

ORACLE SECURE ENTERPRISE SEARCH 10G



主要特性

可搜索信息库

- Web 服务器支持的 HTML 页面
- 数据库表 – 搜索 Oracle 数据库和其他任何支持 ODBC 标准的数据库。数据库表可以位于 Enterprise Search 自己的数据库实例中或者是通过网络访问的远程数据库的一部分。纯文本列和字段列都能被抓取
- 文件 – 本地或远程文件服务器可以通过 file:// 访问协议搜索
- 电子邮件 – 电子邮件和邮件列表可以按照 IMAP 邮件协议进行数据抓取
- Oracle 10g 应用服务器 Portal 信息库 – 包括私有的和公用的 Portal 页面、文件夹、子文件夹和文本项
- 桌面内容与类似企业版 Google Desktop 的桌面搜索引擎集成，后者可选配用于搜索本地下载的电子邮件、本地文件、访问过的网页等
- Secure Crawler SDK 允许定制针对其他数据源的 Crawler

Oracle Secure Enterprise Search 10g 是 Oracle 的一个独立产品，用于实现企业所有信息资产安全、高效、易用的搜索。主要特点包括：

- 能够在 Intranet Web 服务器、数据库、本地文件或文件服务器、IMAP 电子邮件、文档管理系统、应用程序和门户中搜索和定位公共的、私有的和共享的内容
- 高度安全的信息抓取、索引和搜索
- 简易的搜索界面，完美的用户体验
- 优异的搜索质量，既使对列内容同时来自公用和私有的数据源，也能将最相关的项目置顶
- 分析搜索结果，理解使用模式
- 秒以下的查询性能
- 使用您现有的 IT 知识就能轻松管理和维护

最高级别的企业安全

通过在 Internet 上使用通用的搜索引擎，一些用户已成为他们自己的信息检索专家。然而，企业 Intranet 内网的高效搜索仍然没有很好地实现。企业内部的搜索引擎和面向消费大众的公共 Internet 搜索引擎有本质不同。

首先，企业业务存储和使用的决策所需的信息散落在许多信息库中，位于大量常常互不兼容的 IT 系统中。大众化的通用搜索引擎搜索 Internet HTML 页面；但是企业搜索引擎还必须配合可搜索的数据库、电子邮件、门户、文件管理系统和定制的应用程序。通用搜索引擎假设所有文档都在 Internet 上公开访问，企业搜索引擎与之不同，必须解决私有文件或指定工作组用户共享的文件的可访问性问题。必须控制对这些文档的访问，确保只有那些用户有权浏览的文档才被搜索引擎返回。

其次，Intranet 内网搜索软件必须易于使用和管理。它必须足够开放，以适应现有的管理体制，应该便于检查和纠错。IT 组织必须能够确保搜索引擎软件不会成为收集其员工个人信息的后门，不会向他们展示第三方的商业信息，也不会成为无法审计其工作的“黑匣子”。

健壮的索引库

另一个要考虑的方面是引擎本身的安全。企业搜索引擎必须在安

最高级别的企业安全

- 搜索非公共数据源只能看见您授权浏览的内容
- 把搜索索引存储于安全的 Oracle 数据库中
- 与基于 Oracle Internet Directory LDAP 的企业安全集成。在您的数据源和企业搜索间共享 LDAP 用户和组。基于 LDAP 组员身份和用户 id 限制搜索结果
- 从数据源得到文档可见性访问控制列表 (ACL) 的多种方法
 - 由数据抓取器基于数据源、用户提供、管理员输入来使用
- 支持安全数据抓取，包括 HTTPS (安全套接字层/加密数据流)、维护会话 cookie、HTTP 证书认证和 HTML 表单登录

极好的用户体验

- 简单直观的用户界面
- 有效相关性的查询结果
- 易于部署和管理

全方面得到加强；必须和它搜索的最安全数据源一样安全。企业不能在允许在开放其数据源给该搜索引擎的同时，无意地在安全链中引入薄弱环节。例如：存储它们的索引到一个安全的存储库中如：Oracle 数据库。而许多其他的搜索引擎则把这些索引暴露在一个简单的文件系统里。

Oracle 的数据库和应用服务器技术被企业广泛用于存储和访问数据和内容。多年来，Oracle 先后投资于安全访问数据和内容、文本搜索和索引、以及连接不同数据源和 IT 系统等新技术的研发。Oracle 能够提供最高级别的企业安全。

