

## СРЕДСТВА АВТОМАТИЗАЦИИ ИССЛЕДОВАНИЯ РУКОПИСНЫХ ДОКУМЕНТОВ ДЛЯ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

### Введение

Использование средств автоматизации экспертной работы, на наш взгляд, не только повышает производительность труда эксперта, но и в целом повышает объективность оценки того или иного комплекса признаков почерка [1, 2, 3, 4]. Один из основных источников данных для проведения экспертиз, решения идентификационных задач, задач розыска и поиска, а также для обеспечения *информационной безопасности* — это биометрическая информация [5]. Технические биометрические системы [1], основанные на рукописных объектах, можно использовать не только для непосредственной идентификации личности по почерку, но и для определения некоторых ее специфических характеристик, таких как возраст, пол, рост и т. п. признаки. Для решения задач информационной безопасности, розыска и поиска преступников возможно использование результатов диагностических почерковедческих исследований рукописных документов, в том числе и в электронном виде. Такими документами могут быть, например, записки с предупреждением о готовящемся террористическом акте или записки с требованием выкупа [5]. В США работники ФБР проводят исследования записок, написанных грабителями банков [6]. Большое количество и разнородность признаков почерка, характеризующих личность исполнителя рукописи, требуют создания специализированных автоматизированных рабочих мест (АРМ) в составе *автоматизированных средств обеспечения информационной безопасности (АСОИБ)*.

В настоящее время различными группами ученых выполняются исследования, направленные на идентификацию пользователя ЭВМ по его биометрическим параметрам [1]. Разработаны алгоритмы идентификации пользователя по клавиатурному почерку, и ведутся также разработки по его идентификации при подписи мышью [1. С. 189–192]. Одним из возможных автоматических средств ввода рукописного документа и представления его в электронном виде (т. е. как *электронный документ (ЭД)*) кроме мыши является также PC Notes Taker — цифровая координатная ручка, являющаяся принципиально новой системой ввода рукописного текста. PC Notes Taker позволяет одновременно получать как обычный бумажный документ с подписью руководителя, так и его электронную копию. Важно отметить, что электронная копия содержит дополнительные признаки почерка, которые не содержатся в его бумажном представлении. Эти признаки отражают динамические свойства почерка, формирующие изображения символов во времени исполнения текста документа. Устройство PC Notes Taker предлагается использовать в составе АСОИБ, например, для защиты информации. Защищать предлагается следующим образом. Пользователь составляет текст документа (т. е. обычный бумажный документ, например ценную бумагу — вексель), содержащий значимую информацию, и скрепляет этот документ своей подписью (индоссамент), что подтверждает его подлинность. Применение PC Notes Taker позволяет одновременно с бумажным подписанным документом получить его электронную копию (с этой подписью), которая может быть передана по каналам связи. В случае искажения информации злоумышленником предлагаемый способ позволяет выявить наличие возможных искажений (несоответствий) информации. Возможными искажениями могут быть следующие несоответствия: пол человека, его возраст или рост, а также язык текста представления информации. Существуют алгоритмы, позволяющие провести в некоторых случаях и саму идентификацию человека, который составил данный документ, не только по его бумажному варианту, но и по электронной копии этого



документа. Если идентификация дала отрицательный результат (т. е. информация в документе представлена не тем лицом, которым должна быть представлена), это означает, что информация была искажена, а сам документ следует признать фальшивым (т. е. неподлинным).

В АСОИБ предлагается использовать алгоритмы, разработанные для области криминалистики и используемые в различных методиках и методических рекомендациях, представленные в таблице 1.

