
**ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ
СИСТЕМЫ**

УДК 15 + 007

© В. Я. Мамаев, Д. А. Горбунов

МОДЕЛИ И АЛГОРИТМЫ КОНТРОЛЯ ЗНАНИЙ ОБУЧАЕМЫХ НА ОСНОВЕ СЕТЕЙ ПЕТРИ

В связи с развитием вычислительной техники и информационных технологий появляется возможность применять искусственный интеллект в автоматизированных обучающих системах. Одной из проблем, которую можно решить в автоматизированных обучающих системах с помощью сетей Петри, является контроль знаний в случае, когда в качестве тестовых заданий применяются задания открытого типа с максимально свободной формой изложения ответа обучаемого. Данная статья посвящена решению проблемы понимания средствами ЭВМ естественного языка и проблемы оценки полученных ответов обучаемых.

ФОРМУЛИРОВКА И АКТУАЛЬНОСТЬ РЕШАЕМОЙ ПРОБЛЕМЫ

В связи с развитием вычислительной техники и информационных технологий появляется возможность применять искусственный интеллект (ИИ) в автоматизированных обучающих системах (АОС). Одной из проблем, которую можно решить в АОС с помощью ИИ на основе сетей Петри (СП), является контроль знаний в случае, когда в качестве тестовых заданий (Тест3) применяются Тест3 открытого типа с максимально свободной формой изложения ответа обучаемого. До сих пор решение этой проблемы сдерживалось из-за непонимания ЭВМ естественного языка (ЕЯ) и невозможности оценки машиной полученных ответов.

ИЗВЕСТНЫЕ ПОДХОДЫ К РЕШЕНИЮ

В системах ИИ (системах машинного перевода, экспертных системах (ЭС)) эту задачу обычно решает лингвистический процессор (ЛП) [1]. Формальный аппарат ЛП — это метаязык лингвистического описания, позволяющий представить информацию о языке в наглядной и математически строгой форме и облегчающий объективную проверку адекватности этого описания.

С точки зрения представления концептуальной модели удобно рассматривать ЛП относительно таких параметров системы, как сложность формального аппарата представления знания, сложность структур данных, временная сложность алгоритмов, универсальность представления знания. При этом формальный аппарат должен обеспечивать лингвистическую естественность, формальную мощь, вычислительную эффективность.

Данные служат для реализации формальной модели на компьютере. В то же время ими манипулируют алгоритмы, поэтому структура данных

является как бы связующим звеном между формальной моделью и формальным аппаратом ЛП.

Концептуальная модель ЛП приведена на рис. 1 и может быть представлена кортежем [1]

$$LP = \langle F, M, S, A, I \rangle,$$

где F — формальный аппарат; M — формальная модель; S — структура данных; A — алгоритм анализа и синтеза; I — интерфейс, преобразующий входные данные во внутреннее представление и наоборот.

Известные способы математического описания содержания ЕЯ в ЛП рассмотрены в [2] и используются: при разработке формальных языков-посредников, в грамматике Монтегю, в теории обобщенных кванторов, в теории структурированных значений предложений, в динамической предикатной логике, в теории представления дискурсов. Однако они требуют различных подходов для описания ЛП, что не позволяет использовать для этого единый формальный аппарат алгоритмов его работы и затрудняет анализ.

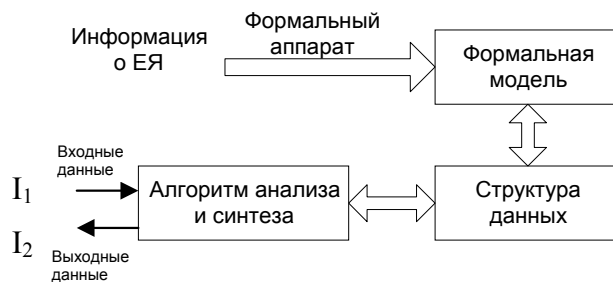


Рис. 1. Концептуальная схема ЛП

**ПРЕДЛАГАЕМЫЙ ПОДХОД,
ЕГО НОВИЗНА И ОСОБЕННОСТИ**

В качестве единого аппарата (метааппарата) для построения лингвистических моделей, для моделирования взаимодействия компонентов системы, таких как формальные модели, данные и алгоритмы, предлагается использовать СП.

