

安全企业搜索

版本 1 (10.1.8)

Oracle 技术白皮书
2007 年 7 月

安全企业搜索

执行概要.....	3
引言.....	3
ORACLE 的 SES.....	4
SES 结构.....	6
搜索器.....	8
搜索器插件 API.....	9
WEB 服务 API.....	10
管理.....	12
搜索质量.....	13
安全搜索.....	15
基于管理的身份验证.....	15
自助式身份验证.....	16
自定义代理或搜索器.....	16
SES 方法.....	16
收集步骤.....	16
分析步骤.....	17
使搜索结果可搜索.....	17
维护步骤.....	17
SES 版本 10.1.8 中的新内容.....	18
安全插件体系结构.....	19
其他改进.....	20
总结.....	20
参考资源.....	20

执行概要

安全企业搜索 (SES) 是 Oracle 的独立搜索解决方案（具有类似于 web 的性能），可以安全地覆盖所有企业数据源，并且易于使用和部署。SES 可提供：

优秀的搜索质量

亚秒级查询性能

高度安全的信息收集、索引和搜索过程

随取随用的用户体验

引言

许多企业拥有的文本数量要远多于结构化数据，但这些文本数据却一直无法得到充分利用。

Internet 搜索已表明，搜索技术可以增加信息量。如果没有搜索引擎，Internet 上仍然有数亿的 web 页，但冲浪者必须事先知道 URL 或者浏览所有目录，才能找到感兴趣的页面。显然，正是搜索技术让 Google 流行起来，让 Internet 更加有用，即使其上的信息量在快速增加。

此外，企业中的信息也比较分散；但是，企业至今也没有得益于优秀的搜索技术带来的信息增长。这在很大程度上归因于内联网和互联网之间的差异。例如：

Internet 上的信息绝大部分都包含在 web 页中。而在内联网中，信息（数据和内容）分散在 web 页、数据库、邮件服务器或者其他协作软件、文档库、文件服务器和桌面中。内联网搜索引擎必须能够通过同一接口搜索企业的 web 内容、应用程序、数据库和邮件。全面、可跨结构化和非结构化数据源以及能够搜索每个角落（深层内联网）是内联网搜索的要点。

Internet 中的所有信息都是公开可见的，与之不同的是，内联网信息需要保密。不同的用户拥有不同的信息访问权限，并且信息资源通常由口令保护。

内联网搜索引擎必须能够实现安全性。如果用户没有被授予查看文档、电子邮件或记录的权限，则甚至该记录是否存在也不应让其知道。访问权限可以更改，并且对不同基础信息存储所作的权限更改必须迅速传播到搜索引擎。

Internet 搜索引擎（如 Google）可以使用 URL 在 web 页之间提供的链接在给定搜索中推断出文档的重要性或相关性。遗憾的是，内联网资源不能通过 URL 链接始终指向彼此：以 PDF 格式创作的文档可能无法链接到其描述的客户的数据库记录。因此，对于内联网搜索来说，需要使用不同的技术来获得高相关性。

尽管关键字搜索可以实现信息增加，但企业通常需要由自己的业务模型指定的更复杂的查询。内联网的搜索引擎应该面向问题，而不是显示几十条搜索结果，并且应该提供分析功能，以允许用户超越关键字搜索并询问复杂问题，而不用触及深层内联网中的隐藏信息。

不同的内联网用户不仅具有不同的资源访问控制权限，而且还根据工作职能具有不同的信息需求。必须个性化搜索结果以满足这些需求。

内联网搜索必须是多语言的。企业在全球的办事处都使用以本地语言表述的内容。

内联网需要更高级别的服务，并且内联网搜索产品的健壮性必须能够满足任务关键的企业软件。

内联网搜索软件必须易于使用和管理。

SES 解决了在公司多个分散的信息库中查找相关信息的问题，并提供了非常直观的搜索和管理界面。

Oracle 的 SES

安全企业搜索：Oracle 的内联网搜索解决方案。

Oracle 开发文本和信息检索技术的时间已逾 15 年。Oracle 数据库一直在使用 Oracle Text（一个综合 API）的基础功能。Oracle9i 中引入了 Oracle Ultra Search，用于实现跨不同信息库的门户搜索，并且可用于 Oracle 数据库、应用服务器和协作套件。构建于这些产品之上，Oracle 的最新企业搜索技术还添加了几个关键功能。

