УДК 004.056.55, 004.82 **DOI** 10.15622/sp.48.4

В.И. Воробьёв, Е.Л. Евневич, Д.К. Левоневский, Р.Р. Фаткиева, Л.Н. Федорченко

ИССЛЕДОВАНИЕ И ВЫБОР КРИПТОГРАФИЧЕСКИХ СТАНДАРТОВ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДОКУМЕНТОВ

Воробьёв В.И., Евневич Е.Л., Левоневский Д.К., Фаткиева Р.Р., Федорченко Л.Н. **Исследование и выбор криптографических стандартов на основе интеллектуального анализа документов.**

Аннотация. В данной статье исследуются проблемы применимости и выбора криптографических стандартов с учетом предпочтений и требований потенциального пользователя. Профили пользователя формируются с помощью онтологических методов. На основе профилей пользователей и характеристик документов формируется набор документов, которые могут подойти конкретному пользователю, и элементы этого набора ранжируются по вероятности соответствия его требованиям. При формировании набора документов используются различные методы фильтрации: коллаборативная фильтрация, анализ и фильтрация контента, а также гибридные методы, совмещающие оба подхода. Таким образом, создается рекомендующая система выбора криптографических стандартов и алгоритмов. При наличии нескольких пользовательских критериев выбора объекта целесообразно использовать интегральный показатель соответствия объекта, который вычисляется в виде взвешенной суммы показателей.

Ключевые слова: криптография, стандарты, рекомендующая система, алгоритмы коллаборативной фильтрации, гибридные методы, онтологии

1. Введение. Современная ситуация в области обеспечения безопасности информационных технологий (ИТ) характеризуется значительной лингвистической неоднородностью стандартов, регламентов, нормативных документов, политик и профилей безопасности, и как следствие, их слабой сбалансированностью и интегрируемостью [1, 2]. Причина прежде всего в том, что объекты защиты представляют собой сложные многоуровневые нелинейные системы с большим числом степеней свободы [3, 4].

Еще более острая проблема заключается в отображении требований каждого стандарта на общее множество требований информационной безопасности, так как некоторые из них повторяются от одного стандарта к другому или взаимно исключают друг друга. При этом встает вопрос о применимости стандартов и конкретных требований к организации в зависимости от формы собственности, сферы деятельности, использования тех или иных видов информации, наличия рисков и т.п. Так, согласно требованиям международного стандарта ISO/IEC 15408-1:2009 [5], требуется как формирование заданий по безопасности, так и построение профиля безопасности. При этом нередко возникают трудности при выборе средств обеспечения криптографической защиты ин-

формации. В частности, трудности могут возникнуть из-за ограничений, связанных с требованиями государственных стандартов при различных формах собственности защищаемого объекта. В связи с этим целесообразно построение *рекомендующей системы выбора криптографических решений*, позволяющей сформировать перечень угроз и произвести выбор на основе опроса пользователей.

2. Основные проблемы. Анализ текстов стандартов на основе онтологических моделей подтверждает их противоречивость и неполноту, что проявляется в несогласованности трактовок базовых концептов стандартов в криптографии, а также в неопределенности некоторых первичных понятий и в вытекающей отсюда неполноте или несогласованности отдельных положений и рекомендаций [1].

Представленные на рынке программные продукты (ПП), используемые для проверки соответствия политики информационной безопасности требованиям стандарта ISO 17799 – Cobra (C & A SystemsSecurityLtd,); CRAMM (CCTA Risk Analysis and Management Method); КОНДОР+ (DigitalSecurity); RiskWatch; ГРИФ 2006 (Digital Security Office); Авангард; Callio Secura(Callio Technologies); Ezrisk (Echelon Consulting); ISRAC (Infosecure Group); PTA (Practical Threat Analysis); RSAM (Relational Security); vsRisk™ (Vigilantsoftware) и др. имеют свои недостатки, например, отсутствие возможности установить весовой коэффициент каждого требования, необходимость специальной подготовки аудитора, высокая стоимость лицензии и ряд других [6, 7].

Высокая значимость обеспечения информационной безопасности, глобальный характер этой деятельности, неоднородность стандартов защиты информации обусловливают необходимость разработки средств, учитывающих особенности пользователей и организаций, сталкивающихся с потребностью в защите информации.

При выборе средств обеспечения информационной безопасности обычно принимаются во внимание следующие аспекты: аппаратная и программная платформа, на которой будет работать средство; сфера согласованности стандартов; масштабируемость; импортно-экспортные возможности; адаптируемость к структуре организации; модель лицензии; простота использования; цена.

Такая постановка задачи основана на анализе документов без учета особенностей пользователей стандартов. Эти особенности связаны с родом деятельности пользователей — физических и юридических лиц, их правовым статусом и ограничениями. Без учета пользовательского аспекта постановка задачи будет неконструктивной.

Требуется согласованная техническая политика для обеспечения определенности и однозначности понятийного аппарата и, в частности,

разработки средств формализации и соответствующей автоматизации этого процесса.

Поэтому предлагается подход на основе построения онтологической модели, описывающей отношения элементов как минимум двух множеств: субъектов (пользователей стандартов, агентов) и объектов (которые в конкретных случаях могут быть стандартами, материалами, документами, сервисами).

