*S.D. Kulik*

**Special Tools For Ensuring Information Security**

*Keywords: information security, question answering systems, information retrieval system*
Summary: This article is about methods of information security. These methods are a part of the automated tools of ensuring the information security. It is offered to use aspecial information retrieval. The results were protected with patents.

*С.Д. Кулик*

# СПЕЦИАЛЬНЫЕ СРЕДСТВА ДЛЯ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В последнее время все большее значения приобретают информационные системы, вопросно-ответные системы (QAsystems), фактографические системы [1-6] и фактографические базы данных (ФБД) [7]. В этих системах эффективность поиска [8] имеет (например, для фактографических систем [9]) большое и порой решающее значение. QA-systems – это одна из разновидностей фактографических систем. Особое место в фактографических системах [9, 10] занимает фактографический поиск (ФП), который является ключевым инструментальным средством фактографических систем.

Специальные средства ФП успешно применяются на практике во многих областях, например в фактографических системах [10], в криминалистике и при поиске различных объектов в исследовании операций [11]. Однако при этом в настоящее время проблемы эффективного применения специальных средств ФП для обеспечения информационной безопасности слабо исследованы и практически не решены. Отчасти это связано с тем, что специалистов по ФП не так уж много, не говоря уже и о публикациях в этой области

В настоящее время в области информационных технологий специалисты среди информационных систем выделяют класс очень важных для практики систем, которые получили название [10] автоматизированные фактографические информационно-поисковые системы (АФИПС). Для повышения эффективности работы автоматизированных средств обеспечения информационной безопасности (АСОИБ) предлагается использовать в ее составе специальные средства фактографического поиска. Качественное отличие фактографической информационно-поисковой системы (ФИПС) от просто информационно-поисковой системы(ИПС) состоит в том, что именно ФИПС позволяет получать ответы на фактографические запросы. На практике одной из главных задач, решаемых фактографическими средствами, является ответ на фактографические запросы, поступающие от пользователя АСОИБ, и обеспечение возможности специализированного фактографического поиска с помощью оператора АСОИБ.

Для АСОИБ предлагается ввести следующие частные показатели, связанные с оценкой эффективности защиты информации в фактографических данных:

$A_1$ – оценка вероятности не искажения информации от внутреннего злоумышленника;

$A_2$ – оценка вероятности не искажения информации от внешнего злоумышленника;

$A_3$ – оценка вероятности не искажения информации от вредоносного *программного обеспечения* (например, компьютерного вредоносного вируса);

$A_4$ – оценка вероятности не искажения информации от вредоносного *аппаратного обеспечения* (например, закладочных устройств, аппаратных закладок, т.е. устройств в электронной схеме, скрытно внедряемых к остальным элементам);

$A_5$ – оценка вероятности не искажения информации от внешнего вредоносного воздействия на: *аппаратуру*, *программное обеспечение*, *фактографические данные* (например, электромагнитное воздействие).

Опираясь на эти введенные показатели и на работы [9, 10, 12], была оценена эффективность ФП в АСОИБ. Подход к оценке информационной безопасности для фактографическихсистемс помощью набора показателей $\{A_1, A_2, A_3, A_4, A_5\}$ является новым. На практике злоумышленник может частично разрушить (исказить) содержимое записи данных в фактографической БД, при этом цель фактографических средств – найти фактографические данные необходимые для эффективного функционирования АСОИБ. На рис. 1 представлена схема специализированной фактографической АСОИБ. Работа такой АСОИБ выполняется следующим образом.

