DOI 10.37882/2223-2966.2023.12-2.26

МЕТОДЫ И АЛГОРИТМЫ ДЛЯ ВЫЯВЛЕНИЯ ПОТЕНЦИАЛЬНЫХ ВИРУСОВ И ЭКСПЛОЙТОВ В ТЕКСТОВЫХ ДАННЫХ ОБРАТНОЙ СВЯЗИ

METHODS AND ALGORITHMS FOR DETECTING POTENTIAL VIRUSES AND EXPLOITS IN TEXT FEEDBACK DATA

A. Rusakov
E. Amelyutin
Ya. Rudkov

Summary. The paper describes one of the possible approaches to identifying potential viruses and exploits in text feedback data. Timely detection of traces and descriptions of viruses and exploits in text feedback data contributes to an increase in the overall level of security. The paper presents an approach for detecting cyber threats by the method of intelligent text analysis, capable of working on big data in order to identify attacks that can be implemented in the form of Internet links, polymorphic malware and Trojan cryptographers engaged in extortion.

Keywords: text mining, information security, detection exploits in the text, feedback.

Русаков Алексей Михайлович
старший преподаватель, МИРЭА — Российский технологический университет
rusal@bk.ru
Амелютин Евгений Вячеславович
Доцент, МИРЭА — Российский технологический университет
amelyutin9@yandex.ru
Рудков Ярослав Игоревич
МИРЭА — Российский технологический университет
yaroslav.rudkov2105@mail.ru

Аннотация. В работе приводится описание одного из возможных подходов к выявлению потенциальных вирусов и эксплоитов в текстовых данных обратной связи. Своевременное обнаружение следов и описаний вирусов и эксплойтов в текстовых данных обратной связи способствует повышению общего уровня безопасности. В работе приводится подход для обнаружения киберугроз методом интеллектуального анализа текстов, способный работать на больших данных с целью выявления атак, которые могут быть реализованы в виде Интернет-ссылок, полиморфных вредоносных программ (polymorphic malware) и троянских программ-шифровальщиков, занимающихся вымогательством.

Ключевые слова: интеллектуальный анализ текстов, информационная безопасность, обнаружение эксплоитов в тексте, обратная связь.

Введение

Современные организации и компании все более активно взаимодействуют с клиентами и пользователями через различные онлайн-каналы обратной связи, включая электронную почту, чаты, социальные сети и другие платформы. Обеспечение безопасности и защита информации, поступающей через эти каналы, становятся важнейшими задачами [1].

Объем данных, генерируемых пользователями, постоянно растет. Автоматизированный анализ и обнаружение угроз в таких данных становятся неотъемлемой частью стратегий информационной безопасности.

Злоумышленники постоянно совершенствуют методы атак и маскировки. Анализ текстовых данных на предмет вирусов и эксплойтов требует постоянного обновления и разработки новых методов для действенной защиты.

Методы обработки естественного языка (NLP—Natural Language Processing») становятся все более мощными и доступными. Их применение для анализа текстовых данных с целью обеспечения информационной безопасности предоставляет новые возможности и перспективы.

Во многих странах ужесточаются законодательные требования, касающиеся обеспечения конфиденциальности и защиты данных. Работа в области обеспечения информационной безопасности и анализа текстовых данных может помочь организациям соблюдать эти нормы и требования.

Таким образом, разработка программного средства интеллектуального анализа текстов каналов обратной связи является не только актуальной, но и имеет широкий практический и теоретический интерес в контексте современных вызовов в области информационной безопасности и обработки текстовых данных.

Алгоритмы для выявления потенциальных вирусов и эксплойтов в текстовых данных обратной связи

Обработка естественного языка (Natural Language Processing, NLP) может быть специально ориентирована

Таблица 1.