Сеть Петри есть двудольный ориентированный граф и определяется [3, 4] как

$$C = \langle P, T, A, Mo \rangle, \tag{1}$$

где соответственно P — конечное множество позиций; T — конечное множество переходов; A — алгоритмические представления, а Mo — начальная маркировка сети.

Предлагаемый в качестве ИИ аппарат СП обладает следующими достоинствами:

- 1) обеспечивает единую форму реализации интерфейса при переходе от одной грамматики к другой;
- 2) легкость описания алгоритмов;
- 3) понятность (доступна пониманию математикам, лингвистам и инженерам);
- 4) модель, построенную на основе СП, легко оценить на устойчивость, обратимость, работоспособность, являющиеся важными свойствами любого программного комплекса.

Еще одним достоинством сетей Петри является возможность моделировать системы с переменной структурой. Для этих целей были разработаны специальные гибкие лингвистические сети.

Определим *позиции* как места, где происходит отделение лексемы от словоформы, а *переходы* — как операции по проверке на отделение соответствующей лексемы, связанной с условиями перехода. *Маркеры* являются носителями информационных ресурсов. Такую сеть назовем *гибкой лингвистической сетью Петри* и опишем в виде кортежа

$$C = \langle P, T, A, Mo, \Omega \rangle, \tag{2}$$

где Ω — таблица или функция, заданная таблично, и управляющая структурой сети по мере и смыслу поступающих сообщений в зависимости от событий, происходящих в реальной системе.

**ОСОБЕННОСТИ РЕАЛИЗАЦИИ
ПРЕДЛАГАЕМОГО ПОДХОДА**

Исходя из выбранного аппарата, в разрабатываемых ЛП было решено применить ряд формализмов.

1. На уровне морфологии была применена атрибутивно-типовая модель, представленная на основе СП и реализованная в виде отношений с помощью реляционных СУБД. На их основе были

смоделированы и алгоритмы морфологии, а также их взаимодействие со структурами данных.

2. На этапах анализа, синтеза и корректировки словоформ были применены лингвистические СП и создан программный комплекс, позволяющий на входе задать с их помощью сетевой шаблон (порядок следования морфем), а на выходе получить готовое разложение сети в виде набора путей.

3. На уровне синтаксиса разработана формальная модель синтаксических отношений с помощью СП, их типовая классификация и алгоритмы анализа. Таким образом, СП впервые столь широко использованы в ЛП.

4. База знаний (БЗ) (содержащаяся в реляционной СУБД) использовалась в рамках предметной области "Воздушная навигация" [5].

5. Реализация алгоритмов произведена с использованием технологии PHP + MySQL + Flash для применения наработок в системе дистанционного образования [6, 7].

В ЛП имеются два анализатора — *морфологический (МА)* и *синтаксический (СА)* (рис. 2) [8]. Каждый из них состоит из набора алгоритмов, реализующих обработку поступающей информации, и структур данных, служащих для извлечения необходимых для этого данных из формальной модели и хранения анализируемой информации. ЛП может работать и как СА, и как МА. В первом случае на вход поступает предложение, а на выходе выдается ею синтаксическая структура (СинтС); во втором — на вход подается словоформа (С), а на выходе выдается ее морфологическая структура (МорфС). При этом МА ЛП осуществляет не только анализ словоформы, но и проверку на корректность, а СА осуществляет еще и классификацию предложения по типам.

