

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 2021 года)	Размер статьи (от 400 слов)	Дата сдачи
11.09.2024	1	Компьютеры на основе тернарной логики и перспективы их развития	20.09.2022	~1176	25.09.2024
25.09.2024	2	Помехоустойчивое кодирование в подсистеме аутентификации	19.12.2022	~1030	09.10.2024
09.10.2024	3	Регулярные выражения для обнаружения Web-рекламы на основе автоматического скользящего алгоритма	2021	~2790	23.10.2024
	4				
	5				
	6				
	7				

Выполнил(а) Дядев В. А., № группы P3131, 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/regulyarnye-vyrazheniya-dlya-obnaruzheniya-web-reklamy-na-osnove-avtomaticheskogo-skolzyaschego-algoritma/viewer>

Теги, ключевые слова или словосочетания (минимум три слова)

Регулярные выражения, оптическое распознавание символов, цифровой маркетинг, Web-реклама, скользящий алгоритм

Перечень фактов, упомянутых в статье (минимум четыре пункта)

1. Семантическая сеть Web-рекламы помогает нам лучше понять тенденции цифрового маркетинга.
2. Все рекламные объявления в интернете делятся на сайты, существующие за счёт рекламы, информационные или корпоративные сайты, а также социальные сети.
3. На данный момент для обнаружения рекламы популярно оптическое распознавание символов (OCR).
4. Selenium Webdriver помогает автоматизировать различные процессы в Web-браузерах, такие как скроллинг сайтов, перемещение между вкладками, заполнение полей.
5. В трёх таблицах на локальном сервере хранится около 600 слов регулярных выражений, которые соответствуют различным словам из цифрового маркетинга, торговым маркам и чаще всего используются в Web-рекламе.
6. Алгоритм был протестирован в браузерах Google Chrome, Mozilla Firefox и Safari на сайтах MSN, Sanborns и AhorraSeguros.

Позитивные следствия и/или достоинства описанной в статье технологии (минимум три пункта)

1. Регулярные выражения помогают бороться с Web-рекламой, а также обеспечивают конфиденциальность данных пользователей.
2. Реализованный алгоритм подстраивается под любой размер экрана.
3. Оптическое распознавание символов достаточно точно извлекает буквенный текст из изображений (85% для рукописного и 90% для печатного текстов).
4. Данный алгоритм показывает достаточно большую достоверность в обнаружении рекламы (75–95%).

Негативные следствия и/или недостатки описанной в статье технологии (минимум три пункта)

1. Оптическое распознавание символов ещё не может охватить все стратегии цифрового маркетинга.
2. Использование Web-скрейпинга в алгоритме.
3. Алгоритм не учитывает персональные расширения браузеров, а также синхронизированные аккаунты с других устройств.
4. Алгоритм осуществляет скроллинг только сверху вниз, что не позволяет анализировать рекламу, которая выдвигается сбоку.

Ваши замечания, пожелания преподавателю или анекдот о программистах¹

User:

- ChatGPT, а ты можешь писать код, ни у кого его не копируя?

ChatGPT:

- Нет, а ты?

User:

- ... 🤔

¹ Наличие этой графы не влияет на оценку