

classifying the input domain

这个文档记录了grep程序的分区信息。

Arlinda在毕业论文中将grep的pattern划分了14个范畴，并且将这14个范畴划分了2组：第一组范畴中的选项不需要与其它范畴中的选项进行组合 就能实例化成有效的测试用例；第二组范畴中的选项必须与其它范畴中的选项进行组合，才能实例化成有效的测试用例。

independent category (A)	dependent category (B)
NormalChar	Bracket
wordSymbol	Iteration
DigitSymbol	Parentheses
SpaceSymbol	Line
NamedSymbol	Word
AnyChar	Edge
Range	Combine

在测试的过程中，每一个测试帧实例化一个测试用例，如果将每一个测试帧作为一个分区，则体现不出APT的特点。因此，我们考虑得到“粗粒度”的分区，使得每一个分区中存在多个测试用例。详细地的分区步骤描述如下。

partition scheme 1

考虑A组中的范畴，忽略B组中的范畴，得到分区模式1。A组中范畴的每一个选项可以单独地实例化为一个有效的测试用例，也可以与A组中的其它选项进行有效地组合，然后实例化成测试用例。我们识别A组范畴中选项的所有有效组合，得到99762个测试帧，具体的选项组合记录在partition_scheme_1.1中。

按照上述方法进行分区，使得每一个分区中的测试用例数目平均不到2个。如果测试帧中包含相同的感兴趣的选项，那么这些测试帧实例化的测试用例具有相似的执行路径。基于上述考虑，我们忽略了测试帧中选项的顺序，得到552个“粗”粒度的测试帧，即此时的分区数目为552（具体的选项组合记录在partition_scheme_1.2中）。

partition scheme 2

考虑B组中的范畴，忽略A组中的范畴，得到分区模式2。该模式存在49648个“有效的”测试帧（具体的选项组合记录在partition_scheme_2.1中），即可以将测试用例划分到49648个分区中，平均每个分区包含3个测试用例。

基于partition scheme 1相同的考虑，我们忽略了“partition_scheme_2.1”中测试帧中选项的次序，得到3380个分区（具体的选项组合记录在partition_scheme_2.2中）。