

Зыков В. И., Антонов С. В.

## АЛГОРИТМ ВОССТАНОВЛЕНИЯ ОДНОЗНАЧНОСТИ СМЫСЛОВОГО ПОНЯТИЯ В СООБЩЕНИИ-112, НАБРАННОМ С ОШИБКАМИ

В статье исследуется необходимость обработки СМС-сообщений о пожаре или чрезвычайной ситуации. Сообщение в диспетчерскую службу передаётся очевидцем, находящимся в состоянии стресса. Ошибки идентичных кнопочных нажатий словаря Т9 и ошибки возможных случайных нажатий соседних букв искажают сообщение до неузнаваемости. На основе предложенного алгоритма проводится восстановление однозначности смыслового понятия.

*Ключевые слова:* Сообщение-112, система приёма экстренных сообщений, Система-112, алгоритм обработки сообщений.

Сообщение о пожаре или чрезвычайной ситуации в диспетчерскую службу, как правило, поступает от очевидца события. Окружающая обстановка способствует тому, что очевидец находится в состоянии стресса. Кроме того, следует учитывать факт, что сообщение может быть передано лицом с ограниченными возможностями.

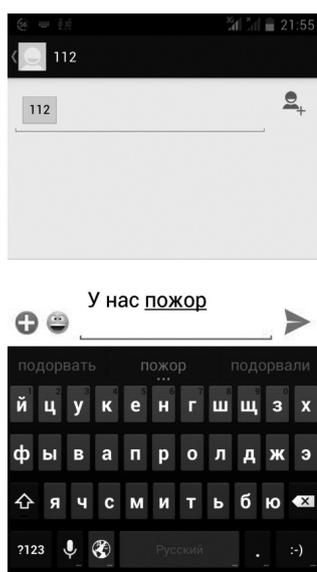
Основная цель модуля Сообщение-112 в Системе-112 – выявление однозначности

смыслового понятия из сообщения, набранного с ошибками [1, 2].

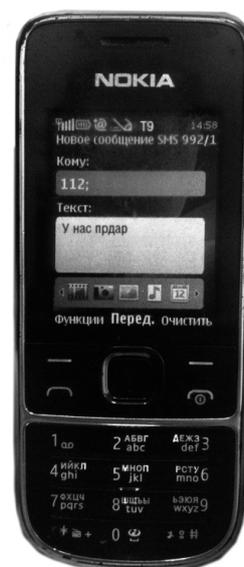
В устройствах, позволяющих осуществлять набор текстовых сообщений, встроены два вида словарей (рис. 1). Первый словарь – орфографический, основная задача которого – контроль соответствия передаваемых сообщений правилам русского языка.

Второй словарь – словарь Т9 (предиктивный ввод текста) – система ускоренного ввода текста в цифровое устройство. Данное программное обеспечение в процессе набора сообщения предлагает варианты окончания слов и фраз, основываясь на имеющихся в его словаре, а также исправляет распространённые ошибки.

Разницу между процессами набора сообщений при использовании двух типов словарей можно продемонстрировать на примере следующего идеального сообщения: «У нас пожар. Скорее приезжайте. Кругом огонь. Ивана засыпало обломками. Присылайте МЧС».



а



б

Рисунок 1. Виды словарей в устройствах сотовой связи:

а – орфографический словарь; б – словарь Т9

При наборе с использованием орфографического словаря сообщение будет выглядеть также, либо будет содержать незначительные ошибки: «У нас **пожор**. **Скарее** приезжайте. **Кругом агонь**. **Ивана засыпало абломками**. **Присылайте МЧС**».

Диспетчер Системы-112 даже при наличии подобных ошибок сможет понять, какие конкретные дежурно-диспетчерские службы (ДДС) следует выслать в данной ситуации. Результаты проведённых экспериментальных исследований показали, что в случае набора сообщения с помощью словаря Т9, оно может выглядеть следующим образом: «У нас **мозгу**. **Синтез** приезжайте. **Кругом обоня**. **Ивана засыпало накопивни**. **Присылайте ОХР**» [3, 4].

Понять, о чём пишет потерпевший, невозможно. Следовательно, диспетчер Системы-112 не сможет правильно отреагировать на сообщении о ЧС и удалит его, посчитав спамом.

