

угольника  $M$ , в результате чего находятся все порождаемые ими элементы покрытия, которые образуют множество  $C = \{C_1, C_2, \dots, C_p\}$ ;

2) из множества  $C$  удаляются все те элементы, которые поглощаются какими-либо другими элементами того же множества.

После этого все оставшиеся элементы множества  $C$  заменяются представляющими их конъюнкциями соответствующих предикатных переменных. Тогда дизъюнкция этих конъюнкций и даст искомую булеву формулу многоугольника в ДНФ.

Изложенный метод проиллюстрируем на примере многоугольника, изображенного на рисунке 6.

Если обход границы вести в прямом направлении, то сторона  $a$ , как было показано выше, порождает элемент покрытия  $C_1$  с границей  $b^*c^*f^*k^*$ . Тот же самый элемент покрытия порождает стороны  $b$ ,  $c$ ,  $e$ ,  $f$ ,  $i$ ,  $j$ . Сторона  $d$  порождает элемент покрытия  $C_3$  с границей  $d^*e^*f^*$ , а сторона  $g$  – элемент покрытия  $C_4$  с границей  $g^*j^*k^*$ . Сторона  $h$  порождает элемент покрытия  $C_6$  с границей  $h^*i^*j^*k^*$ , а сторона  $l$  – элемент покрытия  $C_2$  с границей  $l^*abc^*$ .

Если обход границы вести в обратном направлении, то сторона  $a$  порождает элемент покрытия  $C_2$  с границей  $al^*cb^*$ , а сторона  $b$  – тот же самый элемент покрытия с границей  $bal^*c^*$ . Стороны  $c$  и  $l$  порождают элемент покрытия  $C_2$  с границами  $cbal^*$  и  $l^*cba^*$  соответственно, а стороны  $d$  и  $f$  – элемент покрытия  $C_3$  с границами  $d^*fe^*$  и  $f^*ed^*$  соответственно. Сторона  $e$  порождает элемент покрытия  $C_3$  с границей  $ed^*f^*$ , а стороны  $i$  и  $j$  – элемент покрытия  $C_6$  (рис. 8) с границами  $ih^*kj^*$  и  $jih^*k^*$  соответственно. Сторона  $h$  порождает элемент покрытия  $C_6$  с границей  $h^*k^*ji^*$ , а сторона  $g$  – элемент покрытия  $C_5$  с границей  $g^*ba^*l^*$ .

Тем самым будет найдено множество  $C = \{C_1, C_2, \dots, C_6\}$ , содержащее все возможные элементы покрытия многоугольника  $M$ . Так как элемент  $C_5$  поглощается элементом  $C_2$ , то он удаляется из множества  $C$ .

В результате булева формула многоугольника в ДНФ будет иметь следующий вид:

$$F = bcfk \vee alcb \vee def \vee hijk \vee gjk.$$

**Заключение.** Предложен приемлемый на практике метод решения задачи построения булевой формулы многоугольника в дизъюнктивной

нормальной форме. Метод основан на последовательном формировании выпуклых компонент, которые представляют в искомой формуле отдельные элементарные конъюнкции и покрывают в совокупности все точки плоскости, принадлежащие исходному многоугольнику.

Достоинством метода является присутствие в полученной формуле только тех предикатных переменных, которые связаны с полуплоскостями, порождаемыми лишь сторонами многоугольника.

Необходимо отметить, что направление исследований, связанное с представлением многоугольников булевыми формулами, открывает новые возможности для решения широкого круга оптимизационных задач, например в области топологического проектирования интегральных схем, путем использования развитого аппарата булевой алгебры.

#### СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

- Шестаков, Е.А. Автоматизированная система подготовки информации для формирования фотошаблонов / Е.А. Шестаков, А.А. Бутов, Т.Л. Орлова, А.А. Воронов // Искусственный интеллект. – Украина: Донецк. – 2008. – № 4. – С. 200–207.
- Фейнберг, В.З. Геометрические задачи машинной графики больших интегральных схем – М.: Радио и связь, 1987. – 178 с.
- Ласло, М. Вычислительная геометрия и компьютерная графика на C++ / Пер. с англ. – М.: БИНОМ, 1997. – 304 с.
- Препарата, Ф. Вычислительная геометрия: введение / Ф. Препарата, М. Шеймос / Пер. с англ. – М.: Мир, 1989. – 478 с.
- Никулин, Е.А. Компьютерная геометрия и алгоритмы машинной графики – СПб.: БХВ-Петербург, 2005. – 576 с.
- Закревский, А.Д. Канонические булевы формулы многоугольников // Информатика. – 2009. – № 2. – С. 93–101.
- Поттосин, Ю.В. Использование булевых функций для представления многоугольников / Ю.В. Поттосин, Е.А. Шестаков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2008. – № 2(3). – С. 106–115.
- Бутов, А.А. Простой метод нахождения булевой формулы многоугольника в дизъюнктивной нормальной форме // Вестник Брестского государственного технического университета. Физика, математика, информатика. – 2011. – № 5. – С. 48–51.