Подобный подход позволяет выполнить выбор того или иного документа и оценить механизмы принятия решения. Групповая оценка продукта (информационной системы) позволяет осуществить категорирование или построение рейтинговой шкалы, что облегчает работу пользователя при выборе сервиса. Фактически, пользователь осуществляет свой выбор с учетом информации, предоставленной рекомендующей системой, а эта информация, в свою очередь, зависит от требований конкретного пользователя. Рекомендующие системы также могут приводить аргументы в пользу рекомендуемых объектов, так как пользователи склонны более доверять обоснованным предложениям [8]. Рассмотрим постановку задачи более подробно.

3. Постановка задачи. Пусть существует m пользователей системы $X_1, X_2, \ldots, X_m$ и n документов $Y_1, Y_2, \ldots, Y_n$, созданы профили пользователей, документов и ограничения поиска. Необходимо для пользователя X_i сформировать множество документов, адекватных требованиям пользователя, и ранжировать их по степени релевантности [9].

В рекомендующих системах используются различные методы принятия решения — методы коллаборативной фильтрации [10, 11], гибридные методы, синтаксические методы обработки текстов.

Для сопоставления пользователей определяют метрику сходства. В коллаборативных алгоритмах используется информация о поведении субъекта в прошлом, например, об использовании конкретных ресурсов, взаимодействии с другими субъектами и об оценках рисков [12]. В этом случае не имеет значения тип объектов, но могут учитываться скрытые факторы, которые сложно было бы учесть при создании профиля. В данных методах профиль пользователя определяется множеством сервисов, которым был присвоен рейтинг этим пользователем.

Контекстно-зависимые рекомендующие системы [13] характеризуются также тем, что учитывают знания об интересах пользователя и контекст конкретной задачи выработки рекомендаций. Это выражается, в частности, в том, что такие системы учитывают атрибуты времени, места, социальный контекст, работают с группами пользователей, используют теги для уточнения свойств пользователей и сервисов.

Развитием этого подхода является гибридный метод, который использует как метрики сходства, так и статистические методы и позволяет добиться более точных и обоснованных предположений.

4. Гибридный метод фильтрации как основа рекомендующей системы. Подход базируется на преимуществах фильтрации контента и коллаборативной фильтрации.

В случае коллаборативной фильтрации предполагается, что похожим пользователям нравятся похожие образцы. При поступлении запроса от пользователя U_i вычисляются метрики сходства u_{ij} для всех $j \neq i$ и формируется множество из k пользователей, наиболее сходных с U_i. Далее формируется множество документов, подходящих этим пользователям, и из него исключаются документы, уже известные U_i. Здесь используется только профиль пользователя. Метрики сходства двух пользователей можно рассчитать, используя пересечения множеств понравившихся документов и интересов пользователей. При этом менее популярным документам можно присваивать более высокий весовой коэффициент. Для принятия решения используется симметричная матрица отношений «пользователь-пользователь»:

$$U = \begin{pmatrix} u_{11} \ldots u_{1m} \\ \ldots\ldots\ldots\ldots \\ u_{m1} \ldots u_{mm} \end{pmatrix}.$$

Фильтрация контента используется для оценки сходства документов. Этот подход предполагает, что если пользователю U_i нравится документ D_j, который похож на документ D_k, то пользователю U_i понравится и D_k. В этом случае, напротив, используется профиль документа. Метрики сходства двух документов d_{ij} можно рассчитать, используя пересечения множеств ключевых слов этих документов и пользователей, выставивших документам близкие оценки. Матрица отношений имеет вид:

$$D = \begin{pmatrix} d_{11} \ldots d_{1n} \\ \ldots\ldots\ldots\ldots \\ d_{n1} \ldots d_{nn} \end{pmatrix}.$$

Гибридный подход предполагает построение матрицы «пользователь-документ», содержащей оценки r_{ij}, данные пользователями с индексами $i = 1 \ldots m$ документам с индексами $j = 1 \ldots n$:

$$R = \begin{pmatrix} r_{11} \ldots r_{1n} \\ \ldots\ldots\ldots \\ r_{m1} \ldots r_{mn} \end{pmatrix}.$$

Если учитываются только непосредственные связи между пользователями и документами, матрица R будет очень разреженной, так как пользователь не имеет возможности оценить все или хотя бы значительную часть документов. Неопределенные элементы матрицы можно заполнить, рассчитав псевдорейтинг — величину, которая является предполагаемой оценкой пользователем U_i документа D_j с учетом похожести документов:

$$r_{ij} = \sum_{k=1}^{n} r_{ik} d_{ij},$$

где r_{ik} — известные значения матрицы «пользователь-документ», d_{ij} — метрики сходства документов D_i и D_j.

Сходство пользователей можно определить, используя корреляционную меру, основанную на скалярном произведении векторов:

$$w = \frac{\vec{v}_i \vec{v}_j}{|\, v_i \,||\, v_j \,|},$$

где w представляет собой косинус угла между векторами предпочтений v_i и v_j пользователей u_i и u_j и лежит в пределах от -1 до 1, где -1 соответствует полной противоположности (противонаправленность векторов), 1 — полному сходству (сонаправленность), 0 — отсутствию корреляции (ортогональность).

Предпочтения пользователя u_{ij} можно определить как:

$$u_{ij} = \frac{\sum_{k=1}^{n} (r_{ik} - \overline{r}_i)(r_{jk} - \overline{r}_j)}{\sqrt{\sum_{k=1}^{n} (r_{ik} - \overline{r}_i)^2 \sum_{k=1}^{n} (r_{jk} - \overline{r}_j)^2}},$$

где

$$\overline{r}_i = \frac{1}{n} \sum_{j=1}^{n} r_{ij}.$$

Здесь метрика сходства u_{ij} определяется как корреляция предпочтений пользователей U_i и U_j и является нормированным скалярным произведением векторов предпочтений U_i и U_j в n-мерном пространстве, где каждое измерение соответствует документу, а соответствующая координата вектора — отношению пользователя к этому документу.

