УДК 004.91:004.056.53

ОБ ИНДЕКСНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ ПРИ ЗАЩИТЕ ДОКУМЕНТОВ НА ОСНОВЕ ЛЕКСИКОЛОГИЧЕСКОГО СИНТЕЗА

¹Трофимова А.В., ^{1,2}Черников Б.В.¹РЭУ им. Г.В. Плеханова, Москва, e-mail: 13goan@gmail.com;²ООО «Газпром ВНИИГАЗ», Москва

В статье исследуется технология защиты текстовых документов на основе лексикологического синтеза, основой которого является набор унифицированной документации по форме и содержанию и формирование индексной последовательности. В унифицированных документах выделяются опорные слова, и по ним строится лексикологическая схема документа. Данная схема распределена по уровням уточнения информации: вид документа, заголовок документа и далее информация, присущая отдельному виду документа. Параллельно созданию документа по лексикологической схеме формируется индексная последовательность этого документа. Цель ее формирования – защищенное хранение и передача документа по каналам связи в виде информационной посылки. В тексте статьи описывается процесс фиксации индексной последовательности, приводятся аспекты работы с индексной последовательностью, которые необходимо структурировать и формализовать. Разбирается пример формирования последовательности по фрагменту лексикологической схемы организационно-распорядительной документации вуза. В процессе исследования определяется необходимость разработки методологической модели для формализации процесса и повышения эффективности создания индексной последовательности.

Ключевые слова: электронный документ, защита сведений, лексикологический синтез, индексная последовательность, информационная посылка

THE INDEX SEQUENCE IN PROTECTION OF DOCUMENTS BASED ON LEXICOLOGICAL SYNTHESIS

¹Trofimova A.V., ^{1,2}Chernikov B.V.¹Plekhanov Russian University of Economics, Moscow, e-mail: 13goan@gmail.com;²ООО «Gazprom VNNIGAZ», Moscow

The article investigates the protection of text documents on the basis of lexicological synthesis. The basis of lexicological synthesis is a set of standardized documentation in form and content and the formation of an index sequence. In standardized documents, reference words are allocated, and a lexicological diagram of the document is constructed on them. This scheme is distributed according to the levels of information clarification: the type of document, the title of the document and further information inherent in a separate type of document. Parallel to the creation of a document on the lexicological scheme, the index sequence of this document is formed. The purpose of its formation is the secure storage and transmission of the document through communication channels in the form of an information message. The process of fixing the index sequence is described. The aspects of work with the index sequence that need to be structured and formalized are introduced. An example of the formation of a sequence according to the fragment of the lexicological scheme of organizational and administrative documentation of the university is analyzed, determined the need to develop a methodological model for the formalization of the process and effectiveness of creating the index sequence. The methodological model of the formation of the index sequence makes it possible to effectively use document protection technology based on lexicological synthesis.

Keywords: electronic documents, data protection, lexicological synthesis, index sequence, information package

В любой современной организации циркулирует информация, относящаяся к конфиденциальной: персональные данные сотрудников и клиентов, коммерческая тайна, служебная тайна. В целях обработки такой информации в соответствии с требованиями законодательства необходимо осуществлять ее комплексную защиту от различных угроз. Основной опасностью является несанкционированный доступ к данным, который влечет за собой нанесение материального и других видов ущерба организациям.

Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных» для комплексной защиты конфиденциальной информации определяет использование организационно-правовых и технических мер в совокупности [1].

Наиболее распространенный технический метод защиты информации – шифрование, но и он имеет свои недостатки, самым существенным из которых следует считать то, что для шифрования необходим уже готовый документ. Перспективным способом для обеспечения безопасности информации при ее хранении и передаче по каналам связи является защита текстовых документов на основе лексикологического синтеза [2]. Использование данной технологии позволяет оперировать не самим документом, даже если он зашифрован, а информационной посылкой, при перехвате которой вероятность раскрытия сведений из документа минимальна.

Актуальность и цель исследования

Обеспечение конфиденциальности информации достигается путем применения

комплексной защиты, в том числе и технических мер. Шифрование информации является самым распространенным методом защиты информации от утечки. Но его использование связано со значительными ресурсными затратами, такими как время и производительность процессоров. Кроме того, шифрованию подлежит уже готовый документ. С целью снижения затрат на защиту информации при передаче электронных документов по каналам связи необходимо рассмотреть альтернативу шифрованию, способ, минимизирующий использование различных ресурсов. Перспективным методом является защита текстовых документов на основе лексикологического синтеза, базисом которого является унификация документов и формирование индексной последовательности, передаваемой по каналу связи [2].