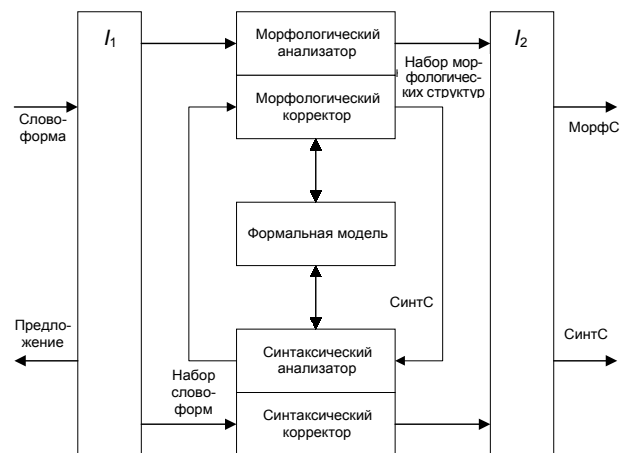


Рис. 2. Архитектура ЛП

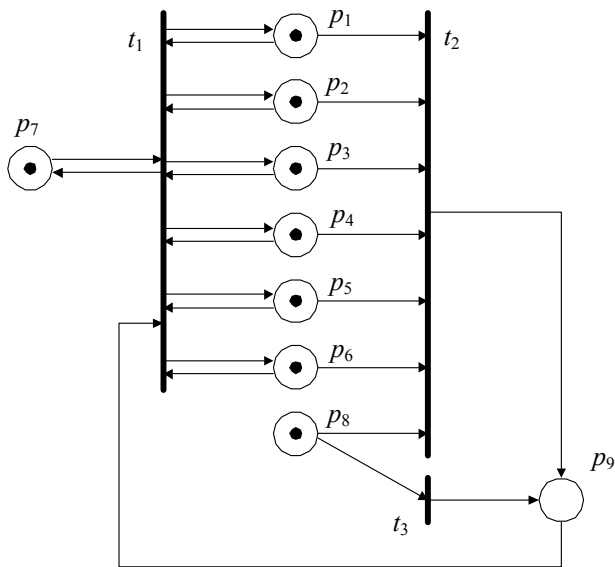


Рис. 3. Сетевая модель извлечения данных с помощью шаблонов. $p_{1-9} \in P$, $t_{1-3} \in T$

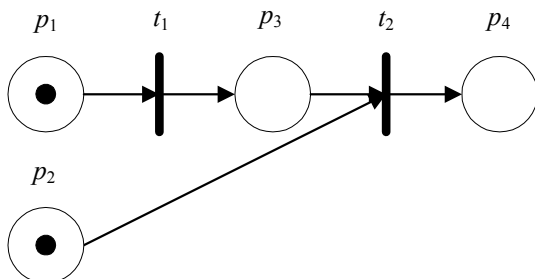


Рис. 4. Модель СинтО. $p_{1-4} \in P$, $t_{1-2} \in T$

Алгоритм *морфологического анализатора* моделируется с помощью СП, а реализуется в терминах, близких к языкам программирования. Принцип действия атрибутивно-типовой модели можно показать на примере сетевой модели (рис. 3) извлечения данных с помощью шаблонов (сетевая модель для атрибутов имеет схожий вид и структуру).

Основной задачей *синтаксического анализатора* ЛП является преобразование МорфС предложений, поступающих с выхода морфологического блока в СинтС. Так как МорфС предложения состоят из МорфС отдельных словоформ, то переход от МорфС предложения к его СинтС осуществляется путем установления синтаксических связей между МорфС слов и между ними самими. При этом МорфС отдельных словоформ служит для их установления. Такие связи принято называть син-

таксическими отношениями (СинтО), и, как показано в [9], все они могут быть смоделированы с помощью СП.

Для анализа простых предложений ЕЯ достаточно около 16 СинтО, что было выявлено путем анализа типов [1]. Все СинтО бинарные и ориентированные. Ориентированность подразумевает то, что все отношения представлены в формате: Y (левая часть)— X (правая часть), где X — главное слово, а Y — зависимое.