Предположение об отношении пользователя U_i к документу D_j выражается величиной:

$$p_{ij} = r_i + \frac{\sum_{k=1}^{m}(r_{kj} - \overline{r}_k)u_{ik}}{\sum_{k=1}^{m}(r_{kj} - \overline{r}_k)}.$$

Для инициализации предполагается определение начальных значений всех элементов матрицы D и по возможности большего количества элементов матрицы R.

Сходство документов d_{ij} предлагается определять путем сопоставления профилей документов. Для этого документам присваиваются атрибуты A_i, определяемые исходя из метаданных, ключевых слов, содержания документа и статистической информации. Каждый атрибут имеет идентификатор (уникальное число или строка). В простейшем случае атрибут может иметь два значения: 1 (true, присвоен) и 0 (false, не присвоен, значение по умолчанию). Тогда наличие атрибута соответствует характеризующему документ утверждению, например:

– документ относится к предметной области пользователя X;

– в документе упоминается X;

– документ положительно/отрицательно оценен пользователем X.

Профиль документа можно определить как множество присвоенных ему атрибутов и их значений, а расстояние между профилями:

$$d(D_i, D_j) = \sum_{k=1}^{l} \alpha(A_k) d(A_k(D_i), A_k(D_j)),$$

где $\alpha(A_k)$ — весовой коэффициент атрибута.

На практике нет необходимости обрабатывать сильно отдаленные документы, поэтому расчет $d(D_i, D_j)$ целесообразно производить для ограниченного числа имеющих наибольший вес атрибутов с расчетом, чтобы выполнялось неравенство:

$$\sum_{k=1}^{l} \alpha(A_k) \le N,$$

где N — константа, ограничивающая точность оценки. В этом случае коэффициент сходства документов можно определить как:

$$d_{ij} = \frac{N}{d(D_i, D_j)}.$$

Применение методов фильтрации позволяет генерировать рекомендации на основании сходства пользователей и документов и с учетом интересов пользователей и их оценок. Следующим этапом развития системы является использование методов контекстного анализа в процессе выработки предложений.

5. Методы учета контекста. Большинство рекомендующих систем опираются на предложение наиболее релевантных образцов конкретным пользователям, при этом контекст, в котором находятся пользователи и сервисы, не принимается во внимание. Такой подход не является достаточным для поставленной задачи.

В рекомендующих системах первого поколения используется оценка функции рейтинга: $R : User \times Item \rightarrow Rating$.

Подобные системы называются также традиционными или двумерными [14], так как учитывают только два измерения — домены пользователей (User) и сервисов (Item). Учет контекста предполагает ввод дополнительных измерений в формулу рейтинга. Это означает, что оценка релевантности сервиса раскладывается по оценкам этого же сервиса в различных контекстах:

$$p_{ij} = \sum_{c=1}^{N_C} p_{ijc}.$$

Рассмотрим применение контекста к решаемой задаче. Пусть имеются отношения:

– *Пользователь* (ID, Имя, Расположение, Организация, Интересы, Профессия);

– *Документ* (ID, Название, Автор, Область, Тип, Время публикации).

Контекст может состоять из нескольких типов, каждый из которых определяет один аспект контекста — временное или пространственное расположение, социальную вовлеченность, цель использования. Для данной задачи можно ввести следующие *типы контекста:*

1. Время — интервал актуальности документа или этап жизненного цикла проекта.

2. Место — атрибут может включать следующие компоненты: *организация* (адрес); *район; населенный пункт*; *область; государство*.

3. Цель — принимает следующие значения*: образовательная; академическая; коммерческая; профессиональная*.

Контекст может иметь сложную и разнообразную структуру. К примеру, информация о пространственном расположении может пред-

ставляться в виде иерархии: *адрес → район → населенный пункт → область → государство.*

Временные атрибуты организуются в иерархии вида: *подэтап жизненного цикла → этап жизненного цикла → степень реализации* или *число → месяц → квартал → год.*

Применительно к построению структуры атрибутов контекста можно выделить два подхода:

– *предметный,* когда множество допустимых значений атрибута и отношений между ними предопределено и не испытывает значительных изменений во времени;

– *интерактивный,* когда предполагается двунаправленное взаимодействие между активностью пользователя и структурой атрибута: контекст влияет на рекомендации, а действия пользователя влияют на структуру контекста.

Для каждого типа контекстных данных определяется одно измерение C_i, каждое измерение характеризуется множеством допустимых значений, упорядоченных в k уровней. Меньшие значения уровня соответствуют меньшей точности и большей общности контекстной информации, а большие значения, близкие к k, определяют контекст наиболее конкретно. Тогда функция рейтинга имеет вид: $R: D_1 \times D_2 \times ... \times D_N \rightarrow Rating$. В этом случае рейтинг представляет собой целую или вещественную функцию, определенную на N-мерном пространстве дискретных значений. Два измерения соответствуют пользователям и сервисам, оставшиеся — типам контекста. Рисунок 1 иллюстрирует функцию рейтинга как гиперкуб.