简单性。一个随取随用的简单 web 用户界面（用于搜索和管理）— 它兼具用户在 Internet 上搜索时所需的简洁外观和易用性。

综合性。能够搜索所有数据源 — web 页、文件服务器或桌面驱动器中的文件、数据库、应用程序、邮件服务器和群件等。

与旧式信息库的连通性。SES 允许公司访问他们最有价值的资产 — 先前驻留在专用信息库中、有关特定业务、流程、产品、客户和文档的信息。连接器包括针对 EMC Documentum、FileNet、OpenText LiveLink、Microsoft SharePoint、IBM Lotus Notes、Oracle 的电子商务套件等的接口。

安全性。能够安全地搜索由口令保护的数据源。Oracle 的搜索技术可以提供基于一次性登录 (SSO) 的安全性（在可用的情况下），还可以采用特定于应用程序的安全性（在 SSO 不可用的情况下）。

高质量的搜索结果。为内联网带来了用户与 Internet 搜索关联的高级相关性。

超越关键字。随着信息量的增加，甚至具有较高的相关性，关键字搜索模式的返回结果开始不断减少。用户需要高级搜索技术，例如，能够为迭代导航分类和集群搜索结果。

SES 已经完全全球化，可以在所有主要语言中执行搜索，包括西欧、中文、日语、韩语、阿拉伯语、希伯来语等。

SES 很健壮，并且已经针对企业进行了强化。通常，一个普通的 Linux 机器每秒可以执行几百个查询。典型的企业内联网通常是 TB 级的，Oracle 的搜索基础架构已经针对多 TB 加载重复部署¹。

¹ IronMountain 案例研究 www.oracle.com/technology/products/text/pdf/oracle_ironmountain.pdf



图 1 显示了运行中的 SES 示例。

SES 结构

Oracle 的安全企业搜索是一个独立的搜索服务器；它以“黑盒”方式运行，可以索引来自搜索器的信息并提供结果。它具有自己的用户界面和管理方式；例如，它不需要您使用 SQL 编程或者以 DBA 身份进行管理。

从体系结构来看，如图 2 所示，该产品由五个不同的组件构成。

搜索器 (Crawler)。SES 搜索器是一个根据设置的日程由 Oracle 服务器激活的 Java 进程。激活后，搜索器会生成一些处理器线程（其数量可配置），读取不同数据源中的文档。搜索器将映射和分析链接关系，以防止循环操作和错误返回。如果搜索器在搜索过程中遇到嵌入的非 HTML 文档，就会使用过滤器自动检测文档类型，并过滤和索引文档。

数据库。Oracle10g 数据库包含 SES 信息库，该信息库用于存储 SES 索引的信息库的相关信息以及搜索引擎“索引”（搜索器收集的信息，由 Oracle10g Text 过滤并索引）。

搜索 UI 和 API。SES 为服务器提供了现成的用户界面。还提供了一个 web 服务 API，用于构建查询索引数据所需的自定义应用程序，并包含有基本搜索表单、高级搜索表单、查询结果显示、帮助页面、反馈页面、URL 注册等界面。

管理工具和界面。SES 管理工具是一个基于浏览器的应用程序，可用于配置和调度搜索器、配置服务器、运行多个报表功能以及其他类似的任务。

联合器 (Federator)。SES 还提供了将查询联合到其他引擎的功能，以实现这些引擎自己的搜索 — 邮件服务器、Internet 搜索引擎和特定应用程序。这些结果可以与 SES 服务器的内部索引提供的结果组合在一起并一同显示。

