**Using Graphics Processors for High Performance IR Query Processing**

**论文研究报告**

首先说，选择这篇论文的原因:

1：在NBLucene中，展望未来的工作中，最后一段话，说实现一个搜索引擎能够支持在GPU系统架构的平台上。

2：参加了实验室组会，第一位同学是介绍GPU的相关知识，由于当时确实没有做好相关的准备，可能对该同学讲述的内容不是很了解，这次正好借助此篇论文了解一下。

1. A summary of the paper.

这篇论文主要讲述了设计了一款基于GPU架构的信息检索架构的实现。除此之外，对于倒置索引压缩，And合并和Or分离逻辑查询，和返回K个分数最高的相关文档以及怎么在以GPU平台上取得尽量高的效率问题进行了较为详细的介绍。

2.    A discussion of the main point/contribution of the paper.

五个主要贡献主要如下所示：

1：这篇论文最主要的是设计了一款基于GPU系统架构的信息检索的架构的实现。（这个架构可以同时运行在以CPU或者GPU任何一方或者两方的系统架构）

2：评估了两个压缩算法，分别为Rice Coder和PForDelta。

3：索引列表合并以及实现在查询中，实现了只有一小部分数据被解压

4：通过两亿五千万页面的相关查询，比较了各自以GPU为基础和以CPU为基础的查询相关性能。

5：在不同的调度策略下，比较了以CPU和以GPU系统架构为基础的相关查询。

3.   A discussion of the strengths of the paper.

1：就作者所写这篇论文之前（2009年），还没有相关出版的论文发表有关GPU执行搜索引擎这一领域，可以说是开创了基于GPU架构的关于搜索引擎实验的先河。

2：首先我觉得逻辑结构组织做的比较好，基本上是一个想法后面跟着一个实验，如果是对比性的话，会有很直观的图表显示。比如说两种压缩算法的对比，执行查询时不同的平台之间的对比，这些通过图表的形式，都给人一种很清晰的感觉。

3：内容介绍较为详细，跟Nblucene的那篇论文对比，可以把每一个步骤很清晰的展示出来。两个例子：

I就比如说索引压缩这一块，先提出了我们为什么要压缩索引 –》 后来跟着介绍了两种压缩算法 -》 然后介绍压缩算法的规则，以及可以有的优化 -》最后还介绍了解压时相应的规则。

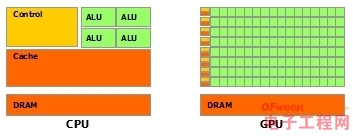
II在引入GPU执行搜索步骤时，先叙述了CPU执行搜索的步骤，让人们了解CPU执行搜索时是什么步骤，然后产生对比，讲述了两个系统架构执行搜索的不同。最后再过渡到在以GPU系统架构上的实验。

4：能够善于从实验中发现问题，并给出相应的原因。比如说比较两款压缩算法在解压的过程中，得出以CPU为基础的解压时只是获得了其相对应的docID的差值，而以GPU为基础架构解压时得到其真正对应的docID，最后给了这种现象的原因。

4.   A discussion of the weakness of the pape。 If none, why not?

1：在刚引入GPU的时候，我觉得可以将CPU和GPU的系统结构给表示出来（或者给出相关的参考资料）。直接上手说 我们要在GPU里做搜索引擎的时候，感觉有点突然。

如下图：



其中绿色的是计算单元，橙红色的是存储单元，橙黄色的是控制单元。

可以简要的说明一下 目前为什么GPU在数据库和科学运算这么强。（GPU则通过较多的流处理器实现大量线程并行，使同时走一条指令的数据变多，从而提高数据的吞吐量，特别适合做普通的科学计算等方面）

2：由于文章是09年的，已经过去很长时间，一些比文中提出的更优秀的压缩方法已经实现，比如说说改进后的NewPFD,还有一些相关论文如：《基于指令级并行的倒排索引压缩》

3：只考虑了能够实现的一些想法，但比如说索引分层（应该是通过指定一个SkipIntervel，来实现分层操作，目的是可以加快查找），和提前终止的相对复杂的技术没考虑。

但这几个原因不影响文章的优秀，确实做出了非常多的贡献。

5.   A discussion of what you learned from the paper.

首先，我对搜索引擎有了一个更深的了解，当用户输入完查询的关键字时，对其后台内部所做的工作有了一个清晰的认识。即 首先将文档建立倒排表（这时候，考虑到内存空间的限制，需要对倒排列表进行压缩，具体文中的方法为 Rice Coding 和PFoUrDelta） -> 作者详细介绍了这两种压缩的编码实现和解码的规则 以及 具体的优化方案 –》 对合并列表也进行详细的介绍 –》对搜索引擎的其他内容比如说查找分块索引结构时，只需要解压一小部分索引即可，提高了大部分的效率 ，对文档的打分，以及在GPU平台实现返回TopK个文档的结构进行了较为详细的介绍。

其次，对于GPU的系统架构，通过搜索网上的资料也有了一定的了解。作者通过与CPU系统架构的对比，直观的反映出GPU的计算优势。

还有，对并行与分布式计算有了一定的理解。

最后，对论文的格式以及作者做学问时严谨的态度有了较深的认识。希望我的研究生阶段能够多读多思考这类作者的文章，提高自己的专业水平和视野。