В этом случае хранилище данных для значений рейтинга удобно организовать с помощью технологии OLAP (Online Analytical Processing) [15]. Рабочие данные представляются в структуре, называемой «OLAP-куб», и аналогичной рисунку 1. OLAP-куб содержит базовые данные и информацию об измерениях (агрегаты). Куб может содержать всю информацию, которая необходима для ответов на любые запросы. При большом количестве агрегатов полный расчет зачастую происходит только для отдельных измерений, для остальных расчет выполняется по требованию.

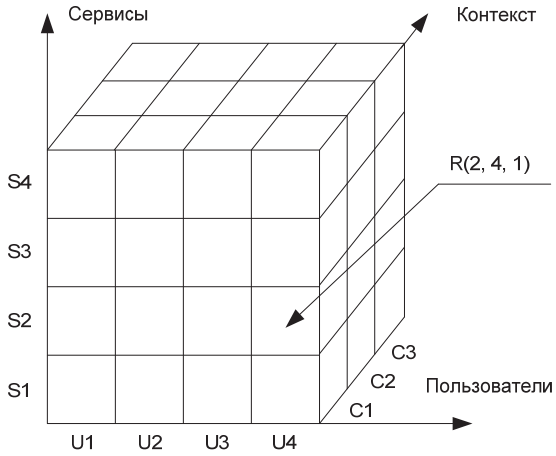

Рис. 1. Иллюстрация функции контекста

Для описания предметной области и формирования структуры контекста применяется инженерия онтологий [16, 17].

6. Инженерия онтологий. Определяется как совокупность процесса разработки онтологий; жизненного цикла онтологий; методов и методологий построения онтологий; набора инструментов и языков для построения и документирования онтологий, импорта и экспорта онтологий разных форматов и языков, поддержки графического редактирования, управления библиотеками онтологий и т.д.

Процесс онтологического моделирования можно разделить на выделение классов (концептов) и их свойств (отношений-слотов). Классы разрабатываемой онтологии описывают понятия предметной области. Каждый из них может иметь свой подкласс, который изображает более подробное описание, чем его надкласс. Задача слота — описать свойства класса и экземпляра. Свойства дают возможность, утверждать общие факты о членах классов. Свойство — это бинарное отношение. Различают два типа свойств:

1. Свойства-значения, отношения между представителями классов и типами данных.
2. Свойства-объекты, отношения между представителями двух классов.

При разработке профиля пользователя предлагается использовать анкетные данные и сведения о его активности. Анкетные данные — круг интересов пользователя, сведения о его образовании, навыках, профессиональной деятельности. Сведения об активности пользователя осно-

вываются на таких его действиях в информационной системе, как: просмотр документа, включая частоту доступа, полноту просмотра и время нахождения на странице; оценку документа; добавление документа в избранное; комментирование документа; создание документа.

Процедуру построения профиля пользователя можно разделить на три этапа:

Этап I. Сбор первоначальной информации о конкретном пользователе. На данном этапе осуществляется сбор информации, поступающей при заполнении анкеты пользователя. Это позволяет осуществить статистическую обработку атрибутов профилей пользователя, а также сформировать группы пользователей по интересам или другим атрибутам, входящим в профиль.

Этап II. Анализ запрашиваемой пользователем информации. Данные о запросах пользователя сохраняются для дальнейшего анализа. Это позволяет осуществить группировку часто встречающихся запросов, осуществить кластеризацию запросов, а также спрогнозировать вероятные запросы пользователя. Использование кластеризации данных позволяет осуществить категорирование пользователей, в том числе, разделение на «новичков в предметной области», специалистов и экспертов.

Такой подход, в свою очередь, позволяет осуществлять адресацию вопросов от новичков к экспертам, выполнять поиск необходимого пользователя, группировать пользователей по командам.

Этап III. Построение онтологии предметной области.

Для построения профиля пользователя была разработана анкета пользователя, а также выделены и описаны основные атрибуты, которые прослеживаются в поведении пользователей на инновационном портале. К базовым атрибутам, не зависящим от особенностей ресурсов информационной системы, относятся: длительность посещения ресурса; длительность посещения страницы; частота посещения ресурса; частота отправления запроса.

7. Синтаксический аспект онтологического моделирования. Переход от текстов стандарта к формализованному онтологическому описанию состоит в определении онтологической тройки:

– множество концепций (терминов);

– множество отношений между концепциями;

– правила логического вывода в сети концепций и отношений (например, правило транзитивности, симметричности, антисимметричности, рефлексивности).

При онтологическом моделировании не обойтись без инструментальной поддержки, обеспечивающей пользователей и лиц, принимающих решения, средствами работы с данными и системами

компьютерной поддержки доказательства утверждений, позволяющими создавать унифицированные программы проверки свойств и проверки доказательств того, что программа соответствует своей спецификации.

В качестве перспективной формальной базы для использования таких систем в СПИИРАН разрабатывается подход, основанный на использовании грамматик и синтаксических методов для представления схем логических выводов и вычислительных схем.

Современные системы компьютерной поддержки доказательства утверждений используют расширяемые системы правил, состоящие из двух частей. Одна часть содержит некоторое постоянное логическое ядро, другая — состоит из правил, задаваемых пользователем. Поскольку набор и состав этих правил может изменяться, необходимы программные средства, позволяющие быстро настраиваться на вводимые изменения и дополнения в схемы правил.