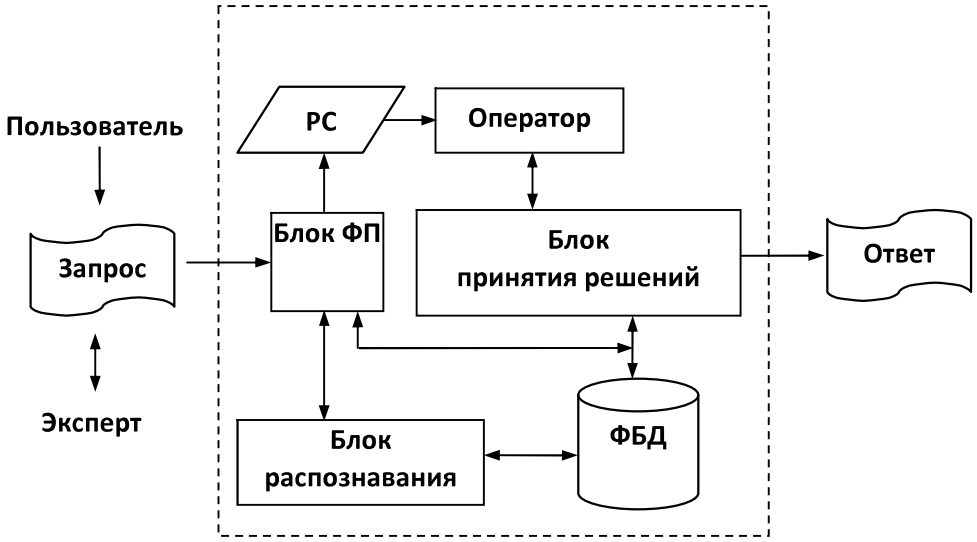


Рис. 1. Схема специализированной фактографической АСОИБ

На практике важные юридически значимые документы, как правило, подписываются (скрепляются рукописной записью (подписью)). Злоумышленник может исказить информацию и подделать подпись. Пользователь АСОИБ хочет убедиться в достоверности информации. Для этого он посылает запрос в АСОИБ. Эксперт, получив документ, выполняет предварительный анализ его реквизитов, элементов защиты и почерка. В итоге заиндексированный документ в виде поискового образа документа (ПОД) поступает в блок ФП, в котором реализована одна из возможных стратегий фактографического поиска. В простейшем случае в блоке ФП может быть реализована двухэтапная стратегия полного фактографического поиска. Суть ее состоит в следующем. На первом этапе в соответствии с ПОД необходимо просмотреть записи ФБД и выбрать (с помощью алгоритма распознавания, например с помощью [13, 14] нейросетевого алгоритма) те из них, которые соответствуют запросу. Список похожих записей (например, их регистрационные номера (РН) и все необходимые фактографические данные для оператора и принятия окончательного решения на поступивший запрос от пользователя) заносятся в рекомендательный список (РС). Каждому выпущему подлинному документу соответствует одна запись в ФБД. Всего таких записей может быть N. Такая за-

пись ФБД соответствует подлинному документу (фактографические сведения о всех подлинных документах заранее заносятся в ФБД). Если в результате поиска для поступившего документа на проверку не обнаружено в ФБД соответствующему запросу подлинного документа (т.е. принято окончательное решение об его отсутствии), то принимается итоговое решение о том, что поступивший документ является фальшивым.

В блоке распознавания текущий ПОД и поисковый образ объекта (ПОО) сравниваются с помощью алгоритма распознавания (например, нейросетевого алгоритма), при этом похожие на запрос ПОО заносятся в РС. При распознавании возможны ошибки первого и второго рода:

$(1–P_1)$ – вероятность *пропуска цели*; $(1–P_2)$ – вероятность *ложной тревоги*.

Будем полагать, что случайные события, соответствующие вероятностям $A_k$ являются независимыми в совокупности событиями. Тогда, опираясь на работу [12], можно показать, что $u_1 = P_1\prod_{k=1}^{5} A_k = P_1A_1A_2A_3A_4A_5$, $u_2 = P_2\prod_{k=1}^{5} A_k = P_2A_1A_2A_3A_4A_5$, где оценки вероятностей $u_1$, $u_2$ вобщем характеризуют эффективность распознавания и защиты информации в фактографических данных.

В случае стратегии полного ФП поиск останавливается, если заполнен выделенный буфер под РС или просмотрена вся область поиска в ФБД (нет больше текущих записей с ПОО для распознавания). После этого РС передается на обработку оператору. Оператор просматривает записи РС и отбирает те из них, которые наиболее соответствуют запросу пользователя. Окончательное решение по запросу пользователя принимается в блоке принятия решения. Если в итоге в ФБД о документе, поступившем в качестве запроса, нет сведений среди подлинных документов, то принимается решение что этот поступивший документ – не подлинный.

Эффективность работы такой специализированной фактографическойАСОИБ зависит от длины буфера, выделенного под хранения элементов РС. Эту длину будем обозначать как $L$. В АСОИБ обработку РС выполняет человек (на схеме АСОИБ он обозначен как оператор). Этот оператор выполняет трудоемкие ручные операции и в связи с этим имеет ограниченные возможности по обработки РС. На практике желательно обеспечить короткие РС, однако слишком короткие РС приводят в итоге к ошибочным решениям (как правило, к пропуску цели). Слишком длинные РС приводят к увеличению времени обработки запроса пользователя. Поэтому далее основное внимание будет уделено именно длине РС (числу элементов РС).

