

по первому основанию значение $A^*(z) = (0, 1, z, z^3 + z^2 + z + 1, z^3 + z^2)$, а соответствующий интервальный полином равен

$$S_{int}^*(z) = z^{10} + z^9 + z^8 + z^7 + z^6 + 1.$$

Подставив в последнее равенство значения $z = \beta_4^1$ и $z = \beta_5^1$, получаем соответственно $S_{int}^4(z) = 1000 = \beta_4^3$, $S_{int}^5(z) = 1110 = \beta_5^{11}$. Полученный результат свидетельствует о том, что элементы $z = \beta_4^1$ и $z = \beta_5^1$, которые являются корнями соответствующих контрольных полиномов $p_4(z) = z^4 + z^3 + 1$ и $p_5(z) = z^4 + z + 1$, не обратили в нуль значения $S_{int}(z)$. Следовательно, нарушение условия (4) свидетельствует о наличии ошибки в $A^*(z)$, т.е. $A^*(z) \notin P_{раб}(z)$.

Обобщая вышесказанное, можно сделать вывод – полиномиальная система класса вычетов, обеспечивая параллельно-конвейерную обработку сигналов, позволяет обнаруживать и исправлять ошибки.

Библиографический список

1. Макеллан Дж., Рейдер Ч.М. Применение теории чисел в цифровой обработке сигналов: Пер. с англ./ Под ред. Манина Ю.И. М.: Радио и связь, 1983. 264 с.
2. Вариченко Л.В., Лабунец В.Г., Раков М.А. Абстрактные алгебраические системы и цифровая обработка сигналов. Киев: Наук. думка, 1986. 248 с.
3. Акуцкий И.Я., Юдицкий Д.И. Машинная арифметика в остаточных классах. М.: Сов. радио, 1968. 440 с.

О.Е. Буцкий

Россия, г. Москва, МИФИ

МЕТОД ОЦЕНКИ СЕМАНТИЧЕСКОЙ ВЗАИМОСВЯЗИ ТЕКСТОВОГО СООБЩЕНИЯ С ЗАДАННОЙ ПРЕДМЕТНОЙ ОБЛАСТЬЮ В КОНТЕКСТЕ ОЦЕНКИ СТЕПЕНИ ЕГО ПОТЕНЦИАЛЬНОЙ ОПАСНОСТИ

В настоящей работе рассматриваются вопросы разработки методов оценки семантической взаимосвязи текстового фрагмента с заданной предметной областью применительно к решению задач выявления потенциально-опасных сообщений, циркулирующих как внутри информационных систем, так и передаваемых в текстовом виде по различным каналам связи.

В последнее время в научной литературе всё чаще встречается понятие «внутренний нарушитель». Внутренний нарушитель – лицо, являющееся легитимным пользователем ИС, использующее свои служебные полномочия для проведения несанкционированных операций над информацией или техническими средствами ИС, приводящим к нежелательным для пользователя или оператора ИС последствиям. Одним из важнейших аспектов деятельности внутреннего нарушителя является нарушение конфиденциальности информации, хранящейся в ИС, посредством передачи её заинтересованному третьему лицу по каналам электронной почты или посредством умышленного понижения степени конфиденциальности. Дополнительную возможность такого несанкционированного распространения информации дает неумышленный сбор авторизованным пользователем информации, имеющей ограниченно конфиденциальный или даже открытый характер, которая при соответствующей обработке и анализе приводила к сведениям явно закрытого характера. Массивы информации, циркулирующие в ИС, могут быть огромны, поэтому мониторинг этих массивов должен быть по возможности автоматизирован. Настоящая работа посвящена именно проблемам адекватного семантического анализа текстов на естественном языке (русском) в контексте по-

строения системы мониторинга информационных массивов, циркулирующих в ИС. Существующие программные продукты, например программный продукт TextAnalyst НПИЦ "Микросистемы" (**Ошибка! Источник ссылки не найден.**, **Ошибка! Источник ссылки не найден.**), предназначенный для автоматизации анализа статичных неструктурированных текстовых данных небольшого объема (порядка мегабайта), продукт компании Инфосистемы Джет «Дозор Джет», поисковые серверы и др. используют в основном статистические методы обработки информации. Принципы их работы состоят в составлении терминологических портретов тематических рубрик и ассоциативных или семантических сетей, связи информационных признаков в которых являются исключительно результатом статистической обработки или поверхностного лингвистического анализа.

Научная новизна работы заключается в подходе к анализу степени принадлежности текстового фрагмента некоторой предметной области. Данный подход сочетает в себе не только статистическую обработку весовых коэффициентов информационных признаков обрабатываемого фрагмента текста, но и извлечение семантических связей между ними. Такой результат достигнут в результате применения алгоритмов глубинного семантического анализа с использованием формализованных методов синтаксического анализа и семантических словарей структуры, описанной в **Ошибка! Источник ссылки не найден.** и **Ошибка! Источник ссылки не найден.** Семантический анализ такой глубины до сих пор применялся только в системах машинного перевода и орфографии, да и то не всегда. Применение этого инструментария применительно к решению задач семантической безопасности применяется впервые.

Суть метода заключается в следующем. При обработке текстовых фрагментов (не важно являются ли они анализируемыми образцами или служат для обучения рубрикатора) строится т. н. семантический граф текста **Ошибка! Источник ссылки не найден.**, **Ошибка! Источник ссылки не найден.** Вторым этапом является извлечение из текста информационных признаков определенной структуры. Отбор информационных признаков происходит по набору масок. Критерием построения маски является целостность и информационная самодостаточность информационного признака. Рассмотрим следующее предложение: «Большая часть работы посвящена тому, что мы сейчас бы назвали семантикой морфологии или просто грамматикой языка». Информационными признаками, удовлетворяющими множеству масок, являются: «Большая часть работы», «семантика морфологии», «грамматика языка». На следующем этапе происходит «взвешивание» информационных признаков. Существует множество подобных алгоритмов, поэтому не будем останавливаться здесь на их описании. Отметим только что вес, как правило, является функцией от локальной и глобальной частоты встречаемости информационного признака и пропорционален «весу текста», задаваемому аналитиком. Таким образом, задача анализа принадлежности анализируемого текстового фрагмента к заданной предметной области сводится к задаче оценки степени корреляции двух векторов. Такая задача может быть решена с помощью методов непараметрической статистики, например, с использованием распределений коэффициента ранговой корреляции Кендалла. Результатом, что очень важно, будет являться количественный параметр, являющийся вероятностью ошибки первого рода. Таким образом, при проведении анализа степени потенциальной опасности некоего текстового информационного сообщения мы получаем конкретную *количественную оценку* степени опасности данного сообщения, разумеется, при наличии качественно обученной по той же схеме совокупности рубрик предметных областей.