

ОПЫТ ТЕСТИРОВАНИЯ СКОРОСТИ ЛЕКСИЧЕСКОГО АНАЛИЗА ПРОГРАММ

В последнее время в сфере языков программирования, используемых для обработки текстов, наибольшую известность имеют Python и Perl, но согласно рейтингу PYPL (Popularity of Programming Language, <https://pypl.github.io/PYPL.html>) все меньше программистов выбирают Perl, отдавая предпочтение языку, занимающему в рейтинге PYPL первое место – Python. Вместе с тем все еще востребованы связанные с обработкой текстов большие проекты, написанные в том числе и на Perl, например, ассемблер NASM (Netwide Assembler, <https://github.com/letolabs/nasm>).

Часто выбор языка программирования лежит в субъективной плоскости, включающей в себя удобство и т.п. но в связи с развитием сферы Big Data на первое место выходит производительность, так что выбор наиболее производительного языка является особенно важным.

Анализ скорости обработки текстов с целью определить наиболее быстрый в сфере обработки текстов язык программирования производится на примере лексических анализаторов программ на языке Pascal как на одном из наиболее простых в анализе искусственных языков. Результаты исследования (<https://github.com/alex-rusakevich/perlvspython>) демонстрируют, что Python (CPython версии 3.8.5) справляется с поставленной задачей в среднем в 1,6 раза быстрее, чем Perl (версии 5.30.0) при обработке текстовых файлов большого размера, но отстает от Perl при обработке текстовых файлов небольшого размера, справляясь с задачей в 1,74 раза медленнее.

Исходя из полученных результатов, можно увидеть, что Python быстрее справляется с задачей анализа большого текста, чем Perl, а значит, является более предпочтительным в сфере анализа больших объемов текстовых данных. Вместе с тем Perl быстрее справляется с файлами небольших размеров, исходя из чего, учитывая скорость написания сценариев, хорошо подходит для небольших задач.