Методы обработки естественного языка и их применимость для выявления угроз информационной безопасности

Метод	Описание	Применимость
Обнаружение ключевых слов	Поиск ключевых слов, связанных с угрозами, в тексте.	Обнаружение текстов, содержащих ключевые слова угроз, например, «вирус», «эксплойт».
Машинное обучение для обнаружения аномалий	Использование алгоритмов обнаружения аномалий для выявления подозрительных текстовых паттернов.	Выявление подозрительных текстовых данных, которые не соответствуют нормальным образцам.
Обнаружение схожих текстов	Сравнение текстов с базой данных известных угроз.	Идентификация текстов, аналогичных известным угрозам, таким как вирусы и эксплойты.
Извлечение информации	Извлечение информации о файлах, URL-адресах и других данных, связанных с угрозами.	Анализ текста на наличие ссылок на вредоносные ресурсы или другой информации, связанной с угрозами.
Модели тематической классификации	Определение текстов, связанных с угрозами, на основе тематики.	Классификация текстов по их тематике, позволяя выявить угрозы на основе содержания.
Обработка метаданных	Анализ метаданных, таких как время, отправитель и получатель текста.	Выявление потенциальных угроз, связанных с метаданными текста.
Лингвистический анализ	Определение попыток обхода системы обнаружения угроз с помощью изменений в языке или структуре текста.	Выявление подозрительных изменений в тексте.
Обработка внешних данных	Интеграция данных из внешних источников, таких как базы данных известных угроз.	Сравнение текстов с известными угрозами из внешних источников.
Обратная связь	Механизмы обратной связи с пользователем для реагирования на угрозы.	Позволяют быстро реагировать на обнаруженные угрозы и предпринимать меры по их предотвращению.

на выявление вирусов, эксплойтов и других угроз информационной безопасности в текстовых данных с использованием различных методов и техник.

Рассмотрим таблицу, представляющая различные методы обработки естественного языка и их применимость для выявления угроз информационной безопасности — таблица 1 [2, 3].

Обратная связь в контексте информационной безопасности представляет собой механизмы, с помощью которых система или организация взаимодействует с пользователями или персоналом для реагирования на обнаруженные угрозы и принятия мер по их предотвращению. Она играет важную роль в обеспечении безопасности информационных систем и защите данных.

Принципы и механизмы обратной связи:

— Обнаружение угроз и инцидентов [4]. Системы информационной безопасности мониторят и анализируют данные и события, чтобы выявлять потенциальные угрозы и инциденты. Это может включать в себя анализ системных журналов, сетевого трафика, сигналов от антивирусных программ и других источников.

— Сигнализация и оповещение. При обнаружении угроз система генерирует сигналы и уведомления, которые могут быть направлены на мониторинговый центр или ответственных сотрудников. Это может включать в себя автоматические уведомления, электронные письма, текстовые сообщения и другие каналы связи.

— Прослеживание инцидентов. После обнаружения угрозы или инцидента система может проследить его ход, включая его источник, характеристики и методы атаки. Это позволяет лучше понять природу угрозы.

— Анализ и реагирование. Обратная связь включает в себя анализ угроз и принятие мер по их нейтрализации. Это может включать в себя блокировку подозрительных активностей, изоляцию уязвимых узлов, обновление средств безопасности и другие действия.

— Сообщение и коммуникация. Пользователям и персоналу предоставляются информация о событиях, угрозах и принятых мерах. Это может включать в себя создание бортовых докладов, уведомлений и документации по инцидентам.

— Обучение и улучшение. Обратная связь позволяет организациям учиться на опыте. Анализ угроз и реакции на них может привести к улучшению системы безопасности, внесению изменений в политику безопасности и обучению персонала.

— Превентивные меры. Важным аспектом обратной связи является применение превентивных мер для предотвращения будущих угроз. Это может включать в себя улучшение средств безопасности, обновление программного обеспечения и ужесточение политики безопасности.

Обратная связь в информационной безопасности позволяет быстро реагировать на обнаруженные угрозы, минимизировать потенциальный ущерб и защищать информацию и ресурсы организации от различных видов атак.

Выявление потенциальных вирусов и эксплойтов в текстовых данных обратной связи может быть выполнено с использованием методов обработки естественного языка (NLP), анализа текста и машинного обучения.

Рассмотрим последовательность действий, которые будут применены для этой цели (рисунок 1) [4, 5]:

Рис. 1. Последовательность действий для выявления потенциальных вирусов и эксплойтов в текстовых данных обратной связи

— Собрать текстовые данные обратной связи от пользователей или клиентов, включая обзоры продуктов, комментарии, письма обращений и другие формы обратной связи.
— Предварительная обработка текста: произвести предварительную обработку текста, включая токенизацию, удаление стоп-слов, приведение к нижнему регистру и лемматизацию.
— Создать векторное представление текста с использованием методов векторизации, таких как TF-IDF или word embeddings (Word2Vec, FastText).
— Использовать обученную модель машинного обучения для классификации текстовых данных на два класса: «потенциально вирус» и «не потенциально вирус».
— Модель может выдавать вероятность того, что текст содержит потенциально вредоносное содержимое. Выбрать порог вероятности для определения подозрительных текстов.
— В зависимости от результата проверки, автоматически реагировать на обнаруженные потенциальные угрозы, включая блокировку доступа, отправку предупреждений и другие меры безопасности.
— Обратная связь и обучение модели: включить обратную связь в процесс обучения модели, чтобы улучшить ее производительность.