Таблица 1. Методическое обеспечение [1]

Название	Год	Примечание	№
<b>Методика</b> вероятностно-статистической оценки совпадающих частных признаков почерка в прописных буквах русского алфавита	1990	<i>Область применения:</i> почерковедческая экспертиза; идентификация личности по почерку и АСОИБ <i>Объекты исследования:</i> векселя, иногда денежные знаки, обрывки текста и информация в ЭД и др.	1
<b>Методика</b> определения пола исполнителя кратких рукописных текстов [3]	1990	<i>Область применения:</i> почерковедческая экспертиза и АСОИБ <i>Объекты исследования:</i> векселя, обрывки текста, информация документа (ЭД) и др. <i>Автоматизация:</i> программа “man-WO-man”	2
<b>Методические</b> рекомендации. Определение языка по текстам, выполненным на основе русской и латинской письменной графики	1991	<i>Область применения:</i> экспертиза текста документа для определения языка текста представления информации <i>Объекты исследования:</i> векселя, обрывки текста, а также информация в ЭД и др. <i>Автоматизация:</i> программа “Ling-R”	3
<b>Методика</b> определения возраста исполнителя рукописных текстов	1995	<i>Область применения:</i> почерковедческая экспертиза и АСОИБ <i>Объекты исследования:</i> векселя, обрывки текста и информация в ЭД и др.	4
<b>Методические</b> рекомендации по определению подлинности бланков ценных бумаг	2000	<i>Область применения:</i> экспертиза ценных бумаг и АСОИБ <i>Объекты исследования:</i> ценные бумаги — векселя, акции, сертификаты и некоторые типы ЭД и др.	5
<b>Методика</b> определения роста исполнителя рукописных документов (ЭД)	≈2009	Методика находится в стадии разработки. Для АСОИБ разработана пробная экспериментальная версия.	6

При создании АРМ были выделены следующие три группы задач [7]: 1 — разработка алгоритма для методики (см. таблицу 1, строка 6) принятия решений о росте исполнителя документа (ЭД) *лицом, принимающим решение* (ЛПР); 2 — разработка структуры АРМ для АСОИБ; 3 — разработка первой версии АРМ как подсистемы АСОИБ. В таблице 2 приведены для *фактографической базы данных* (ФБД) ее краткое название и соответствующее *программное обеспечение* (ПО), позволяющее решать многие задачи, связанные, например, с анализом и обработкой информации, представленной в соответствующем документе. В таблице 3 приведены программы для работы с ценными бумагами и программы для криминалистики. Эти программы



(таблицы 2–3) и реализованные в них алгоритмы могут использоваться и в АРМ для АСОИБ. Однако обзор различных источников показал, что отсутствует программное обеспечение для АРМ эксперта, особенно для принятия решения по документам, содержащих КРТ.

Опыт [1, 2, 3, 4, 8] разработки различных систем выявил, что в качестве показателей эффективности средств автоматизации эксперта разумно использовать:  $T = \sum_x A_i \cdot top_i$  – временные и  $C = \sum_x B_i \cdot cor_i$  – денежные затраты на проведение экспертизы, где суммирование ведется по множеству операций  $X$  с учетом весовых коэффициентов  $A_i$  и  $B_i$ . Программы имеет смысл разрабатывать при помощи языка программирования РНР и использовать СУБД FireBird (как СУБД с открытым кодом, не требующую сложного администрирования), что позволяет эффективно организовать одновременную работу как локальных, так и удаленных пользователей АСОИБ.

В настоящее время реализованы следующие режимы АРМ: режим определения пола по почерку и режим определения языка текста представления информации. Следующий этап разработки АРМ “FHWE v. 1.0” – это этап модификации существующих (см. таблицу 1, строки 1, 2, 3, 4) и разработка новых (см. таблицу 1, строка 6) алгоритмов решения диагностических задач при проведении экспертиз. Базовая структура АРМ как подсистема АСОИБ представлена на рис. 1.

База данных **DBSSFHWE** и ПО **SSFHWE** подсистемы определения пола исполнителя краткого рукописного текста, БД **DBRSFHWE** и ПО **RSFHWE** подсистемы распознавания для АРМ эксперта успешно прошли государственную экспертизу и были защищены охранными документами Роспатента [8, 10, 11, 12]. Полученные результаты позволили разработать специальное устройство определения фальшивых рукописных документов на русском языке [13]. В состав этого устройства (как и в АРМ) входят: блок вычисления решающей функции, блок сравнения с порогом и блок принятия решения.