Эксперименты показали, что в случае быстрого набора сообщения с помощью экранной клавиатуры, человек может промахнуться и случайно нажать соседнюю с нужной букву (рис. 2). Слово, набранное с двумя и более ошибками от нажатия соседних букв, не исправляется орфографическим словарём и будет отличаться от правильно набранного.

Сообщение с возможными случайными нажатиями может выглядеть следующим образом: «Ц **евс прдар**. **Скорее ппикцзайие**. **Кругом ннннь**. **Ивана засыпвло облрсуми**. **Прмчюпйте мчс**».

Такое сообщение также невозможно понять, как и в случае с набором через Т9. Получив такое сообщение, диспетчер Системы-112 не сможет правильно на него отреагировать и удалит, посчитав спамом.

Данная проблема может быть решена, если Система-112 будет оборудована модулем обработки Сообщения-112. Диспетчер



Рисунок 2. Возможные ошибочные нажатия при наборе буквы «а»

Системы-112 увидит оригинальное сообщение, которое отправил потерпевший, и его обработанную версию: «Ц **евс прдар** (возможно: **пожар**). **Скорее ппикцзайие** (возможно: **приезжайте**). **Кругом рнонт** (возможно: **огонь**). **Ивана щачыпвло** (возможно: **засыпало**) **облрсуми** (возможно: **обломки**). **Прмчюпйте мчс**».

Для примера возьмём два сообщения, первое из которых содержит ошибки идентичных кнопочных нажатий словаря Т9, второе – ошибки возможных случайных нажатий соседних букв, которые представлялись ранее:

1) «У нас **мозгу**. **Синтез** приезжайте. **Кругом обоня**. **Ивана засыпало накопивни**. **Присылайте ОХР**»;

2) «Ц **евс прдар**. **Скорее ппикцзайие**. **Кругом ннннь**. **Ивана засыпвло облрсуми**. **Прмчюпйте мчс**».

Рассмотрим работу модуля обработки Сообщения-112 в Системе-112 на примере первого сообщения. После проверки минимальной информации [1] сообщение попадает в блок «Выделения ключевых слов» (рис. 3).

Для обработки текстовых сообщений применяется алгоритм, который предусматривает четыре внешних цикла. В **первом цикле** проводится поиск ключевых слов. Для этого в сообщении поочерёдным перебором анализируется каждое слово (3 и более символов) и проводится сравнение его с ключевыми словами базы данных. Алгоритм обработки текстовых сообщений представлен на рисунке 4.

При совпадении слова из сообщения с ключевым словом формируется матрица  $R_i$  и проверяется следующее слово поступившего сообщения. Матрица  $R_i$  принимает вид:



