0508-phenomizer讨论

* 相似分数：
* 返回一个高分，当查询中的每个hpo在疾病中都有一个好的match
* P-Value：
* 直觉上，真正的疾病中不太可能观察到随机产生的一套临床特征（hpo），所以输入与真正疾病相关的一套临床特征所取得的分数会更高。
* 如果一个分数在随机情况下很少见，我们认为它是统计上显著的。
* P-Value矫正：
* 同时检验多个假设时，（例如同时判断多个疾病是否显著），如果保持单检验的alpha，仅仅由于偶然导致“至少有一个检验判断错误”的概率很大，需要矫正p-value / 调整alpha
* 效果衡量：
* 根据文献，收集一些疾病的临床特征出现频率，生成“模拟患者”进行测试
* 比较返回中正确疾病的排名
* 三个方法：
* FV，OSS（直接用相似分数），OSS-PV（相似分数的p-value）
* 优势：对以下因素更鲁棒 ——
* 噪声：当患者身上有无关的临床特征时
* 不准确性：一个临床特征被用它的父类描述