认证和搜索结果过滤

Oracle Secure Enterprise Search 10g 广泛深入地诠释了安全搜索。您的所有资源——Intranet 网页、数据库表记录、Oracle Application Server 门户页面、电子邮件和文件或指定信息库中的文档——都能被保护。搜索结果经过过滤，保证结果页只显示用户有权访问的文档链接。Oracle 提供了三种方法完成安全结果过滤：

- Secure Enterprise Search 能够将工作基于集中认证方案，像 Unix、Microsoft 或者 Oracle Internet Directory LDAP 登录等，以辨识哪些数据源能被指定用户访问。举例来说，如果您的 LDAP 权限不允许您访问 Oracle Portal，那 Oracle Secure Enterprise Search 就不会显示任何来自 Oracle Portal 系统的搜索结果。

- 对于更加精确的安全要求，Oracle Secure Enterprise Search 能够存储与每个文档相关的访问控制列表 (ACL) 信息，作为搜索引擎索引的一部分。同时支持“ACL Crawling”和“ACL Stamping”。在 ACL Crawling 中，Oracle Secure Enterprise Search 直接从您的信息库中获得每个文档的 ACL。这并不是总能实现的，当数据源没有基于文档模型，比如一个应用程序生成的动态内容。在 ACL Stamping 中，搜索管理员通过输入允许搜索特定数据源的 LDAP 用户和组的“授权”列表，在搜索管理控制台直接指定认证角色。比如，数据抓取中得到的所有文档都能被组 1、组 2 和组 3 的用户标记为可搜索；或者用户 1、用户 2 和用户 3 都能被授予搜索某数据源所有文档的权限；如此等等。

- ◆ Oracle Secure Enterprise Search 能够通过源系统的认证机制过滤搜索请求。这称为查询时间认证。搜索引擎索引返回结果列表后，搜索引擎重新访问每个项目的来源，传输用户凭证来检查用户是否（仍然）有权查看该项目。这样，即使用户的访问权限在上次数据抓取以后有所改变，也不会导致不安全。

身份管理

优异 Intranet 搜索质量

- 有多项针对 Intranet 专利的新的相关模型
- 凝聚了 Oracle 在全文搜索方面 15 年的投资和经验
- 网页链接分析
- 冗余和近似冗余文档的发现
- 使用“渐进放松”技术优化搜索结果，用于元数据提取和文本数据及文档元数据的混合
- 数据源能够提供定制元数据影响搜索结果
- 包括所有常用语言语形学支持的智能匹配
- 查询日志分析能够用于反馈命中和不中的情况给搜索引擎
- 管理员能够对搜索提出建议链接

毫秒级查询性能

- 新的内部索引设计技术实现百万文档查询的高性能和高吞吐量

快速见效

- 几分钟内即可完成搜索的安装和配置，几小时内即可完成数据源的抓取并得到第一次查询的结果

为了避免访问控制信息的冗余，Oracle Secure Enterprise Search 10g 集成了 Oracle 的企业身份管理解决方案—Oracle Internet Directory (OID)，也能和其他身份管理解决方案如 Microsoft Active Directory 同步。

网络服务查询 API 允许指定搜索用户。这样您就能确保只有在企业 LDAP 服务器存在并能通过认证的用户才能提交搜索。

极好的用户体验

在 Oracle Secure Enterprise Search 10g 中，用户将得到与大众化的 Internet 搜索相同的直观体验。用户使用基于浏览器的界面键入类似的关键词，得到来自 Intranet 内网所有数据源的高度相关的结果。Oracle SES 提供了开箱即用的用户界面也能定制搜索结果外观。下面图 1 是 Oracle SES 默认基本搜索的用户界面。