Введем следующие позиции и переходы: p_1 — содержит маркер с атрибутом X , который, возможно, является сказуемым; p_2 — содержит маркер с атрибутом Y , который, возможно, является подлежащим; p_3 — содержит маркер, если СинтО может быть установлено; p_4 — вспомогательная позиция; t_1 — проверка условия, что X — сказуемое; t_2 — проверка условия, что Y — подлежащее (существительное, местоимение или числительное).

На рис. 4 показано, как СП моделирует СинтО. При наличии в позиции p_1 сказуемого X срабатывает переход t_1 . Если это происходит, то при наличии в позиции p_2 подлежащего Y срабатывает переход t_2 . При этом устанавливается, что первое подлежаще-сказуемое отношение имеет место. Аналогично моделируются остальные 15 синтаксических отношений.

ПРИМЕРЫ ПРИМЕНЕНИЯ ПРЕДЛАГАЕМОГО ПОДХОДА

Рассмотрим применение СП для реализации морфологического и синтаксического анализаторов на примере четырех открытых ТестЗ, взятых из [5].

В табл. 1 в столбце "Текст ТестЗ" цифрами обозначены правильные варианты ответов. На основании перечня возможных ответов опишем глоссарий (табл. 2). Для каждой из словоформ представим ее лексическую структуру (табл. 3). Морфологическая структура рассмотрена в табл. 4.

В данном случае словоформа разбирается однозначно, однако при морфологическом анализе довольно частым явлением является морфологическая омонимия, т. е. ситуация, когда словоформа может быть разобрана более чем одним способом. Например, для слова *магнитный* имеем омонимы, приведенные в табл. 5.

Описание лексической структуры может быть представлено в виде (3):

$$\begin{aligned} \text{Причастие} = \\ = \langle \text{Префикс} \rangle \& \langle \text{Основа} \rangle \& \langle \text{Суффикс} \rangle \& \\ \& \langle \text{Окончание} \rangle \& \langle \text{Частица} \rangle. \end{aligned} \quad (3)$$

Оно позволяет создать экземпляры шаблонов.

Табл. 1. Тестовые задания (Тест3) в открытой форме

№	Название теста	Текст теста	Ответы
1	Метеорологическое направление ветра	Метеорологическим направлением ветра называется <u>1</u> , заключенный между <u>3</u> <u>2</u> <u>4</u> и <u>2</u> , откуда дует ветер	1. угол; 2. направление; 3. северный; 4. меридиан
2	Навигационное направление ветра	Навигационным направлением ветра (НВ) называется <u>1</u> , заключенный между <u>2</u> , принятым для ориентации <u>2</u> <u>3</u> , и <u>2</u> , куда дует ветер	1. угол; 2. направление; 3. полет
3	Магнитное склонение	Магнитным склонением (ΔM) называется <u>2</u> между <u>1</u> и <u>3</u> меридианами в данной точке	1. истинный; 2. угол; 3. магнитный
4	Азимутальная поправка	Азимутальной поправкой (ΔA) называется <u>1</u> между <u>3</u> и <u>2</u> меридианами	1. угол; 2. истинный; 3. условный

Табл. 2. Глоссарий для Тест3 из табл. 1

№	Словоформа	Часть речи
1	угол	Существительное
2	направление	—"
3	северный	Прилагательное
4	меридиан	Существительное
5	полет	—"
6	истинный	Прилагательное
7	магнитный	—"
8	условный	—"

Табл. 3. Лексическая структура словоформ из глоссария табл. 2

№	Словоформа	Префикс	Основа	Тема	Суффикс	Окончание	Частица
1	угол	—	угол	—	—	—	—
2	направление	—	направл	—	ен	ие	—
3	северный	—	север	—	н	ый	—
4	меридиан	—	меридиан	—	—	—	—
5	полет	—	полет	—	—	—	—
6	истинный	—	истин	—	н	ый	—
7	магнитный	—	магнит	—	н	ый	—
8	условный	—	услов	—	н	ый	—