Верхнеуровневый процесс использования технологии лексикологического синтеза выглядит следующим образом. На стороне отправителя в автоматизированном режиме формируется документ, и параллельно этому процессу по определенным правилам фиксируется последовательность символов, называемая индексной последовательностью. При ее перехвате восстановить документ практически невозможно без такой базы знаний, как у отправителя. После создания документа индексная последовательность передается по каналу связи, и на ее основе получатель восстанавливает также по определенным правилам исходный документ. Эффективность защиты текстовых документов на основе лексикологического синтеза во многом зависит от процесса формирования индексной последовательности. Необходимо ли уделять формализации данного процесса основное внимание?

Цель исследования – определить аспекты формирования индексной последовательности, которые необходимо формализовать для эффективного использования технологии создания документов на основе лексикологического синтеза.

Правовая основа защиты информации

Началом защиты конфиденциальной информации является ее правовое обеспечение. Основополагающий правовой документ – Конституция Российской Федерации от 12 декабря 1993 г. [3]. В ст. 23 и 24 сказано: «Каждый имеет право на неприкосновенность частной жизни, личную и семейную тайну, защиту своей чести и доброго имени. Каждый имеет право на тайну переписки, телефонных переговоров, почтовых, телеграфных и иных сообщений. Сбор, хранение, использование и распространение

информации о частной жизни лица без его согласия не допускаются» [3].

К информации конфиденциального характера в соответствии с Указом Президента РФ от 6 марта 1997 г. № 188 «Об утверждении перечня сведений конфиденциального характера» относятся: персональные данные, коммерческая тайна, служебная тайна, профессиональная тайна, тайна следствия и судопроизводства, а также сведения о сущности изобретения, полезной модели или промышленного образца до официальной публикации [4]. Согласно ст. 16 Федерального закона от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации» «Защита информации представляет собой принятие правовых, организационных и технических мер, направленных на:

1) обеспечение защиты информации от неправомерного доступа, уничтожения, модифицирования, блокирования, копирования, предоставления, распространения, а также от иных неправомерных действий в отношении такой информации;

2) соблюдение конфиденциальности информации ограниченного доступа;

3) реализацию права на доступ к информации» [5].

Каждый из видов информации конфиденциального характера, и требования к его защите определяются также другими нормативными актами, к которым можно отнести следующие: Федеральный закон РФ от 21.07.1993 № 5485-1 «О государственной тайне», Федеральный закон от 29.07.2004 № 98-ФЗ «О коммерческой тайне», Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных», Федеральный закон от 02.12.1990 № 395-1 «О банках и банковской деятельности» и другие. Например, согласно Федеральному закону от 27.07.2006 № 152-ФЗ «О персональных данных» «персональные данные – любая информация, относящаяся к прямо или косвенно определенному или определяемому физическому лицу (субъекту персональных данных)», а также операторы и иные лица, получившие доступ к персональным данным, обязаны не раскрывать третьим лицам и не распространять персональные данные без согласия субъекта персональных данных, если иное не предусмотрено федеральным законом [1].

Нормативно-правовые акты дают основу для построения системы защиты информации.

Организационные меры защиты информации

Система защиты информации состоит из множества методов и средств защиты.

В работе [6] М.В. Тулупов наиболее важным из них выделяет комплекс специальных мер правового и административного характера, организационных мероприятий.

Практически во всех организациях первым эшелоном защиты конфиденциальной информации являются именно организационные меры. К ним следует отнести такие мероприятия, как контроль действий сотрудников, аудит текущей системы защиты информации, разработку локальных нормативных документов, регламентирующих работу с конфиденциальной информацией, и другие. Однако организационные меры защиты не обеспечивают безопасность на техническом уровне, не учитывают технические каналы утечки информации.

Технические меры защиты информации

В состав мер по обеспечению безопасности персональных данных, реализуемых в рамках системы защиты персональных данных с учетом актуальных угроз безопасности персональных данных и применяемых информационных технологий, входят [7]:

- идентификация и аутентификация субъектов доступа и объектов доступа;
- управление доступом субъектов доступа к объектам доступа;
- ограничение программной среды;
- защита машинных носителей информации, на которых хранятся и (или) обрабатываются персональные данные;
- регистрация событий безопасности;
- антивирусная защита;
- обеспечение целостности информационной системы и персональных данных;
- обеспечение доступности персональных данных;
- защита технических средств;
- защита информационной системы, ее средств, систем связи и передачи данных;
- управление конфигурацией информационной системы и системы защиты персональных данных и др.

При передаче данных по каналам связи наиболее распространенным методом защиты является шифрование содержимого документа, а также организация виртуальных частных сетей. В работе [8] указаны основные недостатки при использовании шифрования, такие как время на организацию процессов шифрования и дешифрования, требования к производительности процессов, сложности управления ключами.