Рисунок 3. Пример первого входящего сообщения в Систему-112

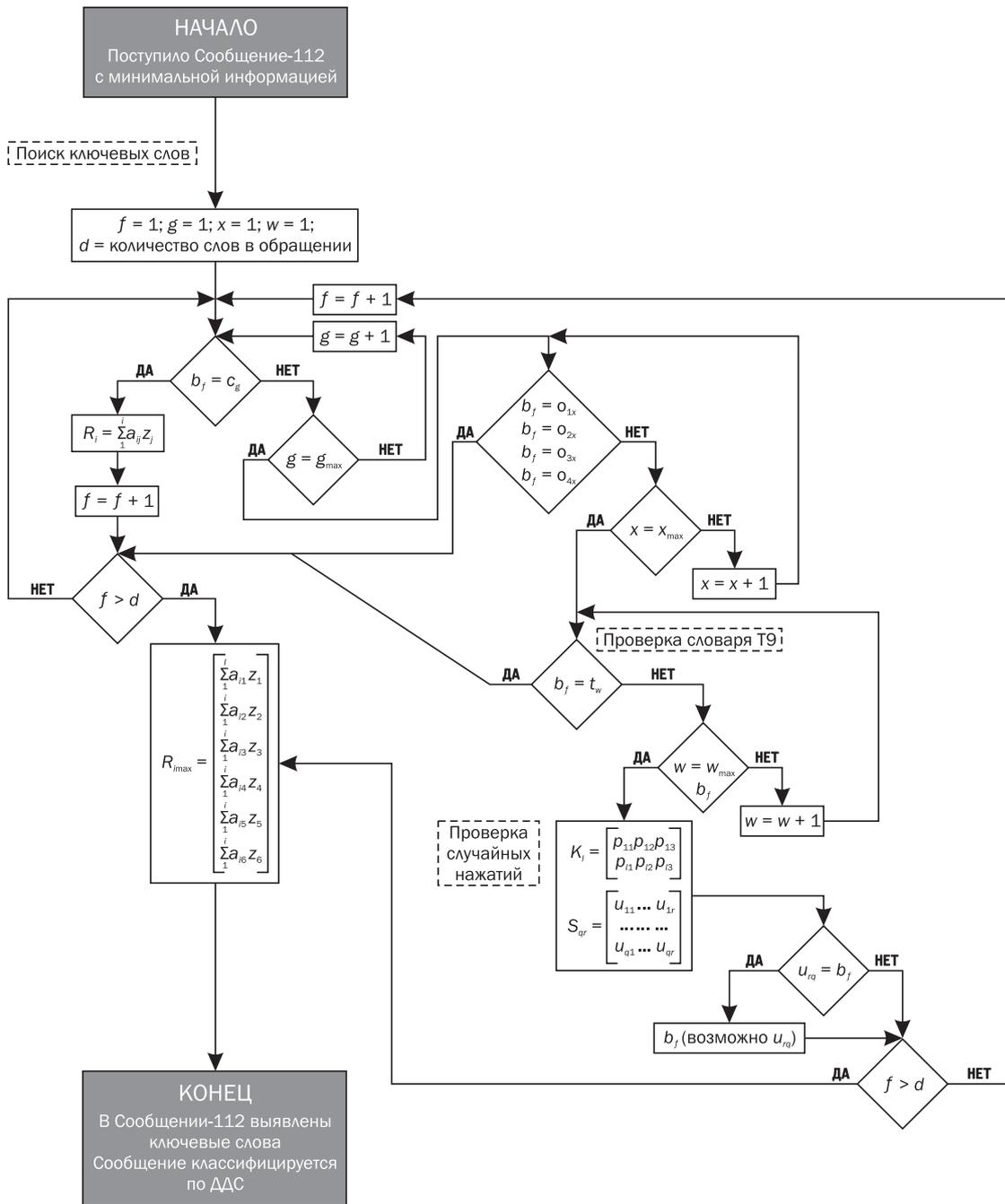


Рисунок 4. Алгоритм выявления ключевых слов в сообщении:

$a_{ij}$  – баллы ключевого слова [ $i$  – строка,  $j$  – столбец];  $z_i$  – ранг важности ДДС;  $b_f$  – слово из сообщения [ $f$  – строка];  
 $c_g$  – ключевое слово из базы данных ключевых слов [ $g$  – номер ключевого слова в базе данных];  
 $g_{max}$  – количество ключевых слов в базе данных;  $d$  – количество слов в сообщении;  $o_{px}$  – варианты ключевого слова при проверке орфографии [ $n$  – вариант проверки орфографии,  $x$  – строка];  $x_{max}$  – все варианты проверки орфографии;  
 $y$  – счётчик проверки орфографии;  $t_w$  – варианты ключевого слова при переборе вариантов словаря Т9  
 $[w$  – номер варианта ключевого слова из базы данных];  $w_{max}$  – все варианты ключевого слова из базы словаря Т9

$$R_i = \sum_1^i a_{ij} \cdot z_j,$$

где  $a_{ij}$  – баллы ключевого слова [ $i$  – строка,  $j$  – столбец];  $z_j$  – ранг важности ДДС.

При несовпадении слова из поступившего сообщения с ключевым словом включается **второй цикл** работы алгоритма, который предусматривает проверку орфографии слов в сообщении. Принцип замены несовпадающих

слов из сообщения на базовые слова отражён в *таблице 1*.

