. Иванов Е.М. К проблеме "вычислимости" функции сознания / Е.М. Иванов. – [Электронный ресурс]. – Режим доступа к док.: www.filosof.historic.ru/books.

3. Успенский В.А. Теорема Геделя о неполноте. (Популярные лекции по математике) / В.А.Успенский. – М.: Наука, 1982. – 112 с.
4. Field H. Tarski A. Theory of Truth // The Journal of Philosophy, LXIX: 13 (1972). – P. 347-375. – Перепечатано в: Meaning and Truth: Essential Readings in Modern Semantics (ed. by J.L.Garfield and M.Kiteley).– N.Y.: Paragon, 1991. – P. 271-296.
5. Хинтиikka Я. Истина после Тарского: пер. с англ. С. Гарина / Я. Хинтиikka. – [Электронный ресурс]. – Режим доступа к док.: <http://korfo.kubsu.ru>
6. Бондарев В.Н. Искусственный интеллект: учебное пособие для вузов / В.Н. Бондарев, Ф.Г. Аде. – Севастополь: СевНТУ, 2002. – 615 с.
7. Хьюбел Д. Глаз, мозг, зрение: пер. с англ. / Д. Хьюбел. – М.: Мир, 1990. – 239 с.
8. Mountcastle Vernon B. An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System / Mountcastle Vernon B. // Gerald M. Edelman and Vernon B. Mountcastle, eds., The Mindful Brain (Cambridge, Mass.: MIT Press, 1978).
9. Хокинс Дж. Об интеллекте: пер. с англ. / Дж. Хокинс, С. Блейкли. – М.: ООО "И.Д. Вильямс", 2007. – 240 с.
10. Джон У. Оллер-мл. Глава 7 из сборника (под ред. Дж. П. Морлэнда) "The Creation Hypothesis: Scientific Evidence for an Intelligent Designer"/ Джон У. Оллер-мл., Джон Л. Олдак. – [Электронный ресурс]. – Режим доступа к док.: www.christianbooks.hotmail.ru
11. Peirce Charles S. Collected Papers of C S. Peirce, ed. Charles Hartshorne and Paul Weiss, 8 vols. (Cambridge, Mass.: Belknap/Harvard University Press, 1931 – 1958), 1:128.
12. Albert Einstein. The Common Language of Science / Albert Einstein // Out of My Later Years (Secaucus, N.J.: Citadel, 1956). – P. 111.

Поступила в редколлегию 6.11.2008

Рецензент: д-р техн. наук, проф. В.А. Краснобаев, Харьковский национальный технический университет сельского хозяйства им. П. Василенко, Харьков.

ПРО ФУНДАМЕНТАЛЬНУ ВЛАСТИВІСТЬ НЕВИМОВНОСТІ РЕПРЕЗЕНТАТИВНИХ ФОРМАЛЬНИХ СИСТЕМ

Ю.В. Паржин

У роботі вперше запропоновано і обґрунтовано модель формальної системи інтелектуального типу (ФСИТ), що містить системи презентації і репрезентації, визначені функції даних систем, доведено необґрунтованість застосування теореми Геделя про неповноту формальної арифметики до ФСИТ, репрезентативні системи яких володіють неповнотою зважаючи на існування фундаментальної властивості α - і β -невимовності, визначеного в роботі.

Ключові слова: формальна система інтелектуального типу, системи презентації і репрезентації.

ABOUT FUNDAMENTAL PROPERTY OF INEXPRESSIVENESS OF REPRESENTATION FORMAL SYSTEMS

Yu.V. Parzhin

In work first offered and grounded model of the formal system of intellectual type (FSIT), containing the systems of presentation and representation, the functions of these systems are certain, the groundlessness of application of theorem of Godel is proved about incompleteness of formal arithmetic to FSIT, the representation systems of which possess incompleteness because of existence of fundamental property α - and β -beyond expression, certain in work.

Keywords: formal system of intellectual type, systems of presentation and representation.