

Университет ИТМО, факультет программной инженерии и компьютерной техники
Двухнедельная отчётная работа по «Информатике»: аннотация к статье

Дата прошедшей лекции	Номер прошедшей лекции	Название статьи	Дата публикации	Размер статьи	Дата сдачи
11.09.2024	1	Системы счисления, которым еще не время умирать.	06.06.2023	~650	09.10.2024
25.09.2024	2	Послание в чаше Петри: кодирование сообщений с помощью бактериальных паттернов	28.09.2022	~2243	09.10.2024
09.10.2024	3	Взлетит или нет — сможет ли новый язык программирования Nare стать альтернативой Си	04.07.2022	~740	23.10.2024
23.10.2024	4	strtree — классификатор строк на основе регулярных выражений	26.06.2024	~1126	06.11.2024

Выполнил _____ Валиев Р.Н. _____, № группы P3131, оценка _____

Ссылка на источник:

<https://habr.com/ru/articles/824544/>

Теги, ключевые слова или словосочетания

1. Классификатор текстов
2. Регулярные выражения
3. Строки на деревьях
4. Строковые алгоритмы

Перечень фактов, упомянутых в статье

1. Strtree — open-source библиотека в Python для классификации строк с помощью автоматически определяемых паттернов.
2. Регулярные выражения подходят для классификации коротких строк и уникальных данных, таких как IP-адреса и модели устройств.
3. В примере с помощью библиотеки strtree ищутся регулярные выражения для классификации наименований смартфонов.
4. Регулярные выражения создаются инкрементно, начиная с пустого паттерна и удлиняя его до достижения минимальной точности или ограничения по длине.

Позитивные следствия и/или достоинства описанной в статье технологии

1. Регулярные выражения подходят для коротких строк и уникальных данных, таких как IP-адреса и модели устройств.
2. Более привычный bag-of-words стирает разницу между строками с одними и теми же символами, но разным порядком в отличие от strtree
3. Strtree поддерживает различные геометрические операции, такие как объединение, пересечение, разность и т.д.
4. Хотя библиотека strtree имеет сложную внутреннюю структуру, интерфейс ее относительно прост для понимания и использования.

Негативные следствия и/или недостатки описанной в статье технологии

1. Strtree может быть не самой простой библиотекой для интеграции и настройки
2. Для оптимальной работы STRtree необходимо предварительно отсортировать данные перед их добавлением в структуру. Это может увеличить время подготовки данных, что в долгосрочной перспективе может казаться неэффективным для некоторых приложений.
3. Когда требуется частое обновление или удаление объектов в индексированной структуре, strtree может оказаться неэффективной, так как такие операции требуют перерасчета и повторной реорганизации дерева, что может приводить к значительным временным затратам.

Анекдот о программах:

Жена говорит программисту:

— Дорогой, сходи, купи хлеб. И если есть яйца, то возьми десяток

Программист заходит в магазин:

— Яйца есть?

— Есть

— Дайте 10 булок