При несовпадении слова из сообщения с ключевым словом во втором цикле проверки орфографии включается **третий цикл** работы алгоритма, который предусматривает проверку слов из сообщения на идентичные кнопоч-

Таблица 1

Проверка орфографии слова из сообщения с помощью различных методов

Слово из сообщения	Действия и результат проверки слова		Ключевое слово в базе данных
Мозгу	Замена	аожар	Пожар
		божар	
		...	
	Удаление	пожая	
		_ожар	
		п_жар	
		...	
	Вставка	пожа_	
		апожар	
		бпожар	
		...	
	Перестановка	пожаря	
		опжар	
		пжоар	
		...	

Таблица 2

Проверка орфографии слова из сообщения на идентичные кнопочные соответствия словаря Т9

Слово из сообщения	Проверка идентичных кнопочных соответствий словаря Т9					Ключевое слово в базе данных
Мозгу	М	М	Д	А	Р	Пожар
	Н	Н	Е	Б	С	
	О	О	Ж	В	Т	
	П	П	З	Г	У	
	м	о	з	г	у	
	п	о	д	а	р	
	.	.	.	.	.	
	о	н	д	а	т	

ные соответствия словаря Т9. Процесс замены несовпадающих слов из сообщения на базовые слова отражён в *таблице 2*.

Выявлены ключевые слова «пожар», «огонь», «засыпало» и «МЧС». В алгоритме включается **четвертый внешний цикл**, вычисляющий главную и второстепенные ДДС, которые должны реагировать на полученное сообщение. В *таблице 3* приведён ранг каждой экстренной службы и балльная оценка каждой службы для данных ключевых слов.

Одно ключевое слово будем считать предпочтительнее другого, если указанные экспертами оценки одного ключевого слова выше оценок другого эксперта, с учётом ранга ДДС. В соответствии с этим, из матрицы  $R_i$  находим максимальное значение в столбце, а именно  $R_{i\max}$ :

$$R_{i\max} = \begin{bmatrix} \sum_1^i a_{i1} \cdot z_1 \\ \sum_1^i a_{i2} \cdot z_2 \\ \sum_1^i a_{i3} \cdot z_3 \\ \sum_1^i a_{i4} \cdot z_4 \\ \sum_1^i a_{i5} \cdot z_5 \\ \sum_1^i a_{i6} \cdot z_6 \end{bmatrix} \quad (1)$$

Порог отбрасывания шума случайных ответов экспертов позволяет выделить только основные ДДС. Применим 10-процентный рубеж для отсева шума (10 % – 139,06 баллов). Это позволит исключить ДДС, выбранные экспертами случайно. Работа алгоритма закончена.

В итоге определена главная служба, реагирующая на сообщение ДДС – пожарная охрана ( $R_{i\max} = 862,8$  балла), и второстепенные: служба скорой медицинской помощи (229 баллов) и служба реагирования в чрезвычайных ситуациях (203,4 балла). Сообщение пересылается в Систему-112 с выделением ДДС, для которых оно предназначено (*рис. 5*).

Рассмотрим работу подсистемы обработки текстовых сообщений в Системе-112

Таблица 3

Выбор главной и второстепенных ДДС

Ключевое слово	Дежурно-диспетчерская служба						Сумма баллов
	Пожарная охрана	Служба реагирования в чрезвычайных ситуациях	Полиция	Служба скорой медицинской помощи	Аварийная служба газовой сети	Служба «Антитеррор»	
	6	3	4	5	1	2	
Пожар	60,00	0,00	20,00	17,60	0,80	0,00	
Огонь	78,80	1,80	1,40	12,20	5,00	0,00	
Засыпало	5,00	66,00	1,00	16,00	0,00	0,00	
Сумма	143,80	67,80	89,60	229,00	5,80	0,00	
Сумма с учётом ранга $R_i$	862,80	203,40	89,60	229,00	5,80	0,00	

на примере второго сообщения: «Ц е в с прдар. Скорее ппикцзайие. Кругом рнонт. Ивана щачыпвл о блрсуми. Прмчыопйте СЯС».

После проверки минимальной информации [1] сообщение попадает в блок «Проверки орфографии» (табл. 1), но блок орфографии не может исправить слова с большим количеством ошибок.

Далее сообщение переходит в блок «Проверки словаря Т9», но ключевые слова не будут выявлены, так как содержатся ошибки возможных случайных нажатий соседних

букв. После блока «Проверки словаря Т9» сообщение проверяется в блоке «Проверки случайных нажатий». Здесь поочерёдным перебором анализируется каждое слово (3 и более символов) и создаётся матрица  $K_j$  согласно вариантам ошибок при наборе желаемой буквы (табл. 4).