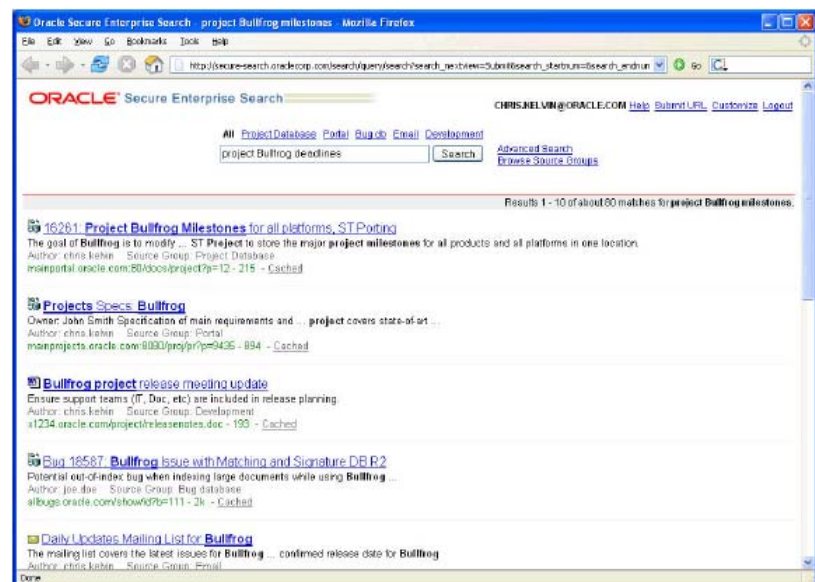


图 1: Oracle Secure Enterprise Search 10g 基本搜索界面

极好的 Intranet 网搜索质量

典型的因特网搜索返回成千上万次命中结果。类似的，随着信息库的增加，全面的 Intranet 搜索返回的搜索命中列表中匹配度较低的“垃圾信息”也会淹没用户。正如我们讨论的，Intranet 也因为不能使用 URL 引用确定页面的普及程度而受到阻碍。您需要新的方法在 Intranet 内“大海捞针”。

Oracle Secure Enterprise Search 10g 有一个新的为企业内网优化的相关模型。

- 集成了 Oracle 在全文检索方面 15 年的投资成果，其算法针对典型的企业内网负载情况进行了重新调整
- 进行了可能的和有用的 URL 链接分析

高级搜索功能

- 基本关键词查询
- 高级参数查询（包括布尔运算符）
- 动态页面摘要。搜索关键字在标题和搜索结果摘录中以粗体显示
- 快照页面
- 快照页面上高亮显示查询语句
- 使用基于网络服务的查询 API 定制搜索结果页的外观和风格

直观的基于 Web 管理环境

- 简单的自动安装
- 从 Web 控制台管理数据抓取活动
- 配置和监控安全搜索
- 报告：最常用的查询、文档没找到、图片链接、无结果的查询等许多报告
- 文档相关性提升。管理员能够提升文档相关性，定制结果集
- 保留 IT 知识

高性能可扩展数据抓取

- 多线程 Java Crawler
- 按不同计划从多个网站和其他数据源收集数据
- 能够通过 Secure Crawler SDK 改装 Crawler 抓取新的信息库
- 页面链接关系可用于更好的相关性分析和影射
- 通过设置“包含”和“不包含”来限制抓取 Intranet 特定的区域

- 包括了消除别名和二义性的功能。Intranet 搜索的很大一部分“垃圾信息”来自于近似文档。如果您创建了一个幻灯片演示文稿并作为附件发送给一个同事，他又抄送给 12 个其他人，他们将它存放在不同位置，搜索引擎是否应该返回这 12 个相同的命中，仅仅因为他们的位置不同呢？Secure Enterprise Search 引入了多个 URL 或位置的相同内容检查机制。
- 在数据抓取时从文档中提取描述文档的元数据——作者、创建日期、MIME 类型、标题等，为数据和元数据建立单独的索引，使用特别的混合机制组合搜索的元数据和数据的相对相关性。还引入“渐进放松”技术找到最优搜索结果
- 数据源能提供定制元数据，并包含于搜索中
- 包括智能匹配，包括所有常用语言的语形学支持
- 能够进行查询日志分析，决定哪个关键字返回差的命中率，将结果反馈给搜索过程，改善有效性。
- 管理员能够为常用搜索词配置“推荐链接”，根据业务需要引导用户到特定的站点

秒级以下查询性能

Oracle Secure Enterprise Search 10g 使用 Oracle 建设高扩展系统的经验提供优异的毫秒级查询性能。使用普通的双 CPU 基于 Intel 的 Linux 服务器，可以在不到一秒时间内搜索百万条文档的资料库。

立竿见影

使用 Oracle Secure Enterprise Search 10g，您无需经历漫长的开发周期就能得到投资回报。该搜索产品基于 Internet 开放标准和“无需 DBA”的观念，易于使用和部署。用户能够安装产品，开始数据抓取，并在很短时间内得到搜索结果。搜索引擎内部所需的所有组件——文本索引、Web 服务器、元数据库、数据抓取器——都捆绑在一起，无需单独安装和配置。对于安全搜索而言，安装 Oracle SES 以后，您还要将其指向合适管理用户身份的 LDAP 目录。

高级搜索

除了基本关键字搜索以外，产品还支持包括元数据搜索的高级搜索和动态页面摘要、快照、高亮显示和定制搜索结果页。下面的图 2 是高级搜索界面。

灵活的 API

- 集成 Oracle Secure Enterprise Search 和您的应用程序，或者通过基于 Web 服务的查询 API 创建新的定制搜索应用程序
- 通过 Secure Crawler SDK 实现定制数据抓取源或访问定制数据源
- 专用的 Enterprise Search 可下载的 Portlet 可将搜索集成到 Oracle 10g Portal 页面中
- 新的显示 URL 特性让您将数据库搜索结果嵌入相关的数据库应用中