Материал поступил в редакцию 06.12.12

#### BUTOV A.A. The method of finding a Boolean formula of a polygon in the disjunctive normal form, without the used of additional predicate

A feasible in practice method of the solution of the constructing a Boolean formula of a polygon in the disjunctive normal form is proposed. The method is based on a sequential formation of convex components, which represent in the desired formula individual elementary conjunctions and cover in aggregate all the points of the plane belonging to the original polygon.

The advantage of the method is the presence in the resulting formula only the predicate variables that are associated with the half-planes generated only by the sides of the polygon.

The method can be used, in particular, in the systems computer-aided design of integrated circuits topology.

УДК 004.514.62

Чеусов А.В.

### БАЗОВЫЙ ЛИНГВИСТИЧЕСКИЙ ПРОЦЕССОР PROSAO И ЕГО ОСНОВНЫЕ ПРИЛОЖЕНИЯ

Современные информационные технологии еще более обострили актуальность автоматической обработки текстовых документов. К традиционным задачам информационного поиска, машинного перевода, классического реферирования добавились задачи автоматической классификации текстовых документов, автоматизации инженерии знаний, интерфейса на естественных языках (ЕЯ) и cross-language функциональности, пользовательски-ориентированного реферирования, прагматического анализа (sentiment analysis или opinion mining) и т.д. Используемые при их решении средства лингвистической обработки, как правило, являются проблемно-зависимыми и, в силу этого, зачастую сильно ограниченными по своей функциональности и недостаточно проработанными с точки зрения качественных показателей, что, в целом, не соответствует их

эффективному использованию и дальнейшему развитию.

Учитывая, что ряд процедур лингвистической обработки текста, а именно его анализа, являются общими для многих приложений, нами сформулирована концепция базового лингвистического процессора (БЛП) [1] и определена его функциональность, которая, в отличие от существующих ЛП, является универсальной по отношению к различным ЕЯ и задачам их обработки, включает лексический, лексико-грамматический, синтаксический и семанτικο-синтаксический анализ текста и обеспечивает построение его формальной, семантико-синтаксической структуры, ориентированной на распознавание в тексте знаний основных типов, и, в силу этого,

Чеусов Алексей Владимирович, ведущий программист ИП «Инвеншион Машин».