Процесс логического вывода удается разбить на два этапа, при этом первый этап (вывод схем правил и схем выводов) приводит к регулярным выражениям, что допускает эффективное использование специального инструментального средства, упрощающего схемы правил и схемы выводов (например, удаление тупиковых и циклических выводов). Второй этап требует использования грамматик, содержащих контекстно-зависимые правила с атрибутами в виде семантик и предикатов, которые также могут быть проанализированы инструментальной системой. При реализации прототипа системы с атрибутами на платформе .NET в качестве основы взят код инструментального комплекса SynGT, разработанного в СПИИРАН [21].

8. Методика выбора алгоритмов. В условиях многокритериального выбора целесообразно применить метод свертки взвешенных показателей (взвешенного среднего арифметического). Применение этого метода для сравнения криптографических решений предложено в работе [18]. Преимущество этого метода в том, что более высокие оценки получают те решения, которые имеют больше критериев с максимальной степенью соответствия. В качестве критериев рассмотрены безопасность, скорость и стоимость, весовые коэффициенты полагаются одинаковыми и равными 1/3.

Затем применяется метод взвешенной метрики. Вычисляется отклонение (дисперсия) решений по отношению к идеальному.

В отличие от данного подхода, в [19, 20] показано, что при выборе функции распределения, описывающей сложные критические объекты, применяется степенная функция распределения. Предлагается следующий способ моделирования неопределенности задания нормирующей функции:

$$F(z,t)=\begin{cases}0, z \le z_{-}\\ \left(\dfrac{z-z_{-}}{z_{+}-z_{-}}\right)^{t}, z_{-} < z \le z_{+}\\ 1, z_{+} < z\end{cases}.$$

В результате появляется возможность построения вектора показателей качества объекта в виде $q=(q_1,...,q_n)$, i=1, …, l для вектора исходных характеристик $x = (x_1, …, x_n)$. Например, для оценки качества применяется набор показателей q в зависимости от характеристик x. После получения набора отдельных показателей выбирается синтезирующая функция $Q(q)=Q(q;w)$, где $w = (w_1, …, w_l)$, $w_1+…+w_l=1$, интерпретируются как весовые коэффициенты, задающие степень влияния отдельных показателей на сводную оценку — в нашем примере усредненное значение показателя качества документа.

При практическом использовании сводных показателей зачастую имеет место дефицит информации, выражающийся в том, что имеется неопределенность выбора функций q, Q и вектора w. Данная неопределенность усугубляется еще тем, что доступная информация не имеет числового характера, то есть квалиметрическая шкала имеет более бедную структуру, чем обычная числовая шкала. Например, показатели качества объекта защиты в зависимости от его месторасположения. В этом случае задача оцифровки состоит в выборе отображения $\varphi(b)$, где b — качественная характеристика (например, баллы).

9. Реализация рекомендующей системы. Рекомендующая система строится на основе информационной системы, оперирующей объектами как самодостаточными сущностями. Классы, определяющие объекты, упорядочены в иерархическую структуру (рисунок 2).

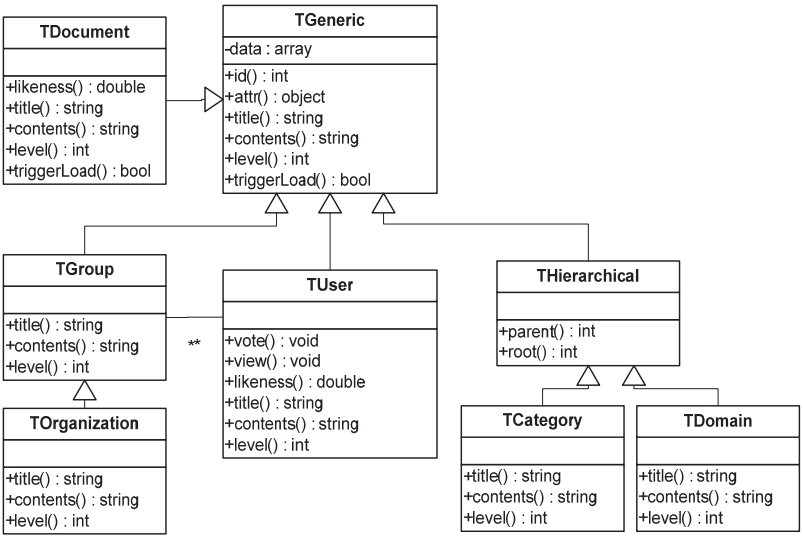

Рис. 2. Фрагмент иерархии классов

Все классы прямо или опосредованно наследуют базовому типу TGeneric. Благодаря этому к основным, не зависящим от спыецифики типа свойствам можно обращаться через унифицированный интерфейс.

Иерархия включает следующие типы, но не ограничивается ими:

– TGeneric — базовый абстрактный класс;

– TUser — пользователь;

– TGroup — группа пользователей;

– TOrganization — организация;

– TDocument — документ;

– THierarchical — абстрактный класс для создания объектов, упорядоченных иерархически;

– TCategory — тип документа;

– TDomain — предметная область документа.

К основным операциям с объектами относятся:

– id — получение идентификатора объекта (числа, уникального в пределах типа);

– attr — получение и установка атрибутов;

– title — получение названия объекта;

– contents – получение HTML-представления объекта;

– level — уровень привилегий, необходимый для доступа к объекту.

Материалы организованы в динамической иерархической структуре (рисунок 3). Они включают:

– публикации пользователей;
– сведения об общественной деятельности;
– служебные и справочные материалы.