Таблица 2. Фактографические базы данных и программное обеспечение систем [1, 2]

№	Программное обеспечение	ФБД	Примечание
1	BUMAGI	IMAGBUM, HTMLBUM	Электронный атлас изображений ценных бумаг РФ
2	man-WO-man	FBD-MWM	Фактографическая система для определения пола исполнителя рукописи (подсистема распознавания)
3	Ling-R	LingRDA, LingRDB	Автоматизированная фактографическая ИПС “Ling” для распознавания языка по печатным или рукописным текстам, выполненным на основе русской или латинской письменной графики
4	BANKIR	FBD-BANK	Фактографическая информационно-поисковая система по коммерческим банкам России
5	BUMAGA (макет системы)	—	Макет АФИПС с фактографическими данными о векселях РФ
6	MA\$TER&B (макет системы)	—	Макет АФИПС с фактографическими данными о ценных бумагах РФ



Таблица 3. Программы [2] для криминалистики и для работы с ценными бумагами (ЦБ)

Номер патента	Язык	Примечание	№
930023	Clipper 5.01	Учет ЦБ (сертификатов акций). Анализ состояния рынка ЦБ. Обслуживание операций при работе с ЦБ.	1
930070	PAL	Система учета ЦБ. Ведение реестра владельцев ЦБ.	2
940388	Clarion Professional Developer 2.1	Чеки России. Учет чеков, не подлежащих приему как изъятые из обращения.	3
960464	MS Excel 5.0, Delphi	АРМ трейдера. Поддержка принятия решений на рынке ГКО-ОФЗ.	4
970096	Power Builder 4.0	Учет бланков строгой отчетности.	5
990500	Paradox v.7.0, Delphi 4, C++ Builder 3.5	Учет бланков строгой отчетности.	6
2000610513	Progress v.8.2A	АРМ учета бланков строгой отчетности.	7
2001610874	Delphi 5, MS Visual C++, Java, PL/SQL	Рабочее место брокера. Торговля ЦБ через сеть Интернет.	8
2001611152	Borland C++ Builder	Справочно-информационная система о бланках, подписях и печатях.	9
960357	C++	Аутентификация почерка.	10
980020	MS Visual C++ 4.2	Идентификация по изображениям в криминалистике.	11
980313	Delphi, C++	Справочная система о подлинных паспортах, водительских удостоверениях и способах их подделки.	12
990182	Delphi	Анализ русскоязычных текстов с целью выявления психологических черт автора.	13
990729	Delphi 4.5	Экспертиза текстов внушения.	14
2000610375	Arc View GIS, Avenu	АРМ оперативного дежурного.	15
2000611305	Visual Studio 98, Oracle PL/SQL	Идентификация наркотических веществ. Рабочее место эксперта.	16

В настоящее время заявка [13] на изобретение (в виде полезной модели) успешно прошла государственную экспертизу, и уже получено положительное решение о выдаче патента на устройство в виде полезной модели.

Рассмотрим более подробно методику [3]. Порядок действий исследователя при использовании методики [3] в АРМ следующий: Шаг 1 — установить пригодность объекта для решения криминалистической задачи; Шаг 2 — выделить признаки и определить (с помощью АРМ) их суммарную информативность (суммы весов признаков); Шаг 3 — выбрать уровень доверия при принятии решения и определить (с помощью АРМ) величину суммарной информативности букв



(суммы весов входов дискриминатора второго уровня); Шаг 4 — принять (с помощью АРМ) решение о поле исполнителя рукописи и оценить вероятность ошибки, связанной с отнесением исполнителя документа к противоположному полу, и сделать общий вывод. Необходимо отметить, что кроме 28 прописных букв в [3] предусмотрено использование строчных букв, которые для простоты вычислений объединены в одну «логическую» букву, обычно обозначаемую латинской буквой «S». Признаки и их информативность определяются по таблицам 1 и 2 методики [3]. Алгоритм (таблица 4) принятия решения в методике определения пола исполнителя кратких рукописных текстов в разработанном АРМ можно представить в виде 21 шага (пункта). Отметим, что все действия по выполнению этих шагов алгоритма (таблица 4) производятся с помощью разработанного АРМ.