Физика, математика, информатика

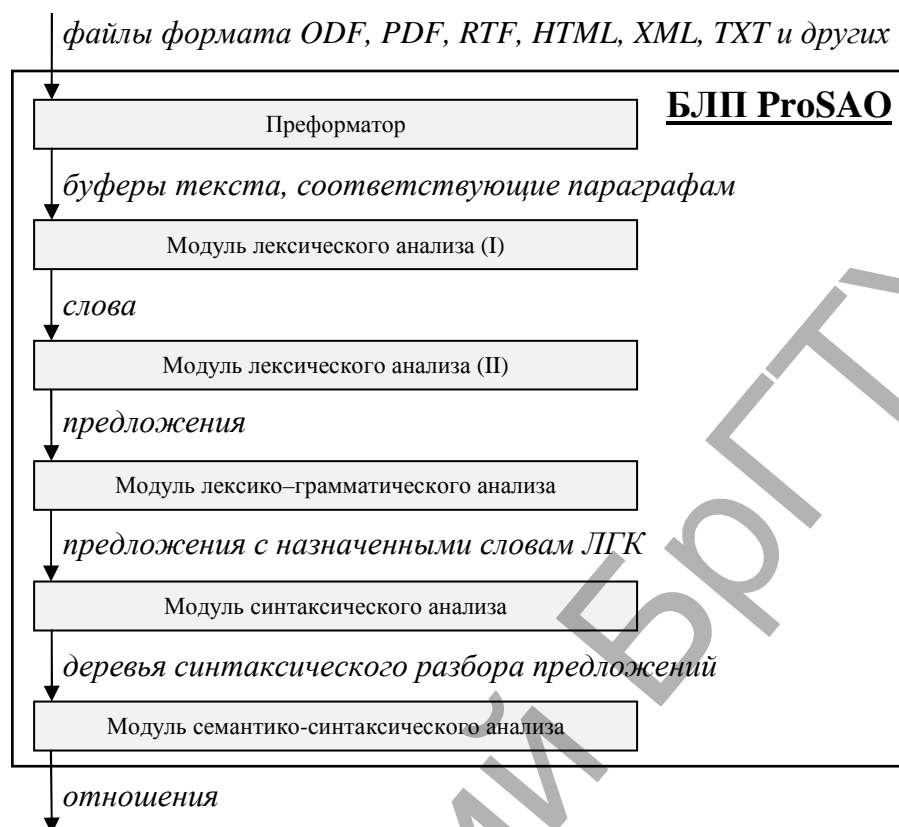


Рис. 1. Структурно-функциональная схема БЛП ProSAO

составляет, в целом, основу реализации эффективных информационных систем. Определены основные типы лингвистических ресурсов лингвистической базы знаний (ЛБЗ) БЛП. Поскольку значительную их часть составляют знания о ЕЯ, формируемые экспертами в виде лингвистических правил (ЛПР), особо значимым при решении задачи разработки эффективного БЛП становится построение для формального описания ЛПР некоторой нотации, максимально соответствующей с требованием ее доступности для использования экспертами, возможностью обобщения разрабатываемых правил и оптимизации скорости их обработки на всех этапах лингвистического анализа (ЛА) текста. Последнее важно с той точки зрения, что БЛП является только определенной частью любого конкретного приложения и, в силу требований, предъявляемых пользователями к промышленным системам, ему доступен сравнительно небольшой ресурс времени для требуемой лингвистической обработки текста. В целях такой нотации разработан язык расширенных регулярных выражений WRE [2]. Построен эффективный алгоритм сопоставления входной цепочки обрабатываемого текста с множеством описанных в ЛБЗ на языке WRE лингвистических правил [3]. Сформулированы основные принципы, которые должны быть положены в основу разработки БЛП, включая ориентацию на обработку произвольных текстов ЕЯ, их многоязычность, независимость ЛБЗ от процедур обработки текста, модульность ЛА текста, обеспечение инструментальных средств качественного тестирования всех компонентов ЛБЗ и алгоритмов ЛА. Полученные теоретические результаты легли в основу построенного нами промышленного многоязычного БЛП ProSAO.

Разработана структурно-функциональная схема БЛП ProSAO (рис. 1). Первоначально текст практически в любом из известных форматов (odf, pdf, rtf, html, doc, txt и др.) поступает на вход БЛП, который выполняет его форматирование и далее — лексический, лексико-грамматический, синтаксический и семантико-синтаксический анализ текста. Результаты всех этих этапов его обработки фактически образуют лингвистический индекс текста. Представленная функциональность ProSAO реализуется соответствующим программным обеспечением БЛП, который имеет в своем со-

ставе еще один важный компонент — модуль технологического обеспечения, представляющий собой совокупность инструментальных средств для разработки как ЛБЗ, так и собственно программного комплекса ProSAO, а также средств его тестирования.

Что касается ЛБЗ, то лексико-грамматический классификатор в нашем случае включает, например, для английского языка, 106 лексико-грамматических классов (ЛГК), базовый словарь (БС) включает порядка 190000 словоформ, синтаксический классификатор содержит порядка 320 классов, семантико-синтаксический — 4 основных и 312 дополнительных классов. Базовый корпус текстов (БКТ) включает более 1300000 словоупотреблений (более 80000 предложений). В состав ЛБЗ ProSAO входят также разработанные экспертами ЛПР лексического анализа текста (459 правил), лексико-грамматического анализа (около 2000 правил) и семантико-синтаксического анализа (около 3000 правил). Учитывая актуальность задачи автоматического распознавания ЛГК слов, отсутствующих в БС, в составе БЛП реализован специальный модуль морфологического анализа, использующий 105 словообразующих префиксов, 49000 постфиксов и 84 ЛПР. Также, в состав ЛБЗ входит множество правил (около 900) преобразования слов в пределах одной парадигмы. Поскольку на этапе лексико-грамматического анализа текста в нашем случае используется также известный вероятностно-статистический алгоритм, основанный на модели Маркова, то в состав ЛБЗ ProSAO включены необходимые для этого статистические данные, полученные из БКТ.