На примере слова «прдар» создаём матрицу возможных вариантов случайных нажатий букв  $K_j$  (рис. 6). В матрицу заносятся «варианты ошибки», соответствующие набираемой букве из слова, и сама буква, так как она может быть правильной.

Матрица  $K_j$  принимает вид:

$$K_j = \begin{bmatrix} p_{11}p_{12}p_{13} \\ p_{21}p_{22}p_{23} \end{bmatrix},$$

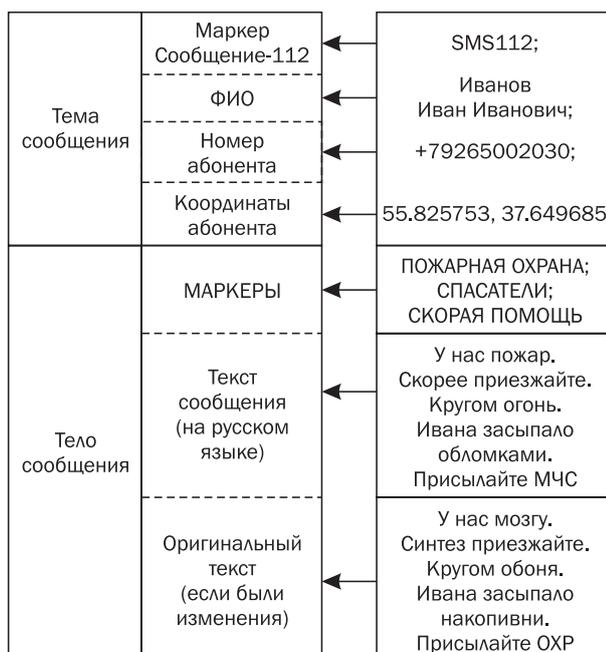


Рисунок 5. Первое сообщение, обработанное алгоритмом обработки и пересылки в ДДС

	п	р	д	а	р
	1	2	3	4	5
1	К	Е	Ш	У	Е
2	А	П	Л	В	П
3	С	М	Ь	Ч	М
4	М	И	Б	С	И
5	И	Т	Ю	М	Т
6	Р	О	Ж	П	О
7	Н	Г	З	Е	Г
8	Е	Н	Щ	К	Н
9	П	Р	Д	А	Р

Рисунок 6. Возможные варианты букв из слова «прдар»

Таблица 4

Варианты ошибок при наборе желаемой буквы

Набираемая буква	Варианты ошибки							
а	у	в	ч	с	м	п	е	к
б	л	ь	ю	ж	д			
в	ц	ы	я	ч	с	а	к	у
г	н	р	о	л	ш			
д	ш	л	ь	б	ю	ж	з	щ
е	к	а	п	р	н			
ж	щ	д	б	ю	э	х	з	
з	щ	д	ж	э	х			
и	п	м	т	о	р			
й	ф	ы	ц					
к	у	в	а	п	е			
л	г	о	т	ь	б	д	щ	ш
м	а	с	и	р	п			
н	е	п	р	о	г			
о	н	р	и	т	ь	л	ш	г
п	к	а	с	м	и	р	н	е
р	е	п	м	и	т	о	г	н
с	в	ч	м	п	а			
т	р	и	ь	л	о			
у	ц	ы	в	а	к			
ф	я	ы	ц	й				
х	з	ж	э					
ц	й	ф	ы	в	у			
ч	ы	я	с	а	в			
ш	г	о	л	д	щ			
щ	ш	л	д	ж	з			
ы	й	ф	я	ч	с	в	у	ц
ь	о	т	б	д	л			
э	з	ж	ю	х				
ю	д	б	э	ж				
я	ф	ч	в	ы				

где  $p_{i1}$  – буквы слова из сообщения [ $l$  – количество вариантов для каждой буквы]. На основе матрицы  $K_l$  создается матрица  $S_{qr}$  всех возможных комбинаций букв и сравнивается с ключевым словом (рис. 7).