Комплексная защита информации обеспечивается совместным использованием как организационных, так и технических мер защиты информации.

Защита данных на основе лексикологического синтеза

За основу данной технологии защиты берется автоматизированное создание документов, которые унифицированы по форме и содержанию. Каждый синтезируемый документ раскладывается на опорные слова, которым соответствуют текстовые фрагменты. Сформированный набор формулировок сохраняется в виде индексированной совокупности элементов (например, в виде базы знаний). Каждой формулировке ставится в соответствие основное слово, выбор которого однозначно определяет наличие конкретной формулировки в документе. Такие слова называются опорными и образуют основу лексикологической схемы формируемого документа (возможный вид лексикологической схемы приведен на рис. 1), которая фактически представляет собой онтологию автоматизируемого комплекса документов. Способы внедрения отдельных частей документа зависят от конкретного типа информации, содержащейся в документе, и отображаются на лексикологическом дереве документа.

Лексикологическое дерево, представленное на рис. 2 на примере документа «Приказ об изменении программы обучения студента», отображает, к какой категории относятся отдельные блоки информации. Например, автор документа в данном случае сопоставляется с наименованием организации, и эти сведения относятся к унифицированной постоянной информации.

Как отмечено в работе [9], документ может содержать четыре типа информации, каждый из которых имеет свои особенности и, соответственно, различается по способу внедрения в документ. К унифицированной постоянной информации относится постоянная и редко меняющаяся информация, например заголовок документа или наименование организации. Унифицированная переменная информация включает формализованные данные, которые хранятся в базе знаний. Переменная вводимая информация представляет конкретизирующие данные для конкретного экземпляра документа. К неунифицированной информации относятся свободные формулировки, вводимые в документ непосредственно с клавиатуры.

Унифицированная постоянная информация внедряется при формировании документа автоматически. Унифицированная переменная информация добавляется путем выбора из предлагаемого списка. Сведения, относящиеся к переменной вводимой и неунифицированной переменной информации, вносятся путем свободного ввода.

Рис. 1. Фрагмент лексикологической схемы организационно-распорядительной документации вуза

При формировании документа создается индексная последовательность, которая, в свою очередь, передается по каналу связи. Структура процесса фиксации индексной последовательности приведена в работе [2]. Индексная последовательность построена таким образом, чтобы без доступа к лексикологической схеме документа было невозможно восстановить его содержимое, тем самым обеспечивая полную конфиденциальность.

Индексная последовательность содержит следующие элементы:

- индекс формы (типа) документа;
- индекс заголовка документа;
- индексы опорных слов;
- зашифрованные свободные формулировки;
- индекс электронной подписи.

Защита документов, обеспечиваемая на основе применения технологии лексикологического синтеза при передаче по каналам

связи и хранении, осуществляется следующим образом. На стороне отправителя формируется информационная посылка, содержащая индексную последовательность, которая и отправляется по каналу связи. На стороне получателя после получения данной посылки происходит автоматическое восстановление документа. Данная технология работает только при наличии у обеих сторон согласованной лексикологической схемы и дерева, без которых создание и последующее восстановление документа невозможны.

Аспекты формирования индексной последовательности

Унифицированный набор документов формируется путем глубокого информационного анализа документации организации, выявления наиболее часто используемых видов документов и их унификации по форме и содержанию.

Рис. 2. Лексикологическое дерево приказа об изменении программы обучения студента

Рис. 3. Элементы технологии лексикологического синтеза

Опорные слова выделяются из унифицированных документов и используются при автоматизированном создании этих документов. Лексикологическая схема документа отображает связь опорных слов, соответствующих формулировок, а также их индексы. Лексикологическое дерево отображает способ внедрения формулировок в зависимости от класса информации, к которому они относятся [10].

Формированию индексной последовательности предшествует реализация самой технологии создания документов на основе

лексикологического синтеза. Элементы описываемой технологии представлены на рис. 3. Опорные слова определяются из унифицированного набора документов. По опорным словам и связанным формулировкам строится лексикологическая схема, на основе которой формируется лексикологическое дерево.

С целью более эффективного использования технологии защиты документов на основе лексикологического синтеза целесообразно в последующем исследовании разработать методологическую модель формирования индексной последовательности,

которая должна включать следующие компоненты:

- структура индексной последовательности;
- алгоритм формирования индексной последовательности.

Процесс формирования индексной последовательности можно представить в виде алгоритмической модели с целью последующей реализации данной технологии защиты документов.

Пример индексной последовательности

Время формирования индексной последовательности зависит от ее структуры. Унифицированный документ состоит из уровней уточнения информации, по которым и осуществляется фиксация индексной последовательности. В качестве примера индексную последовательность, формируемую по лексикологической схеме (рис. 1), можно записывать следующим способом (нумерация элементов внутри блоковых фигур означает просто номер компонентов внутри блока и не имеет отношения к формируемой индексной последовательности).