Последовательность действий позволит автоматически выявлять потенциальные вирусы и эксплойты в текстовых данных обратной связи с целью предотвращения угроз информационной безопасности.

Анализ модели выявления потенциальных вирусов и эксплойтов в текстовых данных обратной связи

Модель для выявления потенциальных вирусов и эксплойтов в текстовых данных обратной связи может включать в себя различные методы машинного об-

Таблица 2.

Сравнение методов для выявления потенциальных вирусов и эксплойтов

Метод	Описание	Преимущества	Недостатки
Модель мешка слов (BoW)	Подходит для выявления паттернов на основе частоты слов.	Прост в реализации, хорошо работает на коротких текстах.	Не учитывает семантику, требует большого словаря.
Модель TF-IDF	Учитывает важность слов в контексте корпуса текстов.	Улучшает веса слов, отлично работает для категоризации текстов.	Требует больше вычислительных ресурсов.
Word Embeddings	Учитывает семантическое сходство слов.	Улучшает понимание смысла текста.	Требует больших объемов данных для обучения.
Рекуррентные нейронные сети (RNN)	Анализируют текст последовательно.	Эффективны для анализа текстовых последовательностей.	Могут быть более трудоемкими при обучении.
Сверхточные нейронные сети (CNN)	Выделяют ключевые признаки из текста.	Работают хорошо для выделения шаблонов в тексте.	Могут потреблять больше вычислительных ресурсов.
Преобученные модели (например, BERT)	Учитывают семантику и контекст.	Производительны и могут работать на разных задачах.	Требуют больше вычислительных ресурсов.

учения и анализа текста. Примером может быть модель, основанная на методах обработки естественного языка и классификации текста.

Сравнение моделей представлено в таблице 2 [5, 6].

Метод TF-IDF (Term Frequency-Inverse Document Frequency) можно использовать для выявления потенциальных вирусов и эксплойтов в текстовых данных обратной связи из-за нескольких факторов [7]:

— Учет важности слов. TF-IDF учитывает не только частоту слова в тексте (Term Frequency) но и важность этого слова в контексте всего корпуса текстов (Inverse Document Frequency). Это позволяет выделять ключевые слова, которые могут указывать на потенциальные угрозы. Эксплойты и вирусы могут часто содержать специфичные ключевые слова или фразы, которые могут быть выделены методом TF-IDF [8].
— Особенности контекста. TF-IDF помогает учитывать контекст, в котором встречаются слова. Это важно для выявления угроз информационной безопасности, так как слова могут иметь разные значения в разных контекстах. Системы безопасности могут отслеживать особенные комбинации слов и фраз, которые часто связаны с угрозами, и использовать метод TF-IDF для их выявления.
— Уменьшение шума. TF-IDF помогает уменьшить влияние общих слов, которые могут быть шумом в анализе текста. Это позволяет сосредоточиться на ключевых словах и фразах, которые могут быть связаны с вирусами и эксплойтами.
— Применимость к большим данным. TF-IDF хорошо масштабируется и может использоваться для анализа больших объемов текстовых данных, что важно для обработки обратной связи из сервисов и документации.

TF-IDF — это статистическая модель, используемая в обработке текста и информационном поиске для оценки важности слов или фраз в документе относительно большой коллекции документов. Она представляет собой метод выявления ключевых слов, фраз и понятий в текстовых данных на основе их частоты в документе и обратной частоты их встречаемости в других документах. TF-IDF помогает определить, насколько термин (слово или фраза) важен для конкретного документа в контексте всей коллекции.

Модель TF-IDF состоит из двух основных компонентов [9,10]:

— Term Frequency (TF, Частота слова). Это мера, оценивающая, насколько часто конкретное слово или фраза встречается в документе. TF вычисляется по формуле:

TF(t) = (Количество раз, когда термин t встречается в документе) / (Общее количество терминов в документе)

Чем чаще слово встречается в документе, тем выше его TF.