Опираясь на работу [12] можно показать, что средняя длина РС $\bar{L}$ (при некоторых ограничениях) для фактографического поиска может быть оценена по следующей формуле:

$$\bar{L} \approx \sum_{m=0}^{L-1}[\lambda_m m] + \sum_{m=L}^{N}[\lambda_m L],\ \lambda_m = \frac{N!}{m!(N-m)!}u_2^{N-m}(1-u_2)^m.$$

Анализ этой формулы показывает, что (при $D = \frac{L}{N}$, $E = (1-u_2)$ и $Q = 0$)

$$\bar{L} \approx N \cdot \begin{cases} Q, \ если\ u_2 = 1; \\ D, \ если\ u_2 = Q; \\ E, \ если\ L = N. \end{cases}$$

Таким образом, варьируя показатели $u_2$ и $L$, можно регулировать среднюю длину РС. Эти данные полностью согласуются с результатами, полученными, например, в работе [10].

Введем следующее обозначение $\rceil x \lceil$ – округление до целой части числа, такое, что $x \approx \rceil x \lceil$. В табл. 1–4 приведены результаты численного исследования $\overline{L}$ при разных значениях показателей $N$ и $u_2$. В табл. 4 представлены результаты исследований при относительно большом значении $N$, а в табл.1–3 – при малом значении $N$.

*Таблица 1. Пример оценки $\overline{L}$ при $N = 50$ и $u_2 = 0,68$*

| Показатели | Значения показателей | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $L$ | 1 | 3 | 6 | 8 | 13 | 25 | 30 | 35 | 40 | 50 |
| $\rceil \overline{L} \lceil$ | 1 | 3 | 6 | 8 | 13 | 16 | 16 | 16 | 16 | 16 |

*Таблица 2. Пример оценки $\overline{L}$ при $N = 50$ и $u_2 = 0,98$*

| Показатели | Значения показателей | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $L$ | 4 | 5 | 7 | 10 | 11 | 12 | 16 | 20 | 40 | 50 |
| $\rceil \overline{L} \lceil$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*Таблица 3. Пример оценки $\overline{L}$ при $N = 50$ и $u_2 = 0,78$*

| **Показатели** | **Значения показателей** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $L$ | 1 | 5 | 7 | 10 | 11 | 12 | 16 | 20 | 40 | 50 |
| $\rceil \overline{L} \lceil$ | 1 | 5 | 7 | 9 | 10 | 10 | 11 | 11 | 11 | 11 |

*Таблица 4. Пример оценки $\overline{L}$ при $N = 100$ тыс. и $u_2 = 0,997$*

| Показатели | Значения показателей | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $L$ | 1 | 5 | 10 | 50 | 100 | 250 | 280 | 285 | 290 | 295 | 300 | 350 | 500 | 1000 | 2000 |
| $\rceil \overline{L} \lceil$ | 1 | 5 | 10 | 50 | 100 | 250 | 279 | 283 | 287 | 290 | 293 | 300 | 300 | 300 | 300 |

Из табл. 1–3 видно, как зависит $\overline{L}$ от изменения $L$. Результаты табл. 4 показывают характерную зависимость для $\overline{L}$ от параметра $L$: сначала имеется почти линейный участок, затем нелинейный и в конце наблюдается насыщение (показатель $L$ растет, а $\overline{L}$ – практически нет).

Выполненные исследования и полученные результаты позволили наметить пути применения фактографического поиска в АСОИБ. Была успешно представлена новая идея АСОИБ с элементами фактографического поиска. Такая АСОИБ позволяет эффективно решать практические задачи обеспечения информационной безопасности. Для АСОИБ получена важная аналитическая оценка предложенного показателя $\overline{L}$ эффективности фактографического поиска. Выполнено исследование этой оценки и выявлены важные свойства. По результатам проведенных исследований были успешно получены необходимые охранные документы Российского агентства по патентам и товарным знакам (РОСПАТЕНТ).

## СПИСОК ЛИТЕРАТУРЫ:

1. Гиляревский Р.С. Основы информатики. М.: Экзамен, 2003. 320 с.