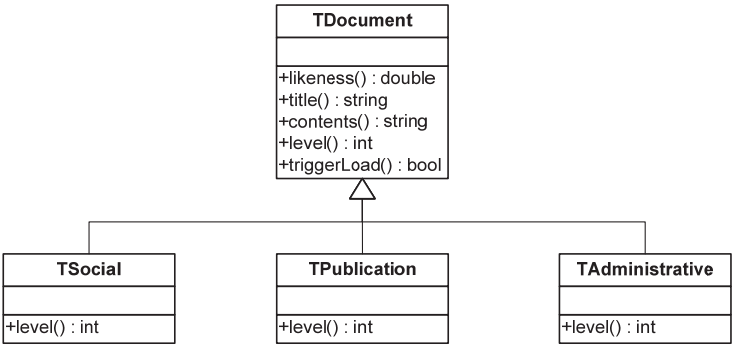

Рис. 3. Организация материалов

Зарегистрированный пользователь системы имеет логин и пароль, которые он сообщает системе в процессе регистрации. Эти данные используются в дальнейшем при входе в систему и удостоверяют его права на доступ к информации и публикацию материалов. Неавторизованные пользователи называются гостями и имеют ограниченный доступ. Пользователям присваивается уровень доступа, который может определять их полномочия в широких пределах.

10. Заключение. Выработка рекомендаций к применению того или иного криптографического стандарта или ресурса информационной системы с учетом индивидуальных особенностей каждого пользователя требует использования статистических методов оценки деятельности пользователя в этой системе, методов обработки неструктурированных данных и контекстного анализа. На этих методах основывается предложенный подход к построению групповой контекстно-зависимой рекомендующей системы. Подход позволяет учитывать функциональные и правовые ограничения субъектов, использующих нормативные документы в области информационной безопасности.

В качестве основного направления будущих исследований предлагается разработка Web-ресурса для гармонизации стандартов и регламентирующих документов на основе их онтологического описания, которое включает построение таксономий терминов предметной области, предикативных отношения между терминами, логический вывод на основе дескриптивной логики, а также средства интеграции онтологий и рекомендующих систем для учета требований пользовfte-

ля. Процессом достижения цели будет построение терминологических таксономий, предикативных отношений, дескриптивных ограничений (ontological restrictions) для вывода новых типов данных и расширение множества правил вывода. Уточнение терминологии и отношений может происходить непосредственно в процессе разработки онтологий (серии онтологий). Все методики предполагается применять для проверки корректности использования криптостандартов в области облачных вычислений.

Литература

1. *Atiskov A.Yu., Vorobev V.I., Fedorchenko L.N. et al.* Theory and Practice of Cryptography Solutions for Secure Information Systems // IGI Global. 2012. pp.101–130.
2. *Sennewald C., Baillie C.* International Security Standards // Effective Security Management (Sixth Edition). 2016. pp. 205–212.
3. *Воробьев В.И., Фаткиева Р.Р.* Природа уязвимостей программного кода // Программируемые инфокоммуникационные технологии. Сборник статей. М.: Радиотехника. 2009. С. 53–55.
4. *Баранов С.Н., Шишкин В.М.* Современные тенденции индустрии разработки программных продуктов // Информационно-измерительные и управляющие системы. 2012. Т. 10. № 5. С. 24–33.
5. ISO/IEC 15408-1:2009: Information technology — Security techniques — Evaluation criteria for IT security — Part 1: Introduction and general model. 2009.
6. Официальный сайт компании «Аудит информационной безопасности». URL: http://www.audit-ib.ru/ (дата обращения: 10.04.2016).
7. IT Governance Green Paper INFORMATION SECURITY & ISO 27001. URL: http://www.itgovernance.co.uk/files/Infosec_101v1.1.pdf (дата обращения: 13.05.2016).
8. Перспективные направления развития науки в Петербурге / Отв. ред. Ж.И. Алферов, О.В. Белый, Г.В. Двас, Е.А. Иванова // СПб.: Из-во ИП Пермяков С.А. 2015. 543 с.
9. *Городецкий В.И., Тушканова О.Н.* Онтологии и персонификация профиля пользователя в рекомендующих системах третьего поколения // Онтологии проектирования. 2014. № 3 (13). С. 7–31.
10. *Wang J., Pouwelse J.* Distributed Collaborative Filtering for Peer-to-Peer File Sharing Systems // Proceedings of the 2006 ACM symposium on Applied computing. pp. 1026–1030.
11. *Melville P., Mooney R.J., Nagarajan R.* Content-Boosted Collaborative Filtering for Improved Recommendations // Proceedings of 18th National ACM Conference of Artificial Intelligence. 2002. pp. 187–192.
12. *Королёва Д.Е., Филиппов М.В.* Анализ алгоритмов обучения коллаборативных рекомендательных систем // Инженерный журнал: наука и инновации. 2013. Вып. 6. URL: http://engjournal.ru/catalog/it/hidden/816.html (дата обращения: 29.02.2016).
13. *Adomavicius G., Mobasher B., Ricci F., Tuzhilin A.* Context-Aware Recommender Systems // AI Magazine. 2011. pp. 67–80.
14. *Wang J., de Vries A. P., Reinders M.J.T.* Unifying user-based and item-based collaborative filtering approaches by similarity fusion // Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM New York. NY. USA. 2006. pp. 501–508.
15. *Cios K.J. et al.* Data Mining: A Knowledge Discovery Approach // Springer. 2007. 606 p.
16. *Gonzalez-Perez C. et al.* An Ontology for ISO software engineering standards: 2) Proof of concept and application // Computer Standards & Interfaces. 2016. vol. 48. pp. 112–123.
17. *Van Ruijven L.C.* Ontology for Systems Engineering // Procedia Computer Science. 2013. vol. 16. pp. 383–392.