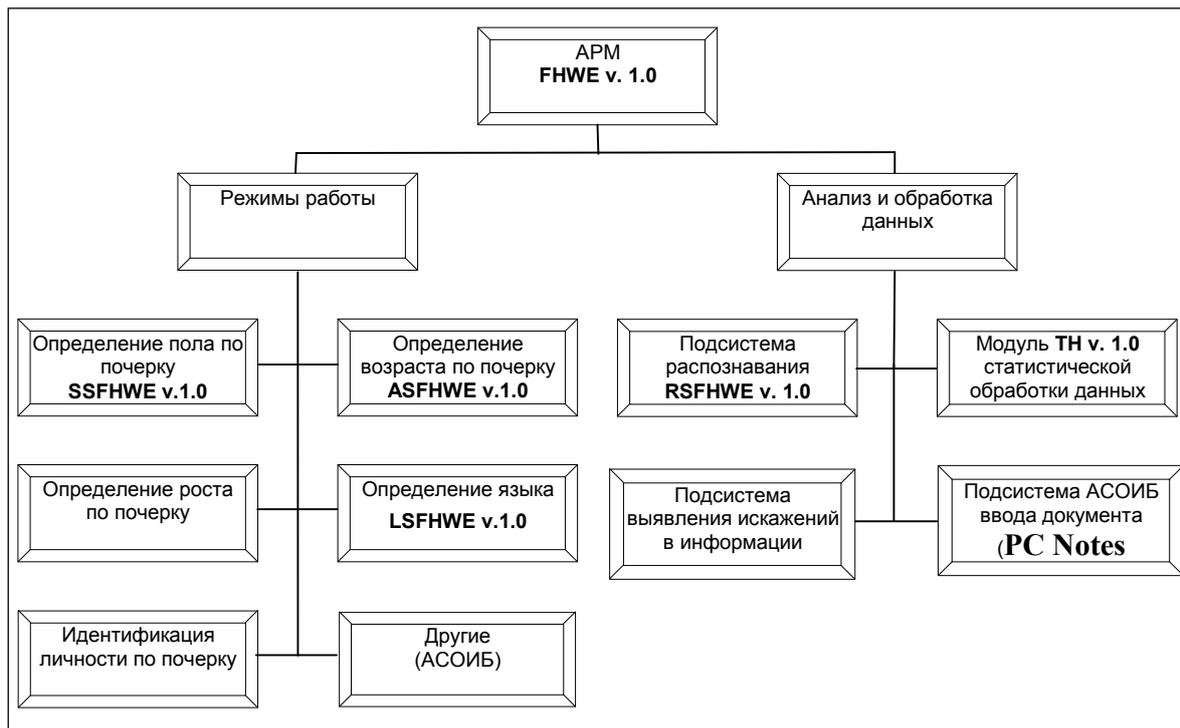


Рис. 1. Базовая структура АРМ как подсистема АСОИБ

Возможный вариант модификации данного алгоритма состоит в том, что вводятся дополнительные уровни доверия, например 0.93 или 0.95 или 0.97 и т. п. Тогда п. 3 данного алгоритма должен начинаться не с 0.9, а с самого большого уровня доверия, например с 0.97, или с другого, например 0.93 и т. п. Другой вариант модификации может состоять в том, что используется иной метод оценки вероятности допустимой ошибки, например с помощью выборки. Еще один вариант модификации состоит в применении другой нейронной сети, например не с пороговой функцией активации.

Таблица 4. Алгоритм принятия решения [3]

№ шага	Описание шага (пункта) алгоритма
1	Выделить признаки букв и определить их суммарные информативности по буквам.
2	В зависимости от знаков сумм информативностей признаков по буквам выдвинуть гипотезу о поле исполнителя текста. Если все суммы положительные, выдвинуть гипотезу о мужском поле исполнителя документа, если все суммы отрицательные, выдвинуть гипотезу о женском поле исполнителя документа, а если есть суммы разных знаков, то проверить обе гипотезы.