В задачи преформатора входит преобразование входного текста в так называемый "простой текст" (plain text) и требуемую кодировку, выделение определенных частей текста, таких как заголовки, абзац и т.п. В нашем случае преформатор реализован с использованием уже существующих внешних библиотек, имеющихся в свободном доступе. Что касается последующих модулей БЛП ProSAO, то они в основном реализованы на основе указанной выше ЛБЗ, упомянутых вероятностно-статистического алгоритма лексико-грамматического анализа и алгоритма сопоставления входного текста с множеством описанных в ЛБЗ на языке WRE лингвистических правил. При этом, на этапе лексического анализа во входном тексте распознаются

границы слов и предложений, лексико-грамматического — ЛГК его слов. На этапе синтаксического анализа строится синтаксическое дерево каждого предложения входного текста, в котором его слова представлены в виде листьев, а другим узлам соответствуют, например, простая именная группа, предложная группа, глагольная группа и т.п., в том числе и собственно предложение (корень дерева). Грамматика синтаксического анализатора в нашем случае имеет вид  $G = (T, N, s, P)$ , где  $T$  — множество терминальных символов, т.е. множество пар (слово, ЛГК),  $N$  — множество нетерминальных символов,  $s \in N$  — стартовый символ грамматики,  $P = \{P_1, P_2, \dots, P_k\}$  — набор правил вывода. Каждое правило  $P_i$  имеет вид  $C_l C_r \rightarrow C_n C_r$ , где  $n \in N$ , а  $C_l, C_r$  и  $C$  — регулярные выражения WRE с алфавитом  $T \cup N$ .  $C_l$  и  $C_r$  задают левый и правый контексты для  $C$  соответственно. В целом грамматика  $G$  подобна контекстно-зависимой грамматике, но от последней ее отличают два существенных обстоятельства. Во-первых, здесь вместо комбинации символов из  $N$  и  $T$  в правилах вывода применяются WRE правила. Во-вторых, внесены дополнительные ограничения на процедуру сопоставления текстового входа левым частям правил из  $P$ , которые обусловлены такой особенностью грамматики, например, английского языка, как то, что он является в основном право-ветвящимся, то есть, атрибуты располагаются в предложении справа от главного слова. На этапе семантико-синтаксического анализа из синтаксических деревьев извлекаются следующие, наиболее значимые с точки зрения знаний основных типов, отношения.

Отношение **SimpleNounPhrase** включает единственный компонент **NounPhrase**, обозначающий простую именную группу, состоящую из так называемого лингвистического главного слова, представленного существительным, личным местоимением или числительным, а также его определений и других слов, например, прилагательных, артиклей, наречий и т.д.

Таблица 1. Отношение **VerbPhrase** включает 14 компонентов

Название компонента	Пример
LeftObject	<b>A planar carbon segment commutator</b> has contact members
Action	A planar carbon segment commutator <b>has</b> contact members.
Adjective	This solution is <b>excellent</b>
RightObject	A planar carbon segment commutator has <b>contact members</b>
Preposition	The contact members are embedded <b>in</b> the carbon segments
IndirectObject	The contact members are embedded <b>in the carbon segments</b>
Adverbial	She runs <b>quickly</b>
LeftObject2	(не заполняется)
Action2	The output buffer can drive a load with different drive levels <b>by changing</b> the transconductance
Adjective2	
RightObject2	The microprocessor is also programmed to request the signal
Preposition2	Temporary liquid holding cavities are formed so as to extend outwardly <b>from</b> the underside
IndirectObject2	Temporary liquid holding cavities are formed so as to extend outwardly <b>from the underside</b>
Adverbial2	Temporary liquid holding cavities are formed so as to extend <b>outwardly</b> from the underside

Структура отношения **VerbPhrase**, выражающих предикацию главных членов предложения, т.е. подлежащего и сказуемого. Единственным обязательным к заполнению компонентом отношения **VerbPhrase** является компонент **Action**. Компоненты **Action2**, **Adjective2**, ..., **Adverbial2** заполняются лишь в случае, когда подлежащее выражается глагольной группой. Указанное глагольное дополнение и все его зависимые компоненты сохраняются именно в