Табл. 4. Морфологическая структура

Лексема без словоизменения	Атрибуты					
	Вид	Залог	Время	Число	Род	Падеж
угол	—	—	—	Ед.	Муж.	Им.
направление	—	—	—	—	Ср.	—
северный	—	—	—	—	Муж.	—
меридиан	—	—	—	—	—	—
полет	—	—	—	—	—	—
истинный	—	—	—	—	—	—
магнитный	—	—	—	—	—	—
условный	—	—	—	—	—	—

Табл. 5. Пример морфологической омонимии для слова *магнитный*

Лексемы	Часть речи	Атрибуты			
		Одушевленность	Род	Число	Падеж
магнитный	Прил.	Неодуш.	Муж.	Ед.	Им.
—	—	—	Ср.	—	—
—	—	—	Жен.	—	—
—	—	—	Муж., жен., ср.	Мн.	—

Табл. 6. Экземпляры шаблонов

Префикс	Основа	Суффикс	Окончание	Частица
#	1	#	+ #	#
#	2	#	ие	#
#	3	н	ые	#
#	4	н	ый	#

Примечание. Здесь используются следующие обозначения: + — наличие лексемы; # — отсутствие лексемы; | — логическое исключительное ИЛИ.

Помимо морфологического анализа система должна анализировать синтаксическую составляющую текста. Синтаксический анализ заключается в проверке правильной последовательности словоформ в определении.

Синтаксическое расположение словоформ представляется в виде матрицы-шаблона разрядностью $n \times m$, где n — количество строк, m — количество столбцов:

$$X_{mn} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1^{-1} & 0 \\ 0 & 1^{-1} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

Коэффициент 1 элемента матрицы (4) говорит

о том, что n -я словоформа может располагаться в m -й ячейке определения; 1^{-1} — информирует, что словоформы могут меняться ячейками без потери смысла определения; 0 — словоформа не может располагаться в данной ячейке ни в коем случае.

Описанная выше методика была реализована в программе на объектно-ориентированном языке программирования Borland Delphi 7.0 и апробирована в СПбГУАП. В программе были реализованы механизмы морфологического и синтаксического анализаторов при ответе на открытое ТестЗ (рис. 5). На первом этапе производится орфографический анализ введенной информации обучаемым, за который выставляется коэффициент M . На втором этапе производится синтаксическая оценка C введенного текста в открытое ТестЗ. По сумме двух коэффициентов система выставляет оценку обучаемому.