3	Выбрать уровень доверия <b>0.9</b> для принятия решения.
4	По таблицам 3 и 6 методики [3] определить значение порога и слагаемых (при $x = 0$ , $x = 1$ ) дискриминантной функции для всех выбранных букв.
5	Выполнить сравнение суммарных информативностей признаков по буквам с соответствующим порогом дискриминантной функции.
6	Если (при гипотезе о мужском поле исполнителя (ГОМПИ)) суммарная информативность признаков больше (при гипотезе о женском поле исполнителя (ГОЖПИ) – меньше) порога, то в качестве информативности буквы взять значение слагаемого дискриминантной функции при $x = 1$ , иначе берется значение слагаемого дискриминантной функции при $x = 0$ .
7	Определить значение суммарной информативности совокупности букв как сумму информативностей букв.
8	Если (при ГОМПИ) значение суммарной информативности совокупности букв больше или равно 0.6 (при ГОЖПИ – меньше или равно -0.6), то гипотеза о поле исполнителя текста подтверждается. Перейти к п. 17.
9	Если текущее значение уровня доверия <b>0.85</b> , то перейти к п. 14.
10	Если текущее значение уровня доверия <b>0.8</b> , то гипотеза о поле исполнителя документа не подтверждается. Перейти к п. 17.
11	Выбрать уровень доверия <b>0.85</b> для принятия решения.
12	По таблицам 4 и 6 методики [3] определить значение порога и слагаемых (при $x = 0$ , $x = 1$ ) дискриминантной функции для всех выбранных букв.
13	Перейти к п. 5.
14	Выбрать уровень доверия 0.8 для принятия решения.
15	По таблицам 5 и 6 методики [3] определить значение порога и слагаемых (при $x = 0$ , $x = 1$ ) дискриминантной функции для всех выбранных букв.
16	Перейти к п. 5.
17	Если необходимо проверить две гипотезы о поле исполнителя документа, а проверена пока только одна, то повторяется выполнение п. 3–16 для второй гипотезы.
18	Если при проверке только одной гипотезы она подтвердилась или если при проверке двух гипотез подтвердилась только одна, то результатом исследования являются результаты проверки этой гипотезы и соответствующие значения уровня доверия и вероятности допустимой ошибки. Принять решение и перейти к п. 20.
19	Если при проверке только одной гипотезы она не подтвердилась или если при проверке двух гипотез не подтвердилась ни одна или подтвердились обе, то по результатам исследования сделать вывод о невозможности принятия решения (НПВ). Перейти к п. 21.
20	По таблице 7 методики [3] найти максимальный порог дискриминантной функции, меньший или равный модулю значения суммарной информативности совокупности букв, и определить значение соответствующей вероятности допустимой ошибки.
21	Сделать общий вывод о принятом решении. СТОП.

Указанные модификации в большой степени касаются методики [3], однако могут быть применены и для других аналогичных методик (см. таблицу 1). Перейдем к подробному рассмотрению простого примера применения данной методики.



### Пример применения методики [3]

Пусть требуется определить пол исполнителя электронного документа (содержащего КРТ), представленного на рис. 2 в виде рукописного текста, выполненного с помощью PC Notes Taker. Пусть в этом документе с помощью некоторых специальных средств выделены характерные признаки букв, определена по заранее рассчитанным таблицам их информативность, а результаты этих действий занесены в таблицу 5. Суммарная информативность признаков по каждой букве электронного документа для удобства представлена в таблице 6.