При этом матрица  $S_{qr}$  принимает вид:

$$S_{qr} = \begin{bmatrix} u_{11} \dots u_{1r} \\ \dots \\ u_{q1} \dots u_{qr} \end{bmatrix},$$

где  $q$  – длина возможных вариантов для каждой буквы;  $r$  – количество вариантов для первой буквы;  $u$  – варианты набора трёх букв с учётом случайных нажатий.

На примере слова «прдар» у буквы  $K$  будет 6561 вариант ( $1 \cdot 9 \cdot 9 \cdot 9 \cdot 9 = 6561$ ). Всего в матрице  $S_{qr}$  будет 59049 вариантов ( $9 \cdot 9 \cdot 9 \cdot 9 \cdot 9 = 59049$ ), которые сравниваются со всеми ключевыми словами из базы данных.

После сравнения всех 59049 вариантов с ключевыми словами выявлено сходство с ключевым словом «пожар». Учитывая, что алгоритм мог работать и с правильно набранным словом, слово из сообщения не заменяется на ключевое, а дополняется найденным вариантом: прдар (возможно: пожар).

Как и в предыдущем случае, выявлены ключевые слова «пожар», «огонь», «засыпало» и «МЧС». Рассчитывается максимальное значение  $R_{i\max}$  (1), вычисляются главная и второстепенные ДДС, которые должны реагировать на полученное сообщение.

С помощью ранее приведённого алгоритма во втором сообщении определена главная дежурно-диспетчерская служба, реагирующая на сообщение – пожарная охрана ( $R_{i\max} = 862,8$  балла), и второстепенные: служба скорой медицинской помощи

	1	2	3	4	5	6	7	8	9
1	КЕШУЕ	АЕШУЕ	СЕШУЕ	МЕШУЕ	ИЕШУЕ	РЕШУЕ	НЕШУЕ	ЕЕШУЕ	ПЕШУЕ
2	КЕШУП	АЕШУП	СЕШУП	МЕШУП	ИЕШУП	РЕШУП	НЕШУП	ЕЕШУП	ПЕШУП
	...	...	...	...	...	...	...	...	...
$q$	КРДАР	АРДАР	СРДАР	МРДАР	ИРДАР	РРДАР	НРДАР	ЕРДАР	ПРДАР

Рисунок 7. Варианты набора трёх букв с учётом случайных нажатий

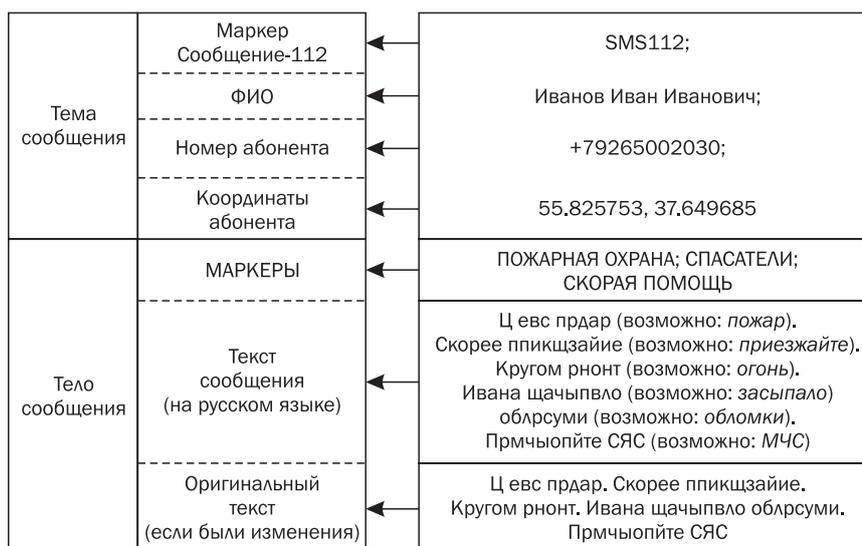


Рисунок 8. Второе сообщение, обработанное алгоритмом обработки и пересылки в ДДС

(229 баллов) и служба реагирования в чрезвычайных ситуациях (203,4 балла).

После обработки исправленное сообщение передаётся в выбранные ДДС (рис. 8).