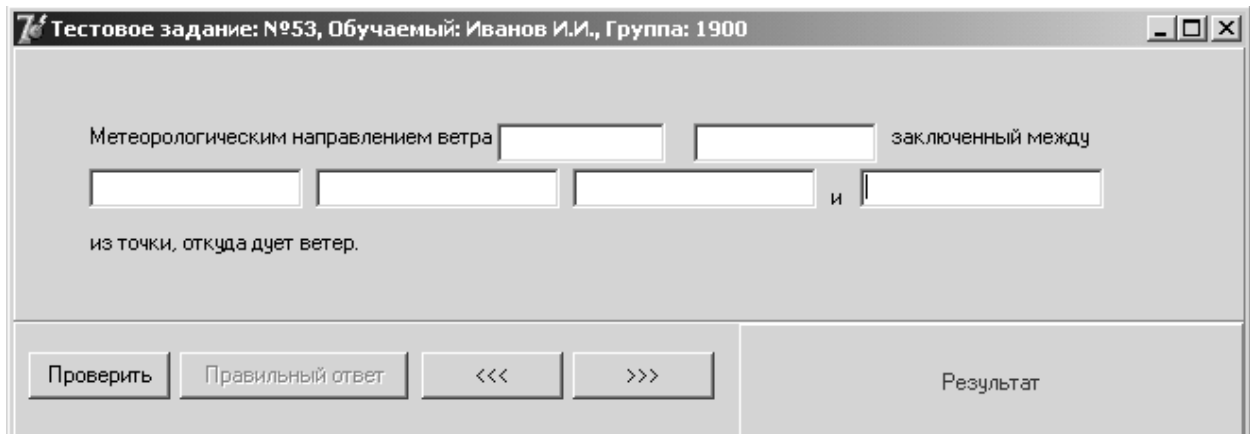


Рис. 5. Исходная форма Тест3

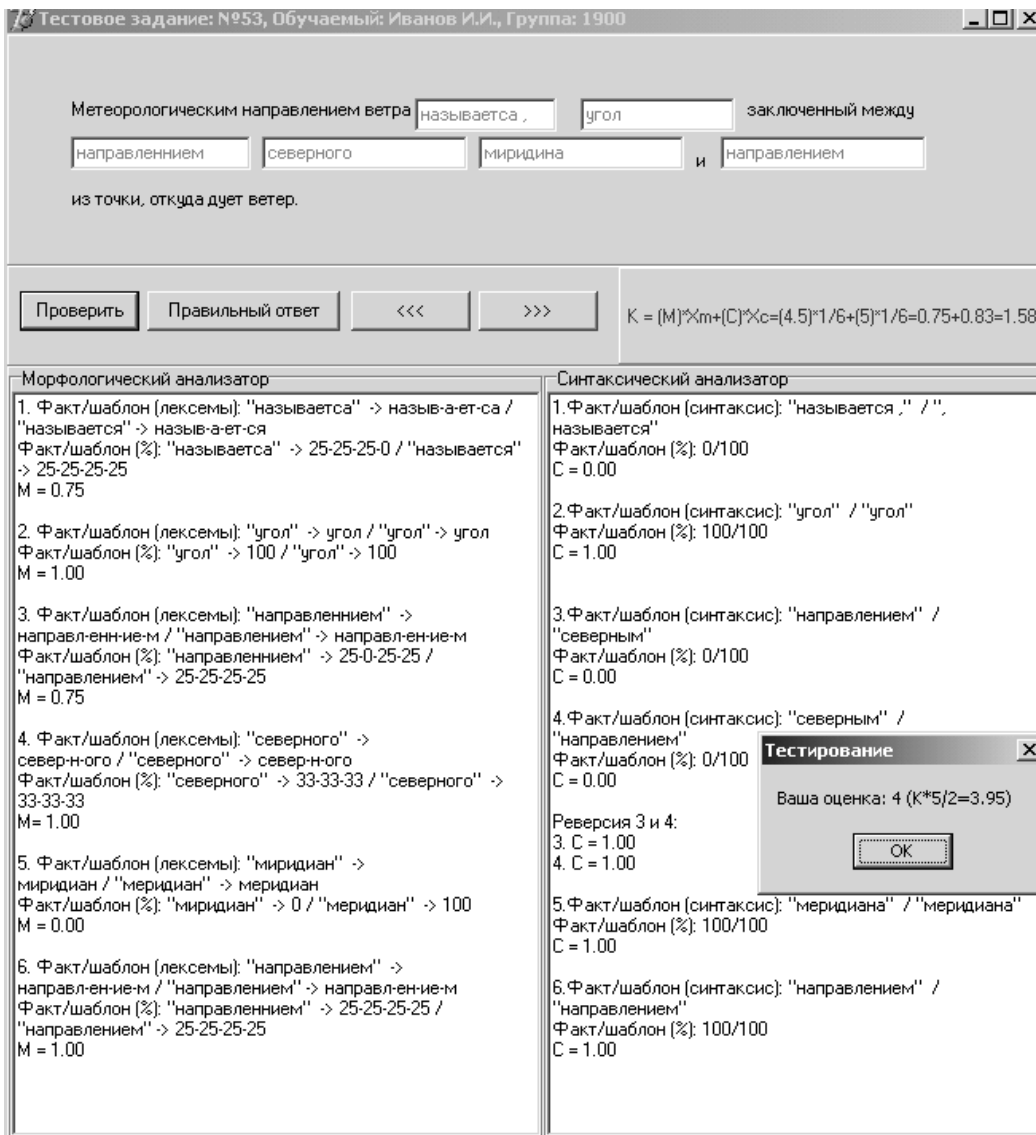


Рис. 6. Форма Тест3 после ввода исходных данных и проверки результата

Тестовое задание: №53, Обучаемый: Иванов И.И., Группа: 1900

Метеорологическим направлением ветра заключенный между и из точки, откуда дует ветер.