18. *Raissi J.* Dynamic Selection of Optimal Cryptographic Algorithms in a Runtime Environment // Proceedings of IEEE International Conference on Evolutionary Computation. 2006. pp. 184–191.
19. *Хованов Н.В.* Оценка сложных объектов в условиях дефицита информации. К столетию метода сводных показателей А.Н. Крылова // Моделирование и анализ безопасности и риска в сложных системах: Сб. научн. трудов 8-й международной научной школы. СПб.: ИПМАШ РАН. 2008. С. 18–28.
20. *Yudaeva M., Hovanov N., Kolesov D.* Double randomized estimation of Russian "Blue Chips" based on imprecise information // Advances in Intelligent Systems and Computing. Springer International Publishing Switzerland. 2014. vol. 299. pp. 155–164.
21. *Fedorchenko L., Baranov S.* Equivalent Transformations and Regularization in Context–Free Grammars / Bulgarian Academy of Sciences // Cybernetics and Information Technologies (CIT). Sofia. 2014. vol. 14. no 4. pp. 29–44.

Воробьев Владимир Иванович — д-р техн. наук, профессор, главный научный сотрудник лаборатории информационно-вычислительных систем и технологий программирования, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: информационная безопасность, облачные и параллельные вычисления. Число научных публикаций — 110. vvi@iias.spb.su; 14-я линия В.О., д. 39, Санкт-Петербург, 199178; р.т.: +7(812)3284369, Факс: +7(812)3284450.

Евневич Елена Людвиговна — к-т физ.-мат. наук, старший научный сотрудник лаборатории информационно-вычислительных систем и технологий программирования, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: распределенные и облачные вычисления, когнитивные технологии, информационная безопасность. Число научных публикаций — 50. eva@iias.spb.su; 14-я линия В.О., д. 39, Санкт-Петербург, 199178; р.т.: 8(812)3284369, Факс: 8(812)3284450.

Левоневский Дмитрий Константинович — научный сотрудник лаборатории информационно-вычислительных систем и технологии программирования, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: информационная безопасность, защита информации, компьютерные сети, моделирование компьютерных процессов, технологии программирования. Число научных публикаций — 15. dl@iias.spb.su; 14-я линия В.О., д. 39, Санкт-Петербург, 199178; р.т.: +7-812-328-43-69, Факс: +7 (812)350-1113.

Фаткиева Роза Равильевна — к-т техн. наук, доцент, старший научный сотрудник лаборатории информационно-вычислительных систем и технологии программирования, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: моделирование информационных систем. Число научных публикаций — 35. rrf@iias.spb.su; 14-я линия В.О., д. 39, Санкт-Петербург, 199178; р.т.: +7-812-328-43-69, Факс: +7 (812)350-1113.

Федорченко Людмила Николаевна — к-т техн. наук, старший научный сотрудник лаборатории прикладной информатики и проблем информатизации общества, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: теория формальных языков и трансляций, регуляризация грамматик, технология разработки трансляторов. Число научных публикаций — 70. lnf@iias.spb.su; 14-я линия В.О., д. 39, Санкт-Петербург, 199178; р.т.: +7(812)3281919, Факс: +7(812)3284450.

V.I. VOROBIEV, E.L. EVNEVICH, D.K. LEVONEVSKIY, R.R. FATKIEVA, L.N. FEDORCHENKO

A STUDY AND SELECTION OF CRYPTOGRAPHIC STANDARDS ON THE BASIS OF TEXT MINING

Vorobiev V.I., Evnevich E.L., Levonevskiy D.K., Fatkieva R.R., Fedorchenko L.N. **A Study and Selection of Cryptographic Standards on the basis of Text Mining.**

Abstract. This paper discusses the problems of application and choice of cryptographic standards taking into account user requirements and preferences. User profiles are created by means of the ontology apparatus. On the basis of user profiles and document features an appropriate set of documents is formed, the elements of which are then arranged according to the degree of compliance to user requirements. Various filtration methods, such as collaborative filtering, content analysis and filtering, as well as hybrid methods combining both approaches, are used. Thus, a recommender system for choosing cryptographic standards and algorithms is built. If there are several user selection criteria, it is reasonable to apply an integral index of object's relevance to user preferences. This index is defined as the weighed sum of the particular indices.

Keywords: cryptography, standards, recommender system, collaborative filtration algorithms, hybrid methods, ontologies

Vorobiev Vladimir Ivanovich — Ph.D., Dr. Sci., professor, chief researcher of computing & information systems and programming technologies laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: information security, distributed and cloud computations. The number of publications — 110. vvi@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)3284369, Fax: +7(812)3284450.

Evnevich Elena Lyudvigovna — Ph.D., senior researcher of computing & information systems and programming technologies laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: distributed and cloud computations, network security, cognitive technologies. The number of publications — 50. eva@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: 8(812)3284369, Fax: 8(812)3284450.

Levonevskiy Dmitriy Konstantinovich — researcher of computing & information systems and programming technologies laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: information security, computer security, computer networks, modeling of information processes, programming technology. The number of publications — 15. dl@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-812-328-43-69, Fax: +7 (812)350-1113.