Таким образом, внедрение модуля обработки Сообщения-112 в Систему-112 по-

зволит лицам, оказавшимся в чрезвычайной ситуации и пребывающим в стрессовом состоянии, а также людям с ограниченными возможностями сообщать о пожарах или ЧС любым доступным способом, в том числе с помощью СМС-сообщений.

#### ЛИТЕРАТУРА

1. Антонов С. В., Зыков В. И. Подсистема обработки текстовых сообщений в Системе-112 // Пожары и чрезвычайные ситуации: предотвращение, ликвидация. – 2013. – № 2. – С. 30–33.
2. Антонов С. В. Определение ключевых слов для дежурно-диспетчерских служб Системы-112 // Пожары и чрезвычайные ситуации: предотвращение, ликвидация. – 2014. – № 2. – С. 29–34.
3. Антонов С. В., Зыков В. И. Обработка текстовых сообщений в системе вызова экстренных служб по номеру «112» //

Мат-лы Междунар. науч.-практ. конф. молодых учёных и специалистов «Проблемы техносферной безопасности – 2012». – М.: Академия ГПС МЧС России, 2012. – С. 130–132.

4. Зыков В. И., Иванников А. П., Левчук М. С. Система радиоканального мониторинга комплексной безопасности объектов в составе ЦУКС // Пожарная безопасность в строительстве. – 2011. – № 3. – С. 24–30.

Zykov V., Antonov S.

## THE ALGORITHM OF RESTORING THE UNAMBIGUITY OF SEMANTIC CONCEPT IN MESSAGE-112 TEXTED INCORRECTLY

## ABSTRACT

**Purpose.** The article examines the necessity of processing text-messages about a fire or an emergency. As a rule, such messages are sent to the dispatcher service by a witness who is under stress. One cannot exclude the possibility of texting a message by a disabled person. Errors in pressing identical buttons of T9 vocabulary and errors in pressing adjacent buttons by mistake can transform the message beyond recognition. Restoring the unambiguity of semantic concept is based on the proposed algorithm.

**Methods.** The research study is based on processing the data received by means of experiments. The algorithm of messages processing, that includes four cycles, performs error correction within T9 vocabulary and eliminates the possibility of pressing buttons by mistake while texting a message.

**Findings.** When installing the designed module of processing Message-112 in System-112, the operator-on duty of System-112 will be able to see the processed version of the original message sent by

the injured person. According to keywords found in the message, the main and secondary duty-dispatching services that must respond to the received message are identified.

**Research application field.** The findings are recommended to be included in the research and development work in the process of designing System-112.

**Conclusions.** Thus, installing the module of processing Message-112 in System-112 will allow people caught up in emergencies and getting under stress as well as disabled people to signal about fires or emergencies by any means available, including via text-messages.

*Key words:* Message-112, the system of receiving emergency messages, the System-112, the algorithm of messages processing.

## REFERENCES

1. Antonov S.V., Zykov V.I. Subsystem of processing text messages within the System-112. *Pozhary i chrezvychainye situatsii: predotvrashchenie, likvidatsiia*. 2013, no. 2, pp. 30–33. (in Russ.).
2. Antonov S.V. Finding keywords for duty dispatcher services of 112 System. *Pozhary i chrezvychainye situatsii: predotvrashchenie, likvidatsiia*. 2014, no. 2, pp. 29–34. (in Russ.).
3. Antonov S.V., Zykov V.I. Processing text messages in a system of emergency call number “112”. *Mat-ly Mezhdunar. nauch.-prakt. konf. molodykh uchenykh i spetsialistov «Problemy tekhnosfernoi bezopasnosti – 2012»*. [Proc. of international sci.-pract. conf. “Problems of technosphere safety – 2012”]. Moscow, State Fire Academy of EMERCOM of Russia Publ., 2012. pp. 130–132.
4. Zykov V.I., Ivannikov A.P., Levchuk M.S. System radio channel monitoring of safety of objects in structure CDCS. *Pozharnaia bezopasnost' v stroitel'stve*. 2011, no. 3, pp. 24–30. (in Russ.).

VLADIMIR ZYKOV	Doctor of Technical Sciences, Professor State Fire Academy of EMERCOM of Russia, Moscow, Russia
SERGEI ANTONOV	State Fire Academy of EMERCOM of Russia, Moscow, Russia