SES 基于 Oracle10g 数据库和一个“独立”版本的 Oracle J2EE 容器“OC4J”（用于实现 web 服务器来处理 HTTP 请求）。在安装期间，SES OC4J “应用程序”（包括搜索和管理环境）将部署在该 J2EE 运行时环境中。Oracle 数据库以自定义方式构建 — 可以配置和调整以满足搜索引擎的特殊需求。

SES 数据库和 web 服务器安装在同一台计算机上。原则上，可以将 SES 数据库和搜索/管理应用程序移到不同的计算机上，但目前不受 Oracle 官方支持。

无需任何自定义 SES 即可连接到 Oracle 的 SSO 基础架构 (OID) — 只需指定几个简单的 Oracle Internet Directory (OID) 连接参数。保护搜索所需的用户和组在 OID 中存储和管理。

SES “OC4J”应用程序通过 JDBC 连接到数据库 — 连接由以下文件定义：

listener.ora、tnsnames.ora、sqlnet.ora（Oracle Net 配置文件）。

路径：OES_HOME\network\admin

data-sources.xml（定义管理应用程序的数据库连接）。路

径：OES_HOME\oc4j\j2ee\OC4J_SEARCH\config

search.properties（定义搜索应用程序的数据库连接）。路

径：OES_HOME\search\webapp\config

上述文件将在安装期间自动配置，请不要更改。如果在安装期间发现其他 Oracle 软件，SES 安装程序将创建一个具有自己的网络配置/端口的新监听程序。

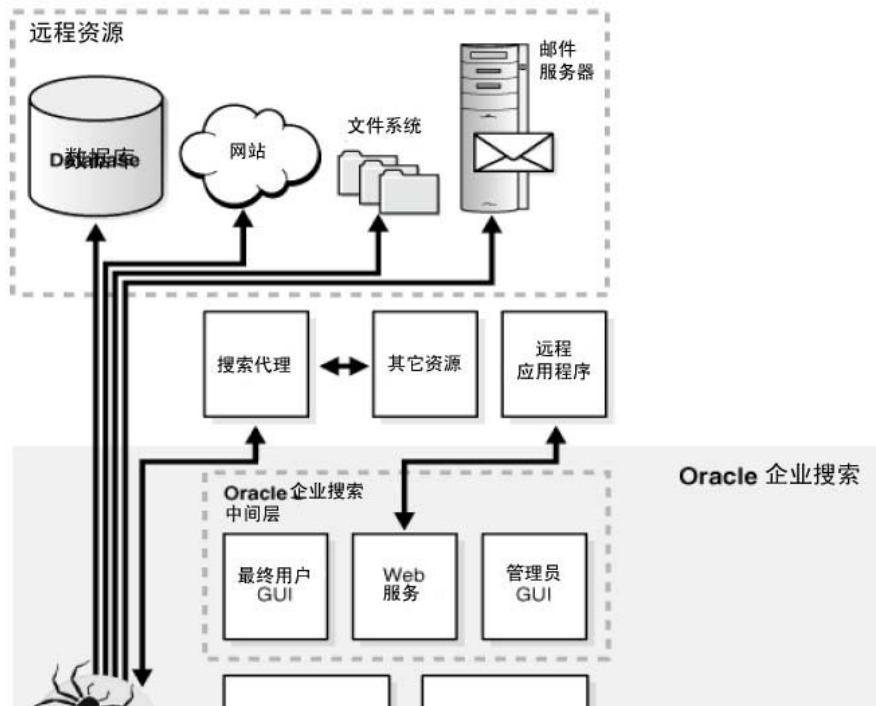


图 2. SES 体系结构

以下部分更详细地说明了其中一些组件和其他方面。

搜索器

SES 搜索器是一个多线程 Java 应用程序，负责从您在配置期间指定的数据源中收集文档。要搜索其他信息库，SES 搜索器允许您定义特定“数据源”。数据源是一种标识信息库的逻辑结构。您可以将单个物理信息库（如数据库）映射到多个数据源（数据源也是您定义元数据的粒度）。SES 可识别以下标准类型的数据源（可通过插件获得更多数据源，如下所示）：

