

Выполнил(а) Коломиец Н. С., № группы P3108, оценка _____
Фамилия И.О. студента не заполнять

Название статьи/главы книги/видеолекции

Использование библиотеки LLVM для генерации кода регулярных выражений

ФИО автора статьи (или e-mail)

Panzerschrek

**Дата публикации
(не старше 2019 года)**
 "07" июля 2021 г.

**Размер статьи
(от 400 слов)**
 5к+

Прямая полная ссылка на источник или сокращённая ссылка (bit.ly, tr.im и т.п.)

<https://habr.com/ru/post/563570/>

Теги, ключевые слова или словосочетания

Программирование, Компиляторы

Перечень фактов, упомянутых в статье

1. Поиск сопоставления текста регулярному выражению реализуется через набор вложенных вызовов функций. Функция для элемента n проверяет текущий символ в строке на соответствие и, если оно найдено, вызывает функцию для элемента $n+1$.
2. В общих чертах, работает следующим образом: строка исходного регулярного выражения разбирается в первое промежуточное представление (линейное). После этого происходит преобразование во второе промежуточное представление (в виде графа). Далее по графу строятся уже LLVM match функции
3. Очевидный способ избавления от рекурсии в поиске сопоставления для последовательностей — заменить рекурсию циклом

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

1. Можно использовать весьма сложные регулярные выражения
2. Регулярные выражения не совсем константны, но в то же время каждое регулярное выражение используется значительное количество раз или используется для поиска в большом объёме данных
3. Встраивание данной библиотеки в компилятор какого-либо языка, который уже и так использует библиотеку LLVM для оптимизации и кодогенерации
4. Использование данной библиотеки как системной библиотеки регулярных выражений (на замену PCRE)

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

1. Можно использовать регулярные выражения, которые являются константами во время сборки программы
2. Разбухание исполняемого файла на два-три десятка мегабайт — для LLVM библиотек, выполняющих оптимизации и компиляцию в бинарный код