— Inverse Document Frequency (IDF, Обратная частота документа). Это мера, которая оценивает, насколько термин уникален и важен среди всех документов в коллекции. IDF вычисляется по формуле:

IDF(t) = log(Общее количество документов / (Количество документов, содержащих термин t))

Чем менее общеупотребительное слово исследуется в коллекции, тем выше его IDF.

— TF-IDF скор. Финальная оценка важности термина вычисляется путем умножения TF и IDF:

$$TF\text{-}IDF(t) = TF(t) * IDF(t) \quad (1)$$

Высокий TF-IDF указывает на важность термина в конкретном документе и в контексте всей коллекции.

Применение модели TF-IDF:

— Информационный поиск — поиск и ранжирование документов по запросам пользователей.
— Кластеризация и категоризация — группировка документов по схожим темам или категориям.
— Классификация текста — определение категории или метки документа на основе его содержания.
— Выявление ключевых слов — идентификация наиболее важных слов и фраз в тексте.
— Анализ тональности — оценка эмоциональной окраски текста.

Модель TF-IDF позволяет автоматически выделять ключевые слова, фразы и понятия в тексте, что делает ее полезной в различных задачах обработки текста и анализа информации.

Анализ семантики текста и модели векторного представления слов (Word Embeddings) являются важными инструментами для более точного понимания смысла и эмоциональной окраски текстов обратной связи [10].

Семантический анализ текста позволяет понимать, о чем именно идет речь в тексте. Это включает в себя выделение ключевых слов и фраз, определение темы и контекста сообщения. Анализ семантики также может помочь определить, является ли текст позитивным, негативным или нейтральным, а также определить эмоциональную окраску текста.

Модели векторного представления слов (Word Embeddings) преобразуют слова и фразы в векторы чисел, представляющие смысл слова и его контекст. Это позволяет машинным алгоритмам понимать смысл слов и их взаимосвязи. Применение моделей Word

Embeddings позволяет выявлять синонимы, антонимы и ассоциации между словами, что улучшает понимание смысла текста. Эти модели также могут использоваться для анализа тональности текста, определяя, является ли текст позитивным, негативным или нейтральным.

Применение анализа семантики и моделей векторного представления слов для текстов обратной связи может помочь:

— Определить, о чем именно пишет клиент, и выявить ключевые аспекты, которые интересуют клиента.
— Понять эмоциональную окраску текста, что поможет в выявлении уровня удовлетворенности клиентов или выявлении проблем и негативных отзывов.
— Определить аспекты продукта или услуги, которые вызывают наибольшие эмоции у клиентов, и принять меры для улучшения качества.

Примером моделей векторного представления слов может быть Word2Vec, FastText, GloVe и другие. Эти модели могут быть обучены на больших корпусах текстовых данных и использованы для анализа семантики текстов обратной связи.

Использование векторных представлений слов (Word Embeddings) в задачах выявления потенциальных вирусов и эксплойтов в текстовых данных обратной связи имеет ряд преимуществ и обоснований [11]:

— Схожесть контекста. Word Embeddings позволяют находить слова, которые схожи по контексту. Это полезно для обнаружения вариаций и схожих выражений, которые могут использоваться в тексте, чтобы скрыть настоящую угрозу.
— Автоматическое извлечение ключевых слов. Word Embeddings помогают автоматически выделять ключевые слова и фразы в тексте. В случае обнаружения вирусов и эксплойтов, это может помочь идентифицировать наиболее важные термины.
— Обнаружение аномалий. Векторные представления слов могут использоваться для обнаружения аномалий в тексте. Если вирус или эксплойт содержит необычные словосочетания или комбинации слов, Word Embeddings могут помочь выявить такие аномалии.
— Контекстуальное понимание. Некоторые вирусы и эксплойты могут использовать контекстуальные зависимости между словами. Word Embeddings, такие как BERT и GPT, обладают способностью учесть контекст в тексте и понимать связи между словами.
— Многоязыковость. Многие модели векторных представлений слов поддерживают несколько языков. Это важно, так как угрозы могут возникать на разных языках, и система должна быть способной выявлять их независимо от языка текста.
— Обучение на специфических данных. Векторные представления слов могут быть обучены на специфических корпусах текста, связанных с информационной безопасностью, что позволяет модели лучше выявлять угрозы и специфические термины.