этой части отношения. Компонент **LeftObject2** в отношении **VerbPhrase** не заполняется, т.к. подлежащее глагольного дополнения всегда отсутствует. Например, из предложения "Temporary liquid holding cavities are formed so as to extend outwardly from the underside" извлекается следующее отношение **VerbPhrase** (приводятся только непустые компоненты):

Название компонента	Пример
LeftObject	Temporary liquid holding cavities
Action	are formed
Action2	so as to extend
Preposition2	from
IndirectObject2	the underside
Adverbial2	outwardly

Отношение **NounPhrase\_additional** описывает именную группу, распространенную предложными, причастными, деепричастными, инфинитивными либо адъективными оборотами, а также предложениями. В определенной степени это отношение близко к отношению **VerbPhrase**. Но глагольный оборот в **NounPhrase\_additional** не может быть представлен в виде глагола в личной форме. Например, из предложения "Information about the films is obtained from newspaper." извлекается следующее синтаксическое отношение типа **NounPhrase\_additional**:

Название компонента	Пример
LeftObject	Information
Preposition	About
RightObject	the films

Отношение **ComplexSentence** идентифицирует сложноподчиненное предложение. Так же как и в случае **VerbPhrase**, это отношение состоит из двух частей, причем для первой и второй семерки компонентов добавлены компоненты **Determiner** и **Determiner2**, заполняемые подчинительными союзами, относящимися к той или иной семерке. Например, из предложения, "If you are going to stay, then I shall stay with you." извлекается отношение **ComplexSentence** следующего вида:

Название компонента	Пример
Determiner	If
LeftObject	you
Action	are going to stay
Determiner2	then
LeftObject2	I
Action2	shall stay
Preposition2	with
IndirectObject2	you

Перечисленные отношения охватывают фактически все информативные единицы входного текста и действительно определяют распознавание основных типов знаний. Так, отношение **SimpleNounPhrase** соответствует объектам, отношение **VerbPhrase** — фактам, а **ComplexSentence** — причинно-следственным отношениям. В зависимости от наполнения **NounPhrase** также может содержать факты или объекты и их атрибуты.

БЛП ProSAO является многоязычным и успешно реализован, наряду с английским языком, для французского, японского, немецкого и китайского языков. Это потребовало разработки для каждого из них своей ЛБЗ. Что касается алгоритмического и программного обеспечения, то здесь незначительные изменения коснулись лексического и лексико-грамматического анализа текста, и они связаны с такими особенностями, как отсутствие пробелов в качестве разделителей между словами для иероглифических языков, опущение диакритических знаков во французском языке, высокая распространенность словосложения в немецком языке и другими. В силу использования полученных нами концептуальных, алгоритмических и технологических решений скорость лингвистического анализа входного текста на сопоставимых

этапах в среднем значительно превосходят показатели таких известных систем, как, например, Connexor [4] (скорость синтаксического анализа в нашем случае выше приблизительно в 6 раз при сопоставимом качестве). Выходом БЛП ProSAO является лингвистический индекс текста. Важно, что он отражает различные уровни глубины его анализа — от лексического до семантико-синтаксического. Эти возможности БЛП успешно использованы, прежде всего, при реализации основной функциональности известной системы автоматизации инженерии и управления знаниями Goldfire[5], используемой для решения инновационных задач многими крупнейшими компаниями мира: распознавание в текстовых документах знаний основных типов, семантическое тегирование текстовых документов и ЕЯ-запросов пользователя в целях вопросно-ответной функциональности с ЕЯ-интерфейсом пользователя и для распознавания атрибутивных знаний и субъективной информации, получение интегрированной оценки информативности отдельных фрагментов текстовых документов на этапе их автоматического реферирования.

#### СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Чеусов, А.В. Базовые лингвистические процессоры: назначение, принципы построения, функциональность, состав, приложения / А.В. Чеусов, И.В. Совпель // Актуальные проблемы теоретической и прикладной лингвистики: материалы Международной конференции – МГЛУ, 2010. – Т. 2. – С. 122–126.
2. Чеусов, А.В. Расширенные регулярные выражения в системах обработки естественного языка / А.В. Чеусов // Информационные системы и технологии (IST'2002): материалы I Международной конференции. – БГУ, 2002. – Т. 3. – С. 21–24.
3. Чеусов, А.В. Принципиальная схема алгоритма сопоставления текстового входа для задачи его автоматического лингвистического анализа / А.В. Чеусов // Информатизация образования. – 2012. – С. 74–85.
4. Домашняя страница компании Connexor. – Режим доступа: <http://www.connexor.com>.
5. Домашняя страница компании Invention Machine. – Режим доступа: <http://www.invention-machine.com>.