联合查询

- 对其他企业搜索服务器或 Internet 搜索引擎的联合查询将结果组合到一个页面中

大量的文档类型

- Microsoft Office® Suite 95/97/2000 、 Adobe Acrobat ®PDF、 Lotus 1-2-3 ®、 Lotus Freehand ®、 Corel Word Perfect ® 等等
- 所有其他常用文档类型
- 上述类型的文档不论是磁盘文件还是电子邮件附件或 Web 网站或其他信息库中都能查询

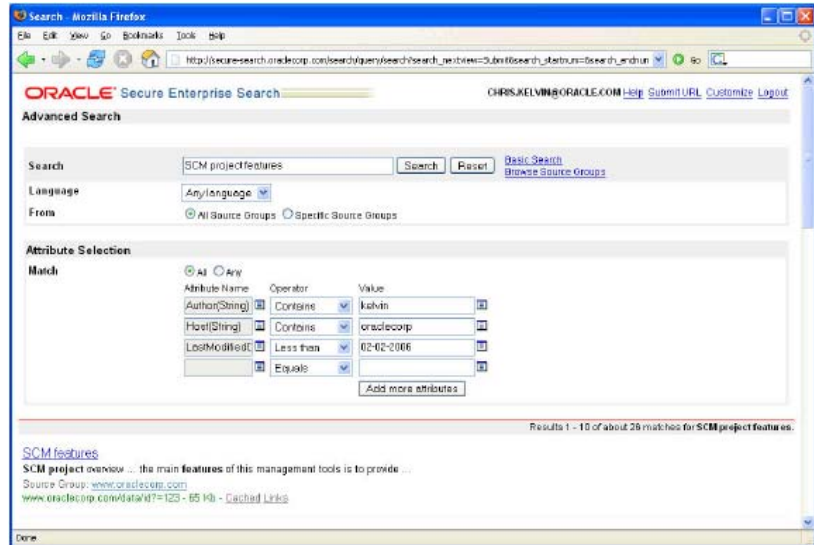


图 2: Oracle Secure Enterprise Search 10g 高级搜索界面

元数据

元数据是实现有效企业搜索的利器之一。

- Oracle Secure Enterprise Search 10g 从文档、数据库表、电子邮件和其他信息库中提取元数据字段
- 提供灵活的元数据映射方法，通过强大的全文和字段文本组合，将提取出来的元数据信息映射到查询属性
- 支持许多不同的属性类型，包括日期范围和值列表
- 基本搜索自动集成元数据搜索，决定哪些文档最相关。也能在高级搜索中显式调用元数据搜索

可以下载工具与桌面搜索工具集成，比如企业版 Google Desktop。

直观的基于 Web 的管理环境

Oracle Secure Enterprise Search 10g 提供简单基于浏览器的管理工具，统一进行软件维护。向导式的页面和对话框简化了确定数据源、指定安全规则、计划数据抓取、监视搜索有效性和调整相关性等任务。

搜索引擎管理员能快速方便地定义进行数据抓取的信息源、计划维护性数据抓取、为管理用户定义用户帐户。用户只需向 Secure Enterprise Search 管理登录网页确认自己为授权的搜索管理员即可。管理员一登录，管理环境就会显示向导式的页面和对话框帮助管理员操作搜索引擎。

Oracle SES 管理环境经过简化，保证企业不需要专门的信息检索

专家来运行和管理搜索引擎。现有的网络或数据中心管理员的 IT 知识足以用来部署和维护企业搜索。下面的图 3 是部分管理界面。

The screenshot shows the Oracle Secure Enterprise Search Administration interface. The browser window title is 'Oracle Secure Enterprise Search Administration : Statistics : Query Statistics : Most Popular Queries - Mozilla Firefox'. The URL is 'secure-search.oraclecorp.com/search/admin/queries/topquery_statistics.jsp'. The page has a navigation bar with 'Home', 'Search', and 'Global Settings'. Below the navigation bar, there are tabs for 'General', 'Sources', 'Schedules', and 'Statistics'. The main content area is titled 'Most Popular Queries' and shows a table of query statistics. The table has columns for 'Query String', 'Count', 'Total (%)', 'Hit Count', 'Click-throughs', and 'Action'. The data is filtered for the period from 2/27/06 to 2/28/06. The table lists 20 queries, with 'ldap' having the highest count of 26.