Таким образом, использование Word Embeddings обосновано для задачи выявления потенциальных вирусов и эксплойтов, так как оно улучшает способность системы понимать смысл текста и выявлять угрозы с высокой степенью точности.

Сочетание Word Embeddings с TF-IDF будет полезным для выявления потенциальных вирусов и эксплойтов в текстовых данных обратной связи. Оба метода имеют свои преимущества, и их совместное использование усилит способность системы к обнаружению угроз информационной безопасности.

Word Embeddings позволяют учесть семантику слов и фраз, что может помочь в выявлении смысла и контекста в тексте. Слова, близкие по смыслу, будут иметь схожие векторные представления. Это полезно для выявления скрытых угроз, где вирусы или эксплойты могут быть описаны с использованием разных слов и синонимов.

TF-IDF позволяет выделить ключевые слова и фразы в тексте на основе их важности. Это может помочь выявить наиболее релевантные термины, которые могут указывать на наличие угрозы. Комбинирование TF-IDF с Word Embeddings позволяет фокусироваться на наиболее значимых словах и контексте в тексте.

TF-IDF может помочь в выявлении аномалий в тексте, так как необычные слова или фразы могут иметь высокий TF-IDF-скор. Это может помочь выявить потенциально вредные угрозы.

Если используется Word Embeddings, которое обладает контекстуальным пониманием, такое как BERT или GPT, оно может помочь понимать зависимости между словами и фразами. Это важно для обнаружения скрытых угроз, которые могут использовать сложные контексты.

Многие модели Word Embeddings поддерживают несколько языков. Это важно, так как угрозы могут возникать на разных языках.

Сочетание Word Embeddings с TF-IDF позволяет создать комплексный подход к анализу текста, который учитывает семантику, важность слов и контекст, что существенно повышает эффективность обнаружения потенциальных вирусов и эксплойтов в текстовых данных обратной связи.

Анализ текстовых данных сервисов обратной связи с учетом аспектов информационной безопасности

Предоставим таблицу с методами классификации текста и их применимостью для выявления угроз информационной безопасности — таблица 3 [11].

Таблица 3.

Методы классификации текста и их применимостью для выявления угроз информационной безопасности

Метод	Применение
Методы наивного Байеса	Фильтрация спама, определение тональности, обнаружение фишинговых писем, зловредных сообщений.
Метод опорных векторов (SVM)	Бинарная и многоклассовая классификация, обнаружение аномалий, защита от несанкционированного доступа.
Глубокое обучение (RNN, CNN)	Классификация текста, анализ контекста и семантики, обнаружение вредоносных паттернов в тексте.
Методы ансамблей (Random Forest)	Повышение точности и надежности классификации, обнаружение вредоносных активностей и сетевых атак.
Обработка естественного языка	Выделение ключевых фраз, анализ семантики и синтаксиса, выявление аномалий и угроз информационной безопасности.
Системы машинного обучения с подкреплением	Адаптивное обучение и принятие решений на основе текстовых данных, выявление уязвимостей в информационной системе.

В контексте информационной безопасности NLP используется для выявления угроз, мониторинга текстовых данных и обнаружения аномалий [8]:

— Выделение ключевых фраз. NLP-системы могут анализировать текстовые данные, чтобы выделить наиболее важные слова и фразы. Это может быть полезным для идентификации ключевых терминов, связанных с информационной безопасностью, или для выявления ключевых слов в текстах, которые могут указывать на возможные угрозы.

— Анализ семантики и синтаксиса. NLP позволяет понимать не только отдельные слова, но и их взаимосвязи в предложениях и текстах. Это позволяет определять смысл и контекст сообщений. Например, анализ семантики может помочь распознать, является ли текст угрожающим или безопасным.

— Выявление аномалий. NLP может быть использован для обнаружения аномалий в текстовых данных. Это включает в себя поиск необычных или подозрительных паттернов в текстах, которые могут указывать на потенциальные угрозы. Например, обнаружение необычных активностей в текстовых сообщениях или сетевых данных.

— Выявление угроз информационной безопасности. Путем анализа текстовых данных NLP может помочь выявить угрозы информационной безопасности, такие как фишинговые письма, вредоносные ссылки, атаки на сети и другие виды кибератак. HLP может помочь идентифицировать характеристики, свойственные таким угрозам, и реагировать на них.

Все эти аспекты обработки естественного языка важные для обеспечения информационной безопасности, поскольку они позволяют анализировать большие объемы текстовых данных, выявлять потенциальные угрозы и реагировать на них, что помогает защитить информацию и системы от различных видов атак.