Материал поступил в редакцию 18.12.12

#### CHEUSOV A.V. Linguistic processor ProSAO and its applications

In this paper a linguistic processor ProSAO is described. ProSAO is the core of the knowledge engineering and management system Goldfire developed by Invention Machine (IHS). Also described are basic design principles, linguistic knowledge base characteristics, applications of ProSAO and format of its output.

УДК 656.13.05

Михневич В.А., Шуть В.Н.

## РЕГУЛИРОВАНИЕ ГОРОДСКОГО ПЕРЕКРЕСТКА НА ОСНОВЕ МНОГОАГЕНТНОГО ПОДХОДА

**Введение.** Светофоры, как средство управления транспортным движением на перекрестках, используются на протяжении нескольких десятков лет без принципиальных изменений, в то время как растущее количество АТС (автономное автотранспортное средство) создает дорожные заторы на перекрестках. Современные автомобили могут быть оснащены всем необходимым оборудованием для автономного движения и планирования пути на основе систем глобального позиционирования. При помощи беспроводного взаимодействия и систем позиционирования современные АТС способны образовывать многоагентную систему, что приводит к более эффективному дорожному движению.

С развитием автономного движения, т.е. без участия человека-водителя, к существующим видам перекрестков (регулируемый и нерегулируемый) добавится регулируемый интеллектуальный. Интеллектуальный перекресток обладает более широкими возможностями, чем просто регулирование движения в сегодняшнем его виде.

Переход со временем на автономные транспортные средства требует кардинального пересмотра принципов управления транспортными потоками в улично-дорожной сети (УДС) города. Светофорный объект будет не только отмерять временные промежутки фаз зеленого и красного сигналов пересекающихся дорог перекрестка, но также планировать оптимальный разъезд, на основе информации о подъезжающих с различных направлений транспортных потоках.

### 1. Постановка задачи пересечения перекрестка

Для загруженных автомагистралей строят дорогостоящие развязки «клеверного» типа, однако большинство современных перекрестков являются одноуровневыми и управляются светофорами. Именно на перекрестках наблюдается наибольшее скопление транспорта. Очереди из автомобильных транспортных средств (АТС) на

перекрестках также пагубно влияют на пропускную способность потока в целом. Из-за очередей АТС приходится разгоняться и тормозить, что приводит к потере времени, а также к дополнительному расходу топлива. Кроме того, согласно статистике 25%–45% всех столкновений АТС приходится при пересечении перекрестков [1]. Поэтому считается, что перекресток является самым слабым звеном в городской дорожно-транспортной системе.

### 2. Описание и взаимодействие агентов в многоагентной системе

Предлагаемая многоагентная система (МАС) как модель состоит из двух видов агентов: агент-менеджер и агент-водитель. Каждая система обслуживает один перекресток, при этом перекресток интерпретируется как агент-менеджер (далее «М»), а АТС — как агент-водитель (далее «В»). МАС является централизованной, в которой «М» управляет агентами «В» посредством сообщений. Коммуникация между агентами может быть 3 типов: «В-М», «М-М», «В-В».

*Взаимодействия типа «В-М» (рис. 1).* Протокол сообщений агентов вида «В-М» описан в отдельном пункте. В общих чертах процедура пересечения перекрестка АТС выглядит следующим образом: «В», подъезжая к перекрестку, высылает сообщение «М» о пересечении перекрестка, передавая необходимую информацию (местоположение, время прибытия на перекресток, скорость движения, полоса прибытия на перекресток, полоса назначения). Способ передачи не уточняется, т.к. это не принципиально важно и может быть реализовано как через беспроводную сеть, так и через интернет. «М» необходимо дать положительный (с информацией о пересечении) либо отрицательный (с причиной отказа) ответ «В». «М» работает как сервер в режиме реального времени. Поэтому время на принятие решения и отправление ответа ограничено. Принятое решение может зависеть от состояния перекрестка (количества АТС в

Михневич В.А., магистрант кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Шуть В.Н., к.т.н., доцент кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Беларусь, БрГТУ, 224017, г. Брест, ул. Московская, 267.