Query String	Count	Total (%)	Hit Count	Click-throughs	Action
ldap	26	8	10	0	
ora error message	12	4	57	0	
rac	11	3	97	0	
plsql	10	3	149	0	
real application clusters	10	3	91	0	
linux	9	3	203	0	
java api	8	2	66	0	
xml	8	2	99	0	
java soap	8	2	54	0	
sql tuning	8	2	1	0	
ora 800	7	2	10	0	
oracle test	7	2	999	0	
security	6	2	230	0	
tkprof help	6	2	18	0	
asp plug in	5	1	2	0	
sql rollout	5	1	9	0	
local news on hox	5	1	66	0	
Done					

图 3: 部分 Oracle SES 管理界面显示前 50 个查询

高性能可扩展的数据抓取

Oracle Secure Enterprise Search 10g 搜索大量不同的信息库。它通过在您的企业内网中抓取并审查存在于不同信息库中所有信息来收集信息。Oracle Secure Enterprise Search 提供:

- 多线程 Java Crawler
- 在多台机器上分发 Crawler 的能力
- 数据抓取框架的安全, 包括 HTTP 基本认证和证书认证, 以及上面集成安全支持部分提及的 ACL 收集
- 高效的过滤。对于二进制文档格式——Office、PDF 文档等而言, 需要过滤器从二进制文档中提取纯文本信息。过滤技术自动辨识文档类型并启动正确的过滤器产生适合索引的文本数据和元数据。在 SES 中, 一百多种常用文件类型有对应的过滤器。

Oracle SES 使用抓取——索引——搜索模式, 确保数据收集一气呵成。文档经过分析, 但是仍然保留原来的物理路径和名称。数据抓取可以安排在低负载时间段进行。

灵活的 API

Oracle Secure Enterprise Search 10g 有一套基于 Web 服务的查询 API，让您创建能够处理任何类型信息的定制搜索。Web 应用程序能够使用这些 API 检索和显示查询结果，让您将 Oracle Secure Enterprise Search 的结果嵌入您的应用程序中。

Secure Crawler SDK 让您指定如何进行定制数据源的数据抓取。Oracle Secure Enterprise Search 可以配置调用您的定制数据抓取器以抓取哪些 Oracle 标准组件中不支持的数据源。

网上下载的 Portlet 让您将 Oracle SES 集成到您的 Oracle 10g Portal 中。

新显示 URL 的特性使搜索可以与您的基于浏览器的应用程序无缝集成。当您想在应用程序环境中显示该应用程序的搜索结果时，Oracle SES 让您设置一个 URL，指定如何在本地调用主应用程序显示结果。

支持的平台

- Linux 和 Microsoft WindowsNT/2000/XP 以及其他常用平台

面向国际市场

- 搜索所有能用 Unicode 2.0 表示的语言
- 对主要语言，包括日语、汉语、韩语、阿拉伯语、希伯来语、西欧和东欧语言等，提供高级支持

联合查询

联合查询允许同时查询 Oracle SES 进行数据抓取的信息库和自己进行数据抓取和索引的异构数据源。很多原因可能会导致信息库不适于数据抓取。如果数据源已经实现了某种搜索，您也许不想从头索引其内容。数据源也许是别人管理的，您也许无法进行数据抓取。更新速度也许太快，无法计划数据抓取。联合搜索的能力补充了 Oracle SES 的集中数据抓取：

- 数据源保留它们自己的全文索引或者进行它们自己的数据抓取
- Oracle SES 从最终用户处接受搜索关键词并把它们分发到其他数据源。这些数据源实现一个使用 Oracle SES 查询 API 的 Web 服务，把结果传回 Oracle SES 显示。样例程序可以下载。

总结

Oracle Secure Enterprise Search 10g 能让您减少花费在您公司信息库上寻找相关文档的时间。它通过 Web 风格的搜索对您的企业内网进行数据抓取、索引和搜索。它让您不再需要对难用的低端的 API 进行编码。它通过提取能在门户应用程序中使用的有价值的原数据来组织多个信息库的内容。

- 它有默认安全机制，能够覆盖您所有的数据源
- 毫秒级查询性能提供了极好的用户体验和优异的搜索质量
- 它能立竿见影，部署周期短，投资回报快

- 高级搜索是简单的基于网络的管理界面，结合了字段元数据搜索和全文搜索。高性能可扩展的数据抓取保证了超大文档集的搜索。Oracle SES 有灵活的 API，能够根据您的特定环境进行配置和扩展
- 它由世界最大的企业软件公司支持，面向国际市场

详细信息

了解有关Oracle Secure Enterprise Search 10g的详细信息，请访问：

<http://www.oracle.com/database/secure-enterprise-search.html>