Анализ текстовых данных сервисов обратной связи с учетом аспектов информационной безопасности — это важная задача, которая помогает организациям и предприятиям не только понимать отзывы и комментарии клиентов, но и обеспечивать безопасность своих систем и данных. Выделим несколько аспектов, которые учтены при анализе таких текстовых данных с учетом информационной безопасности: выявление угроз и атак, обнаружение фишинговых попыток, защита личных данных, мониторинг соблюдения политики безопасности, реакция на инциденты [10].

Сервисы обратной связи могут быть использованы для распространения фишинговых писем и ссылок. Анализ текстов позволяет выявить подозрительные сообщения, которые могут включать фишинговые URL-адреса или попытки мошенничества.

Анализ текстовых данных также может помочь в обнаружении утечек личных данных, таких как номера социального страхования, банковские реквизиты, адреса и другие конфиденциальные информационные данные. Выявление таких случаев и их предотвращение является важной частью обеспечения информационной безопасности.

Путем анализа текстов можно отслеживать соблюдение политики безопасности в организации. Это может включать в себя выявление случаев нарушения правил и стандартов, связанных с информационной безопасностью.

Для анализа текстовых данных с учетом аспектов информационной безопасности организации могут использовать различные методы, включая обработку естественного языка (NLP), машинное обучение и алгоритмы обнаружения аномалий. Такой анализ помогает не только понимать мнение клиентов, но и обеспечивать безопасность информации и операций компании.

Заключение

Таким образом при создании методов обнаружения киберугроз с помощью интеллектуального анализа текстов (NLP), способных обучаться на больших данных с целью выявления атак, которые могут быть реализованы в виде Интернет-ссылок, полиморфных вредоносных программ (polymorphic malware) и троянских программ-шифровальщиков, занимающихся вымогательством, могут быть представлены:

1. Метод атрибутно-ориентированного распознавания Интернет-ссылок в формах обратной связи с использованием частотных шаблонов, который может провести оценку атрибутов и отличить доброкачественные, фишинг, и вредоносные ссылки.
2. Метод детектирования полиморфных вредоносных программ в формах обратной связи, позволяет обнаруживать полиморфные вредоносные программы посредством анализа хэшей PE секций и поиска похожих секций в библиотеке вредоносного кода. После подтверждения детектирования PE файла, новые хэши секций добавляются в библиотеку.

ЛИТЕРАТУРА

1. Introduction to Natural Language Processing (NLP) [Электронный ресурс]. Режим доступа: https://towardsdatascience.com/introduction-to-natural-language-processing-nlp-323cc007df3d
2. Top 10 Pre-Trained NLP Language Models[Электронный ресурс]. Режим доступа: https://insights.daffodilsw.com/blog/top-5-nlp-language-models
3. Щипина Л.Ю. Информационные технологии в лингвистике. Учебное пособие — 2013. С.43–52.
4. Зубова И.И. Информационные технологии в лингвистике: Учебное пособие. — МГЛУ. — Мн., 2001.
5. J. Chai and J. Lin, «The role of natural language conversational interface in online sales: a case study», International Journal of Speech Technology., vol. 4, pp. 285–295, Nov. 2001.
6. Abdul-Kader, S., & Woods, J. (2015). Survey on Chatbot Design Techniques in Speech Conversation Systems. International Journal of Advanced Computer Science and Applications, 6(7). http://doi.org/10.14569/ijacsa.2015.060712.
7. What is Amazon Lex? [Электронный ресурс] — Режим доступу до ресурсу: https://docs.aws.amazon.com/lex/latest/dg/what-is.html.
8. M. Bates (1995). «Models of natural language understanding» [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC40721/
9. What is Amazon Comprehend? [Электронный ресурс] — Режим доступу до ресурсу: https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html.
10. Poibeau T., Saggion H., Piskorski J., Yangarber R. Multi-source, Multilingual Information Extraction and Summarization. Theory and Applications of Natural Language Processing. Springer, Berlin, Heidelberg, 2013. — 257 c.
11. Klopfenstein, L., Delpriori, S., Malatini, S., Bogliolo, A.: The rise of bots: a survey of conversational interfaces, patterns, and paradigms. In: Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 555–565. Association for Computing Machinery (2017).

Журнал «Современная наука: актуальные проблемы теории и практики»