ir4

1. 哪个系统返回的结果更好？

我认为第二个列表更好。因为在达到某个相同的召回率时，第二个往往有更高的精度。且在更短的序列中找回了所有的相关文档。

1. 第三代搜索引擎是什么样的？特征是什么？

个性化索引——针对于每个用户、当前场景，猜测适合于当前场景的结果并提供最适合与单个用户的返回结果，并能与其他应用软件串联。

多类型、跨类型信息检索——声音、视频、图像、文字的融合搜索。

特征：大规模数据分析、分布式计算、模式识别……

1. 如何分配url？

综合考虑各计算机的性能（效率）、计算机上剩余的任务进行分配，最大化爬虫的运行速度。

（同一域名分给同一机器。）