Fatkieva Roza Ravilievna — Ph.D., associate professor, senior researcher of computing & information systems and programming technologies laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: modeling of information systems. The number of publications — 35. rrf@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-812-328-43-69, Fax: +7(812)3501113.

Fedorchenko Ludmila Nickolayevna — Ph.D., senior researcher of applied informatics and society informatization problems laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: theory of formal lan-

guages and translations, regularization of grammars, development of compilers. The number of publications — 70. lnf@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)3281919, Fax: +7(812)3284450.

References

1. Atiskov A.Yu., Vorobev V.I., Fedorchenko L.N. et al. Theory and Practice of Cryptography Solutions for Secure Information Systems. IGI Global. 2012. pp.101–130.
2. Sennewald C., Baillie C. International Security Standards. Effective Security Management (Sixth Edition). 2016. pp. 205–212.
3. Vorobiev V.I., Fatkieva R.R. [Nature of program code vulnerabilities]. *Programmiruemye infokommunikacionnye tehnologii. Sbornik statej* [Programmed information and communication technologies. Collected papers]. M.: Radiotehnika, 2009. pp. 53–55. (In Russ.).
4. Baranov S.N., Shishkin V.M. [The Actual Trends of the Software Industry]. *Informatsionno-izmeritelnye i upravlyayushchie sistemy – Measurement and control systems*. 2012. vol. 10. no. 5. pp. 24–33. (In Russ.).
5. ISO/IEC 15408-1:2009: Information technology — Security techniques — Evaluation criteria for IT security — Part 1: Introduction and general model. 2009.
6. Oficial'nyj sajt kompanii Audit informatsionnoy bezopasnosti [Official web site of the Information security audit company]. Available at: http://www.audit-ib.ru/ (accessed 10.04.2016). (In Russ.).
7. IT Governance Green Paper INFORMATION SECURITY & ISO 27001. Available at: http://www.itgovernance.co.uk/files/Infosec_101v1.1.pdf (accessed 13.05.2016).
8. *Perspektivnyie napravleniya razvitiya nauki v Peterburge. Otv. red. Zh.I. Alferov, O.V. Belyj, G.V. Dvas, E.A. Ivanova* [Actual trends of science development in St. Petersburg. Edited by Zh.I. Alferov, O.V. Belyiy, G.V. Dvas, E.A. Ivanova]. SPb.: Iz-vo IP Permyakov S.A. 2015. 543 p. (In Russ.).
9. Gorodetsky V.I., Tushkanova O.N. [Ontology-based user profile personification in 3g recommender systems]. *Ontologiya proektirovaniya – Ontology of design*. 2014. vol. 3(13). pp. 7–31. (In Russ.).
10. Wang J., Pouwelse J. Distributed Collaborative Filtering for Peer-to-Peer File Sharing Systems. Proceedings of the 2006 ACM symposium on Applied computing. pp. 1026–1030.
11. Melville P., Mooney R.J., Nagarajan R. Content-Boosted Collaborative Filtering for Improved Recommendations. Proceedings of 18th National ACM Conference of Artificial Intelligence. 2002. pp. 187–192.
12. Korolyova D.E., Filippov M.V. [Analysis of collaborative recommender system learning algorithms] *Inzhenerny zhurnal: nauka i innovatsii – Engineering journal: science and innovation*. 2013. vol. 6. Available at: http://engjournal.ru/catalog/it/hidden/816.html (дата обращения: 29.02.2016). (In Russ.).
13. Adomavicius G., Mobasher B., Ricci F., Tuzhilin A. Context-Aware Recommender Systems. *AI Magazine*. 2011. pp. 67–80.
14. Wang J., de Vries A. P., Reinders M.J.T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM New York. NY. USA. 2006. pp. 501–508.
15. Cios K.J. et al. Data Mining: A Knowledge Discovery Approach. Springer. 2007. 606 p.
16. Gonzalez-Perez C. et al. An Ontology for ISO software engineering standards: 2) Proof of concept and application. *Computer Standards & Interfaces*. 2016. vol. 48. pp. 112–123.

17. Van Ruijven L.C. Ontology for Systems Engineering. Procedia Computer Science. 2013. vol. 16. pp. 383–392.
18. Raissi J. Dynamic Selection of Optimal Cryptographic Algorithms in a Runtime Environment. Proceedings of IEEE International Conference on Evolutionary Computation. 2006. pp. 184–191.
19. Hovanov N.V. [Complex objects estimation in conditions of lack of information] *Trudyi 8-y mezhdunarodnoy nauchnoy shkolyi «Modelirovanie i analiz bezopasnosti i riska v slozhnyih sistemah»* [Proceedings of the 8th international scientific school "Modeling and analysis of security and risks in complex systems"]. SPb.: IPMASh-HRAN, 2008. pp. 18–28. (In Russ.).
20. Yudaeva M., Hovanov N., Kolesov D. Double randomized estimation of Russian "Blue Chips" based on imprecise information. Advances in Intelligent Systems and Computing. Springer International Publishing Switzerland. 2014. vol. 299. pp. 155–164.
21. Fedorchenko L., Baranov S. Equivalent Transformations and Regularization in Context–Free Grammars. Bulgarian Academy of Sciences. Cybernetics and Information Technologies (CIT). Sofia. 2014. vol. 14. no 4. pp. 29–44.