Представленная индексная последовательность сформирована на основе фрагмента лексикологической схемы организационно-распорядительной документации вуза (рис. 1).

Первому уровню присваивается индекс 1, которому соответствует вид доку-

мента «Приказ», второму уровню – индекс 2 с соответствующим заголовком «Об изменении программы обучения», третьему – код студента 048967, 2 курс, очно-заочная форма обучения с индексом 2, факультет, направление, уровень обучения. Основа обучения не выбирается, следовательно, ставится знак «^». Четвертый уровень не уточняется, указывается знак «*». Пятому уровню соответствуют индексы направления, уровня обучения и срока обучения.

Результирующая индексная последовательность имеет вид

1:1:048067-2-2-4-1-1-1-^*:1-1-406,

где уровни уточнения информации разделяются знаком «:», внутриуровневое разделение обозначается символом «-», отсутствие уточняющей информации на уровне «*», отсутствие информации внутри уровня – «^».

Свободные формулировки из документа в зашифрованном виде можно размещать в конце индексной последовательности с указанием позиции их расположения. Детальное рассмотрение процесса формирования индексной последовательности является предметом для дальнейшего исследования, к основным задачам которого следует относить сокращение размера формируемой индексной последовательности (включая фрагменты неунифицированной информации и ее размещение в информационной посылке).

ПРИКАЗ

09 марта 2017 г.

№

г. Москва

Об изменении программы обучения

ПРИКАЗЫВАЮ:

1. Перевести студентов 2 курса очно-заочной формы обучения факультета дистанционного обучения Университета, обучающихся по направлению "Экономика" (бакалавриат) на внебюджетной основе, в число студентов 1 курса очной формы обучения факультета математической экономики, статистики и информатики Университета, обучающихся на внебюджетной основе по следующему списку:

- на направление "Экономика" (бакалавриат), со сроком обучения 4 года 6 месяцев:

1.

Рис. 4. Фрагмент приказа об изменении программы обучения

Лексикологической схеме соответствует фрагмент документа, представленного на рис. 4. За черными рамками скрыта конфиденциальная информация университета, а также возможные сведения, относящиеся к категории персональных данных.

Заключение

1. Эффективность формирования индексной последовательности зависит от ее структуры и способа ее фиксации. Индексная последовательность должна содержать минимум знаков и символов, но быть читаема и восстанавливаема на стороне получателя при наличии согласованной лексикологической схемы.

2. Представлен пример формирования индексной последовательности на основе организационно-распорядительной документации вуза. При определении лексикологической схемы и дерева всего набора документации вуза технология лексикологического синтеза позволяет создавать и передавать документы в определенном формате, обеспечивающем необходимый уровень защиты передаваемых сведений от несанкционированного доступа.

3. В дальнейшем исследовании необходимо разработать методику фиксации индексной последовательности и решить следующие задачи:

- разработка структуры индексной последовательности;
- разработка алгоритма формирования индексной последовательности при защите документов на основе лексикологического синтеза;

- формирование правила работы со свободными формулировками (неунифицированной информацией) и их шифрования.

Список литературы

1. Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных» // Российская газета от 29 июля 2006 г. № 165.
2. Черников Б.В. Технология защиты информации текстовых документов // Известия Южного федерального университета. Технические науки. – 2014. – № 6 (155). – С. 173–185.
3. Конституция Российской Федерации от 12 декабря 1993 г. // Парламентская газета от 23 января 2009 г. № 4.
4. Указ Президента РФ от 6 марта 1997 г. № 188 «Об утверждении перечня сведений конфиденциального характера» // Собрание законодательства Российской Федерации. Изд-во «Юридическая литература». – 10 марта 1997. – № 10, ст. 1127.
5. Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации» // Парламентская газета от 03 августа 2006 г. № 126–127.
6. Тулупов М.В. Роль и место организационно-правовой защиты информации в обеспечении информационной безопасности предприятия // Гаудеамус. – 2014. – № 2 (24). – С. 204–206.
7. Приказ ФСТЭК России от 18 февраля 2013 г. № 21 // Российская газета от 22 мая 2013 г. № 107.
8. Баричев С.Г., Гончаров В.В., Серов Р.Е. Основы современной криптографии. – М.: «Горячая линия – Телеком», 2001. – 120 с.
9. Черников Б.В. Способ автоматизированного лексикологического синтеза документов с защищенной информацией при передаче их по каналам связи. Патент РФ № 2331104, 2008.
10. Черников Б.В. Лексикологический синтез документов в комплексах информационных систем. – М.: ИД «ФОРУМ», 2017. – 336 с.