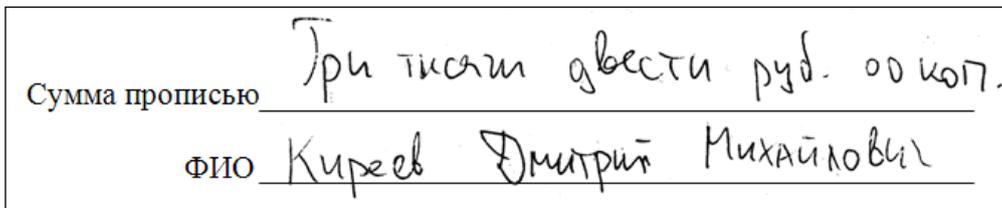


Рис. 2. Пример рукописного документа, полученного с помощью PC Notes Taker

Таблица 5. Частные признаки букв (Рис. 2) документа и их информативность

Буква	(Номер признака из [3]; Информативность из [3])
S	(3; 1.39) (18; 1.43) (29; 1.50)
Д	(15; 0.44) (19; 0.44) (20; 0.76) (32; -0.81)
К	(1; 2.23) (8; 0.49) (15; 0.42) (18; 2.05) (48; 3.00)
М	(1; 2.38) (10; 0.89) (20; -0.06)
Т	(2; 1.13) (34; 1.54) (47; 0.40)

Из данных, представленных в таблицах 5, 6, видно, что все суммы информативностей признаков по буквам положительные, следовательно, в блоке АСОИБ выдвигается гипотеза о мужском поле исполнителя электронного документа. Результаты вычисления суммарной информативности совокупности букв при гипотезе о мужском поле и уровне доверия 0.9 представлены в таблице 6.

Таблица 6. Выбор параметров дискриминатора при гипотезе о мужском поле и уровне доверия 0.9

Буква	Суммарная информативность признаков по каждой букве	Знак сравнения с порогом	Порог	Слагаемые дискриминантной функции		Информативность буквы	Пояснения
				x=0	x=1		
S	4.32	>	0.98	-1.07	<b>2.27</b>	2.27	4.32 > 0.98
Д	0.83	<	0.90	<b>-0.72</b>	2.25	-0.72	0.83 < 0.90
К	8.19	>	2.07	-0.68	<b>2.21</b>	2.21	8.19 > 2.07
М	3.21	>	1.27	-0.33	<b>2.36</b>	2.36	3.21 > 1.27
Т	3.07	<	3.12	<b>-0.32</b>	2.33	-0.32	3.07 < 3.12
Суммарная информативность совокупности букв						<b>5.80</b>	2.27-0.72+ 2.21+2.36-0.32



Результаты вычисления (по таблице 7 из [3]) вероятности ошибки, связанной с отнесением АСОИБ исполнителя электронного документа к противоположному полу, представлены в таблице 7.

Таблица 7. Определение значения вероятности допустимой ошибки

Модуль суммарной информативности совокупности букв	Порог	Вероятность ошибки
5.80	2.30	0.005

По результатам выполненного исследования объекта (см. рис. 2) можно сделать вывод, что исполнитель электронного документа именно мужчина, а вероятность допустимой ошибки 0.005. Аналогично применяют методику [3], используемую в АРМ как подсистеме АСОИБ, и в других случаях на практике. Если до проведения исследования было известно, что документ и должен был быть составлен мужчиной, то в АСОИБ принимается решение об отсутствии искажений в информации документа. Если до проведения исследования было известно, что документ должен был быть составлен женщиной, то в АСОИБ принимается решение о возможном наличии искажений в информации документа (в этом случае необходимо провести дополнительные исследования на подтверждение наличия искажений).

### Выводы

Существующие методики (и содержащиеся в них алгоритмы) для экспертов требуют модификации, так как с момента их разработки прошло достаточно много времени. Предложены три варианта возможной модификации существующих методик с точки зрения модификации алгоритма принятия решений. Обзор источников показал, что отсутствует ПО (и необходимые алгоритмы) для реализации АРМ эксперта, особенно для принятия решения по информации, содержащейся в документе, представленном кратким рукописным текстом. Состояние дел в области экспертиз указывает на необходимость создания автоматизированных средств проведения экспертиз по рукописным документам как в электронном, так и в бумажном виде.

Разработана базовая структура АРМ как подсистемы АСОИБ, ПО и БД для него. В качестве показателей эффективности средств автоматизации принято использовать временные и денежные затраты. Эксперименты с устройством PC Notes Taker показали, что оно может быть использовано в АСОИБ для защиты электронных документов от подделки.