Проверить Правильный ответ <<< >>> Результат

Рис. 7. Форма Тест3 с правильным ответом

Система подробно описывает все ошибки обучаемого (рис. 6) и показывает правильный ответ (рис. 7).

ВЫВОДЫ

По результатам исследований с учетом практического применения реализованного подхода:

- уменьшилось потребное учебное время инструктора по сравнению с обучением без АОС;
- увеличилась объективность контроля знаний;
- дифференцировалась скорость обучения;
- расширилась интеллектуальность АОС (функцию анализа Тест3 выполняет ЭВМ, в том числе анализ открытых Тест3).

Предлагаемый к использованию аппарат позволяет подойти к анализу ЕЯ с точки зрения системы ИИ. Аппарат позволяет описывать сложные процессы распознавания ЕЯ с помощью СП и реализовать тестовый контроль знаний на основе Тест3 открытой формы. Применение СП дало возможность использовать единый подход и формат для всех структурных единиц ЛП.

СПИСОК ЛИТЕРАТУРЫ

1. Апресян Ю.Д., Богуславский И.М., Иодмин Л.Л. и др. Лингвистический процессор для сложных информационных систем. М., 1992. 256 с.
2. Фомичёв В.А. Формализация проектирования лингвистических процессоров. М.: Макс Пресс, 2005. 386 с.
3. Котов В.Е. Сети Петри. М.: Наука, 1984. 160 с.

4. Питерсон Дж. Теория сетей Петри и моделирование систем: Пер. с англ. М.: Мир, 1984. 264 с.
5. Мамаев В.Я., Синяков А.Н., Петров К.К., Горбунов Д.А. Воздушная навигация и элементы самолетовождения: Учебное пособие. СПб: СПбГУАП, 2002. 256 с.
6. Мамаев В.Я., Синяков А.Н., Петров К.К., Горбунов Д.А. Воздушная навигация и элементы самолетовождения (DOSHR): Электронное учебное пособие. Зарегистрировано в Реестре программ для ЭВМ: Свидетельство № 2003610194 от 17.01.2003.
7. Горбунов Д.А., Мамаев В.Я., Петров К.К. Электронный учебник оператора-навигатора (ETON). Зарегистрирован в Реестре программ для ЭВМ: Свидетельство № 2005612096 от 17.08.2005.
8. Вудс В.А. Сетевые грамматики для анализа естественного языка // Кибернетический сборник. М., 1976. Вып. 13. С. 120–158.
9. Кузьмук В.В. Применение модифицированных Е-сетей для построения параллельных алгоритмов // Докл. АН УССР: Сер. А. 1985. № 8. С. 186–203.

Санкт-Петербургский государственный университет аэрокосмического приборостроения (ГУАП)

Материал поступил в редакцию 20.12.2006.

TEST MODELS AND ALGORITHMS FOR KNOWLEDGE PRESENTED AS PETRI NETWORKS

V. Ya. Mamaev, D. A. Gorbunov

State University of Aerospace Instrumentation, Saint-Petersburg

Due to the development of computer engineering and data processing technology, it has become possible to create computerized training systems based on artificial intelligence. One of problems that can be solved in the computerized training systems by means of Petri networks is testing the knowledge base in case test tasks are of the open type and the form of the trainee answer is maximally free. The paper represents solutions of the problems of computerized natural language interpretation and assessment of the trainees' answers.