

使用OracleText构建全文搜索应用程序 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/207/2021_2022__E4_BD_BF_E7_94_A8Orac_c102_207484.htm

Oracle Text 是一种功能强大的搜索技术，它内置于 Oracle 数据库的所有版本(包括免费提供的快捷版 (XE))中。它所提供的开发 API 使软件开发人员能够轻松实现功能齐备的内容搜索应用程序。Oracle Text 可用于搜索结构化和非结构化文档，是对 SQL 通配符匹配的补充。Oracle Text 支持使用基本的布尔运算符(AND、OR、NOT、NEAR 等)将多个搜索条目组合到一起，此外，它还具有更高级的功能，如 soundex 和模糊搜索，以及结果排序等。该技术支持数百种文件类型，包括 Microsoft Office 和 PDF。Oracle Text 适合多种与搜索相关的使用情况和存储结构。Text 的应用领域包括电子商务、文档和记录管理，以及问题跟踪等。可检索的文本可以结构化形式驻留在数据库中，也可以非结构化形式驻留在本地文件系统中或 Web 上。Oracle Text 提供完整的基于 SQL 的搜索 API，该 API 包含自定义查询运算符、DDL 语法扩展、一组 PL/SQL 过程和数据库视图。通过 Text API，应用程序开发人员可完全控制索引、查询、安全、演示以及有时会需要的软件配置，在开发即需即用的非自定义软件时尤为如此。通过即需即用的软件产品，您希望使软件的配置尽可能简约，即使这意味着要在产品开发中多做一些前期工作也是如此。降低应用程序的复杂性通常会在产品生命周期的后期，尤其是在支持、维护和未来产品开发阶段中见到成效。Oracle Text 还支持文档级授权，而文档级授权通常很难在统一的同时保持高性能。借助

Text，组合了关系数据与非结构化数据的混和查询也得到了很好的支持。对于授权，这意味着您可以将全文搜索和授权合并到一个查询中。独立结果集和获得最终结果所需的过滤阶段的数量可最大程度地缩减，从而简化了应用程序的开发。Oracle Text 使应用程序开发人员从繁琐的开发中解脱出来，可以集中精力进行性能优化。Oracle Text 也是编程语言不可知的，并且也可以同样卓越的表现作为 PHP 以及 Java 应用程序工作。前一段时间，我需要提高企业内容管理 (ECM) 系统的搜索功能。我首先对 Oracle Text 的使用进行了评估。评估证明 Oracle Text 是一种构建应用程序搜索非常可行的技术：它具有高级搜索功能，支持大量不同的文件类型，可高度自定义，同时高度可伸缩。原有搜索技术的一个缺点是，您需要在数据库外部运行文件内容搜索，然后运行数据库元数据搜索，对结果进行授权，最后，合并独立的结果集。使用 Oracle Text，所有这些操作都可以在数据库中进行。ECM 系统已经使用 Oracle 数据库来存储元数据。因为此项技术已经推出，因此客户自然会选择使用，而且它也不会给客户增加任何成本。在数据库中执行自由文本搜索查询的简单方法类似于：`SELECT * FROM issues WHERE LOWER(author) LIKE %word1% AND LOWER(author) LIKE %word2% ...` 使用这种方法，每一列都需要与每个关键字单独进行匹配。在每一列中，可以任何顺序与关键字进行匹配。然而，关系数据库的设计使得它不会像上面那样有效地执行查询，而且使用这种方法会产生极其不可伸缩的应用程序。当然，您可以设计自己的索引和搜索解决方案，但是，那样您可能不会优化使用您的资源，在您已经为将搜索技术作为数据库的一部分付出了

成本的情况下尤为如此。 本文讨论 Oracle Text 在虚拟的问题跟踪应用程序中的使用。 在这个应用程序中，用户可以创建一些包含元数据和可选附加文件的问题。 该应用程序利用 Oracle Text 实现元数据和可选附加文件内容的全文搜索功能。 此处给出的示例已经在 Linux 的 Oracle 数据库 XE 上进行了测试，这些示例应该同样也可在其他 Oracle 平台上运行良好。 索引进程与搜索 Oracle Text 为可检索的数据项建立索引之后，用户才能够通过搜索查找内容。 编制索引是确保搜索性能的常用方法。 Oracle Text 的索引进程是根据管道建模的，在这个管道中，从数据存储检索来的数据项经过一系列转换之后，其关键字会添加到索引中。 该索引进程分为多个阶段，每个阶段都由一个单独的实体来处理，并可由应用程序开发人员来配置。 Oracle Text 具有适合不同用途的不同索引类型。 对于大型文档的全文搜索，适合使用 CONTEXT 索引类型。 该索引进程包括以下几个阶段： 数据检索：只是将数据从数据存储(例如 Web 页面、数据库大型对象或本地文件系统)中取出，然后作为数据流传送到下一个阶段。 过滤：过滤器负责将各种文件格式的数据转换为纯文本格式。 索引管道中的其他组件只能处理纯文本数据，不能识别 Microsoft Word 或 Excel 等文件格式。 分段：分段器添加关于原始数据项结构的元数据。 词法分析：根据数据项的语言将字符流分为几个字词。 索引：最后一个阶段将关键字添加到实际索引中。 索引构建完成后，应用程序即可通过普通的 SQL 查询执行最终用户输入搜索。 安装 Oracle Text 在默认情况下，Oracle Text 随 Oracle 数据库 XE 一起安装。 如何使用的是其他数据库版本，您需要自己安装 Oracle Text 功能。 安装了此功能后，

您只需创建一个普通的数据库用户，并赋予该用户 CTXAPP 角色。这样，用户即可执行特定的索引管理过程：`CREATE USER ot1 IDENTIFIED BY ot1. GRANT connect,resource,ctxapp TO ot1.` 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