

Lucene2.2发布：加入词条载荷、函数查询及性能提升 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/472/2021_2022_Lucene22_E5_c67_472937.htm Lucene Java 2.2已经可以使用啦！Lucene是一个完全由Java编写的高性能、全特性的全文检索引擎库。在这一版中有许多新特性，其中包括：词条载荷（payloads）允许用户将任意二进制数据和索引中的任意词条（term）相关联；函数查询（function queries）为文档匹配度（score）的计算方式提供更高的可控性（从Solr中并入）；通过NFS进行“瞬间（point-in-time）”查询将类似于快照的功能引入NFS；用来操作预分析字段的新API可以让用户操作预先分析的文档字段而无须另加空分析器代码；公开的Maven发行版Lucene最新版本中的所有模块都可以通过Maven Repository获得。

InfoQ采访了Lucene项目的代码贡献人及项目管理委员会（Project Management Committee，PMC）成员Grant Ingersoll，进一步了解了这个发布版。在谈话过程中，Ingersoll希望InfoQ注明他所说的观点和意见都属于个人性质，不代表Lucene PMC的官方观点。InfoQ了解到，Lucene 2.2的发布标志着该项目的发布周期朝着一个更短的按季度发布的方向飞跃。Ingersoll相信更频繁的发布会带来很多的好处，比如说使为社区提供Bug修复和新特性变得更快。发布过程也因为得到Maven的支持而变得更加有效率，因此未来Maven用户将能更快速地获得发布版本。InfoQ请Ingersoll详细地描述了词条载荷的特性，他说：词条载荷是一个允许信息在索引中按逐词条储存的新特性。例如，当索引Web页面时，储存某个关键词的额外信息可能会很有用，例如这个关键词关联的URL

或者经过文字分析后得出的权重系数。在更高级的应用中，为了突出语句中的名词成分相对于其它成分的重要性，储存语句中这个关键词出现的部分可能会很有帮助。我今年在ApacheCon Europe会议上的演讲中就有几张讲述词条载荷的幻灯片，感兴趣的读者可以去看看。他还描述了源于Solr的新的函数查询功能：新的搜索函数包

(`org.apache.lucene.search.function`) 允许开发人员在计算文档相关度使用某个字段的实际内容。例如，如果你在文档的字段中储存精度和纬度信息，你就可以使用这些字段中的信息影响文档的排名。就是说，如果你搜索星巴克，你可以在结果中将离用户较近的分店（假定你知道用户的地点）排在较远分店的前面。另一个例子是使用价格或者利润信息来影响排名（即给能为公司带来更大利润的产品打更高的评分，并不是说道义上我同意这样做，但它确实可以做到）。随后，在Ingersoll被问到用户对Lucene的后续版本会有什么期望时，他指出，在使用了Michael McCandless领导的几项新的内存管理技术之后，索引性能将会得到显着提升。他还提到，在最近发布的Lucene中已经加入了不少性能改善，用户可以自己去体验一下这些变化。最后，Ingersoll说，对Java 5的支持和更灵活的索引过程将是Lucene未来可能出现的特性。这一版提供了一个全面的更新日志，列出了这个版本中所有已修复的Bug、特性和优化情况。和以前的版本一样，2.2版也可以读取和导入以前版本的索引，不过一旦进行了转换，索引将不能被以前的版本（如2.1版）使用。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com