Университет ИТМО, факультет программной инженерии и компьютерной техники

Двухнедельная отчётная работа по «Информатике»: аннотация к статье

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Дата прошедшей лекции | Номер прошедшей лекции | Название статьи/главы книги/видеолекции | Дата публикации (не старше 2022 года) | Размер статьи (от 400 слов) | Дата сдачи |
| 10.09.2025 | 1 | Компьютеры на основе тернарной логики и перспективы их развития | 2022г. | 1122 | 16.09.2025 |
| 24.09.2025 | 2 | 1. Resource Reduction in Multiplexed High-Dimensional Quantum Reed-Solomon Codes | 2022г. | 4690 | 22.09.2025 |
| 8.05.2025 | 3 | Применение Регулярных Выражений Для Обработки Текстовых Данных | 2022г. | 1700 | 8.05.2025 |
|  | 4 |  |  |  |  |
|  | 5 |  |  |  |  |
|  | 6 |  |  |  |  |
|  | 7 |  |  |  |  |

Выполнил(а) Трикашный М. Д. , № группы *P3114* , оценка

Фамилия И.О. студента не заполнять

|  |
| --- |
| **Прямая полная ссылка на источник или сокращённая ссылка (bit.ly, tr.im и т.п.)**  *https://cyberleninka.ru/article/n/primenenie-regulyarnyh-vyrazheniy-dlya-obrabotki-tekstovyh-dannyh* |
| **Теги, ключевые слова или словосочетания (минимум три слова)**  **Регулярные выражения, обработка, текстовые данные** |
| **Перечень фактов, упомянутых в статье (минимум четыре пункта)**   1. **Определение:** Регулярное выражение — это строка специального вида, состоящая из обычных символов и метасимволов (символов-джокеров). 2. **Назначение:** Регулярные выражения предназначены для описания языков (шаблонов текста). 3. **Три основные задачи:**    1. Проверка текста на соответствие шаблону (валидация).    2. Поиск, анализ и замена фрагментов текста.    3. Лексический и синтаксический анализ (упрощение написания трансляторов). 4. **Стандарты:** существуют разные стандарты, в основном POSIX и PCRE. 5. **Использование в языках программирования:** Регулярные выражения входят в базовый состав многих языков (JavaScript, PHP), но в некоторых требуют подключения библиотек. 6. **Примеры применения:**    1. **Валидация:** Проверка корректности email и пароля при регистрации пользователя.    2. **Обработка текста:** Автоматическое исправление ошибок форматирования (лишние пробелы, пробелы вокруг знаков препинания), удаление спама (ссылок), удаление дублирующихся слов.    3. **Лексический анализ:** Создание частотных словарей (разбиение текста на слова), распознавание специфических шаблонов (например, цветов в HEX-формате).    4. **Синтаксический анализ:** Подсветка синтаксиса, выделение определенных конструкций в тексте (например, содержимого в скобках). 7. **Расширенные возможности:** использование бэкреференсов (например, \1 для поиска дубликатов) и утверждений нулевой ширины (например, (?=.\*[0-9]) для проверки наличия цифры) позволяет работать с более сложными языками, включая контекстно-свободные и контекстно-зависимые. 8. **Актуальность:** регулярные выражения широко используются в системах искусственного интеллекта для анализа текстовых потоков и распознавания текстовых образов. 9. **Происхождение:** изначально регулярные выражения появились в области математической лингвистики. |
| **Позитивные следствия и/или достоинства описанной в статье технологии (минимум три пункта)**   1. **Унификация и верификация данных:** позволяют привести данные к единообразной форме и проверить их корректность на раннем этапе (валидация email, паролей), что систематизирует информацию. 2. **Автоматизация обработки текста:** значительно экономят время на рутинных задачах: поиск и замена, исправление опечаток и форматирования, модерация контента (удаление спама). 3. **Повышение эффективности работы с данными:** помогают избежать ошибок и дублирования записей в базах данных, что оптимизирует работу информационных систем. 4. **Упрощение сложных задач:** делают более доступными разработку таких сложных инструментов, как трансляторы, интерпретаторы, системы проверки орфографии и синтаксиса. 5. **Интеллектуальный анализ данных:** являются инструментом для классификации информационных потоков, построения частотных словарей и анализа текстовой информации, что используется в машинном обучении и AI. 6. **Гибкость и мощь:** с помощью метасимволов и расширенных конструкций можно описать чрезвычайно сложные и разнообразные шаблоны для поиска. 7. **Широкая распространенность:** являются стандартным инструментом во многих языках программирования и средах, что облегчает их изучение и применение. |
| **Негативные следствия и/или недостатки описанной в статье технологии (минимум три пункта)**   1. **Сложность чтения и написания:** сложные регулярные выражения (например, для валидации email) могут быть очень трудными для восприятия, отладки и модификации другими разработчиками. Статья прямо показывает, насколько громоздким может быть даже относительно простой шаблон. 2. **Риск ошибок:** из-за сложности синтаксиса легко допустить ошибку, которая приведет к неправильному срабатыванию или, что хуже, пропуску ошибочных данных. 3. **Проблемы с производительностью:** неоптимально составленные регулярные выражения (особенно с жадными квантификаторами или вложенными конструкциями) могут работать очень медленно на больших объемах текста. 4. **Необходимость экранирования:** специальные символы (метасимволы) необходимо экранировать, чтобы они представляли сами себя, что добавляет сложности и может привести к ошибкам. 5. **Различия в диалектах:** хотя синтаксис в целом одинаков, существуют различия между стандартами (POSIX, PCRE) и реализациями в разных языках программирования, что может вызвать проблемы при переносе кода. 6. **Ограниченность выразительной мощности:** классические регулярные выражения предназначены для описания регулярных языков. Хотя расширения (как показано в статье) позволяют выйти за эти рамки, для анализа сложных конструкций (например, натурального языка с его контекстной зависимостью) их может быть недостаточно, и требуются более мощные инструменты (например, формальные грамматики). |
| **Ваши замечания, пожелания преподавателю *или* анекдот о программистах[[1]](#footnote-1)**  Регулярки это что-то с чем-то, на моей памяти, пока самое сложное задание, которое я делал (3 доп).  Хотелось бы, чтобы в варианте с прилагательными условие было конкретнее. |

1. Наличие этой графы не влияет на оценку [↑](#footnote-ref-1)