По результатам проведенных исследований АСОИБ была подана заявка на выдачу патента на изобретение в виде полезной модели и получено положительное решение о выдаче этого патента. Основные полученные результаты защищены охранными документами Роспатента.

### СПИСОК ЛИТЕРАТУРЫ:

- Андрианов В. В., Галушкин А. И., Кулик С. Д. и др. Нейрокомпьютеры в биометрических системах. Кн. 26 / Под ред. А. И. Галушкина. М.: Радиотехника, 2007. — 192 с.
- Кулик С. Д. Объекты интеллектуальной собственности России (обзор программного обеспечения): Научное издание. М.: Компания Спутник+, 2001. — 159 с.
- Чельшев М. М., Кулик С. Д., Мирошникова С. Ю., Левицкий А. Б. и др. Методика определения пола исполнителя кратких рукописных текстов: Учебное пособие. М.: ВНКЦ МВД СССР, 1990. — 185 с.
- Кулик С. Д. Проектирование АФИПС криминалистического назначения // Безопасность информационных технологий. 2002. № 1. С. 78–81.
- Кулик С. Д., Никоненц Д. А. Автоматизация криминалистического исследования рукописных документов и вопросы безопасности // Научная сессия МИФИ-2008. XV Всероссийская науч. конф. «Проблемы информационной безопасности в системе высшей школы». Сб. науч. трудов. М.: МИФИ, 2008. С. 88–89.



6. Chaski C. E. Steps Toward a Science of Authorship Identification // National Institute of Justice Journal. September. 1997. P. 15–20.
7. Кулик С. Д., Никонцев Д. А., Воронкова М. М., Шелякова П. А. Разработка подпрограмм определения пола по почерку и языка текста в АРМ эксперта-криминалиста “FHWE V. 1.0” // Научная сессия МИФИ-2008. Сб. науч. трудов в 15 т. М.: МИФИ, 2008. Т. 13. С. 81–82.
8. Кулик С. Д., Никонцев Д. А. Свидетельство на программу РФ №2007614121 «Подсистема распознавания для АРМ эксперта-криминалиста v.1.0» (RSFHWE).
9. Кулик С. Д., Никонцев Д. А. Автоматизация криминалистического исследования рукописных текстов при помощи нейронных сетей // Научная сессия МИФИ-2008. X Всероссийская науч.-технич. конф. «Нейроинформатика-2008». Сб. науч. трудов. М.: МИФИ, 2008. Ч. 1. С. 15.
10. Кулик С. Д., Никонцев Д. А. Свидетельство на программу РФ № 2008610383 «Подсистема определения пола исполнителя краткого рукописного текста для АРМ эксперта-криминалиста v.1.0» (SSFHWE).
11. Кулик С. Д., Никонцев Д. А. Свидетельство на БД РФ № 2008620047 «БД подсистемы определения пола исполнителя краткого рукописного текста для АРМ эксперта-криминалиста v.1.0» (DBSSFHWE).
12. Кулик С. Д., Никонцев Д. А. Свидетельство на БД РФ № 2007620333 «БД подсистемы распознавания для АРМ эксперта-криминалиста v.1.0» (DBRSFHWE).
13. Кулик С. Д., Никонцев Д. А., Ткаченко К. И., Жижилев А. В. Заявка на выдачу Свидетельства на полезную модель, РФ (RU), кл. МПК<sup>7</sup> G 07 D 7/00. Устройство определения фальшивых рукописных документов на русском языке. Заявка № 2007147832/20; Заяв. 25.12.2007 г.; Приоритет от 25.12.2007 г. (РОСПАТЕНТ).
14. Кулик С. Д., Никонцев Д. А. Разработка подсистемы распознавания “RSFHWE V. 1.0” в АРМ эксперта-криминалиста “FHWE V. 1.0” // Научная сессия МИФИ-2008. Сб. науч. трудов. М.:МИФИ, 2008. Т. 13. С. 31.