2. Когаловский М.Р. Перспективные технологии информационных систем. Сер.: ИТ-Экономика. М.: ДМКПресс, 2003. 288 с.
3. Chorng-Shyong Ong, Min-Yuh Day, Wen-Lian Hsu. Development of an evaluation model for Question Answering Systems // IEEE International Conference on Information Reuse and Integration, IRI 2008 (13-15 July 2008), 2008, pp. 178-183.
4. Лахути Д.Г. Автоматизированные документально-фактографические информационно-поисковые системы // Итоги науки и техники. Сер.: Информатика. Т. 12. М.: ВИНИТИ, 1988. С. 6-79.
5. Солтон Дж. Динамические библиотечно-информационные системы. М.: Мир, 1979. 557 с.
6. Ланкастер Ф.У. Информационно-поисковые системы: характеристики, испытание и оценка. М.: Мир, 1972. 308 с.
7. Гайдамакин Н.А. Автоматизированные информационные системы, базы и банки данных. М.: Гелиос АРВ, 2002. 368с.
8. Кнут Д.Э. Искусство программирования. Т.3: Сортировка и поиск. М.: Вильямс, 2004. 822 с.
9. Соколов А.В. Информационно-поисковые системы. М.: Радио и связь, 1981. 152 с.
10. Кулик С.Д. Оценка эффективности поисковых операций// Прикладная информатика, 2014. №6(54). С. 60-69.
11. Кулик С.Д., Никонец Д.А., Ткаченко К.И., Жижилев А.В. Устройство определения поддельных документов // Безопасность информационных технологий, 2009. №1. С. 114-115.
12. Колмогоров А.Н. Основные понятия теории вероятностей. М.: ФАЗИС, 1998. 144 с.
13. Иванов А.И. Нейросетевые алгоритмы биометрической идентификации личности /Под ред. А.И. Галушкина. М.: Радиотехника, 2004. 144 с.
14. Хайкин С. Нейронные сети: полный курс. М.: Вильямс, 2006. 1104 с.

## REFERENCES:

1. Gilyarevskiy R.S. Osnovy informatiki. M.: Ekzamen, 2003. 320 p.
2. Kogalovskiy M.R. Perspektivnyye tekhnologii informatsionnykh sistem. Ser.: IT-Ekonomika. M.: DMK Press, 2003. 288 p.
3. Chorng-Shyong Ong, Min-Yuh Day, Wen-Lian Hsu. Development of an evaluation model for Question Answering Systems // IEEE International Conference on Information Reuse and Integration, IRI 2008 (13-15 July 2008), 2008, pp. 178-183.
4. Lakhuti D.G. Avtomatizirovannyye dokumental'no-faktograficheskiye informatsionno-poiskovyye sistemy // Itogi nauki i tekhniki. Ser.: Informatika. T.12. M.: VINITI, 1988. pp. 6-79.
5. Solton Dzh. Dinamicheskiye bibliotechno-informatsionnyye sistemy. M.: Mir, 1979. 557 p.
6. Lankaster F.U. Informatsionno-poiskovyye sistemy: kharakteristiki, ispytaniye i otsenka. M.: Mir, 1972. 308 p.
7. Gaydamakin N.A. Avtomatizirovannyye informatsionnyye sistemy, bazy i banki dannykh. M.: Gelios ARV, 2002. 368 p.
8. Knut D.E. Iskusstvo programmirovaniya. T. 3 : Sortirovka i poisk . M .: Vil'yams , 2004. 822 p.
9. Sokolov A.V. Informatsionno-poiskovyye sistemy. M.: Radio i svyaz', 1981. 152 p.
10. Kulik S.D.Otsenka effektivnosti poiskovykh operatsiy // Prikladnaya informatika, 2014. №6(54). pp.60-69.
11. Kulik S.D., Nikonets D.A., Tkachenko K.I., Zhizhilev A.V. Ustroystvo opredeleniya poddel'nykh dokumentov // Bezopasnost' informatsionnykh tekhnologiy , 2009. №1 . pp. 114-115.
12. KolmogorovA.N. Osnovnyyeponyatiyateoriiveroyatnostey. M.: FAZIS, 1998. 144 p.
13. Ivanov A.I. Neyrosetevyye algoritmy biometricheskoy identifikatsii lichnosti / Pod red . A.I. Galushkina . M .: Radiotekhnika, 2004. 144 p.
14. Khaykin S. Neyronnyye seti: polnyy kurs. M.: Vil'yams, 2006. 1104 p.