Web 站点 — 利用 HTTP 协议将 web 站点定义为数据源，如 www.oracle.com。

数据库表 — SES 可以搜索 Oracle 数据库和其他支持 ODBC/JDBC 标准的关系数据库。要搜索的数据库表可以驻留在 SES 自身的数据库实例中，或者成为通过网络访问的远程数据库的一部分。SES 允许搜索全文本列和“字段文本”列。字段文本列允许您将数据库列映射到 SES 属性（如 author、title），以创建一组针对数据库内容的索引。

文件 — 文件必须可由搜索计算机直接访问。还可以搜索远程文件，只要它们可以通过 file:// 协议搜索到。

文件必须可由每个搜索器计算机（在本地或通过网络远程）访问。

电子邮件。SES 可以连接到 IMAP 电子邮件服务器，并索引用户的所有电子邮件。要索引邮件列表，可以选择创建一个订阅感兴趣邮件列表的特定 IMAP 帐户。

OracleAS Portal 实例

SES 使用第三方过滤器从文档中检索文本和元数据，并自动识别文档类型。过滤器可处理常见文档格式，如 PDF 和 MS-Office。还支持过滤用 ZIP 实用程序压缩的文档。

为了维护全面、最新的搜索结果，SES 使用了*同步计划*。例如，电子邮件搜索结果可以不断更新，但不需要频繁收集已发布的内容。每个同步计划可以附加一个或多个数据源。

为了限定只搜索企业网络的特定部分，或者确保搜索不会返回错误内容，并追踪指向内联网以外位置的链接关系，SES 可让您为搜索指定所谓的“包含”域和“排除”域。SES 支持“实例快照”，您可在其中创建主 SES 实例的只读快照，以进行查询处理或备份。在主实例损坏并且您希望使用快照作为新的主实例时，这非常有用。

还可以指示 SES 搜索器收集 URL，而无需进行索引。这个*数据采集模式*允许您检查文档 URL 及其状态，移除不需要的文档，以及启动索引。

搜索器插件 API

搜索器可以通过“插件”扩展（插件是 Oracle 提供的或客户开发的 Java 类，与搜索应用程序运行在同一个 OC4J 容器中）。SES 为许多企业内容源（如 EMC Documentum、Lotus Notes 和 Microsoft Sharepoint 等）提供了一系列现成插件（又称为连接器）。

插件通过 SES 管理 GUI 部署 — 在定义后，它们将作为新数据源列出。在进行配置后，插件可以将数据提供给搜索器，并且可以像其他数据源一样进行索引。

从技术上讲，插件负责收集指向要索引文档的 URL — 它们将这些 URL 传递给搜索器以进行索引。以下是一些方法的简短示例，如果客户希望实现自己的插件，则必须提供这些方法：

open: 初始化

startCrawling: 进行获取文档所需的设置

stopCrawling: 不再发送文档

isDeltaCrawlingCapable: 指定代理是否可以仅返回在特定日期后创建的文档

fetch: 返回下一个文档 URL

received: 确认已经获取一个 URL

getCredential: 返回访问该 URL 所需的任何用户名和口令

getCookies: 返回访问 URL 所需的 cookie 流

getAttributeLOV: 返回所有源属性的值列表

close: 关闭并清理

WEB 服务 API

诸如 SES 之类的搜索引擎通常集成在现有的客户 web 站点和门户站点中。理想情况下，最终用户从搜索表调用搜索，甚至不会意识到 Oracle SES 在后台处理他们的搜索请求。结果列表的“外观”必须与调用搜索的门户站点相符。为此，SES 提供了基于 SOAP 和 WSDL 之类的标准的 web 服务接口。

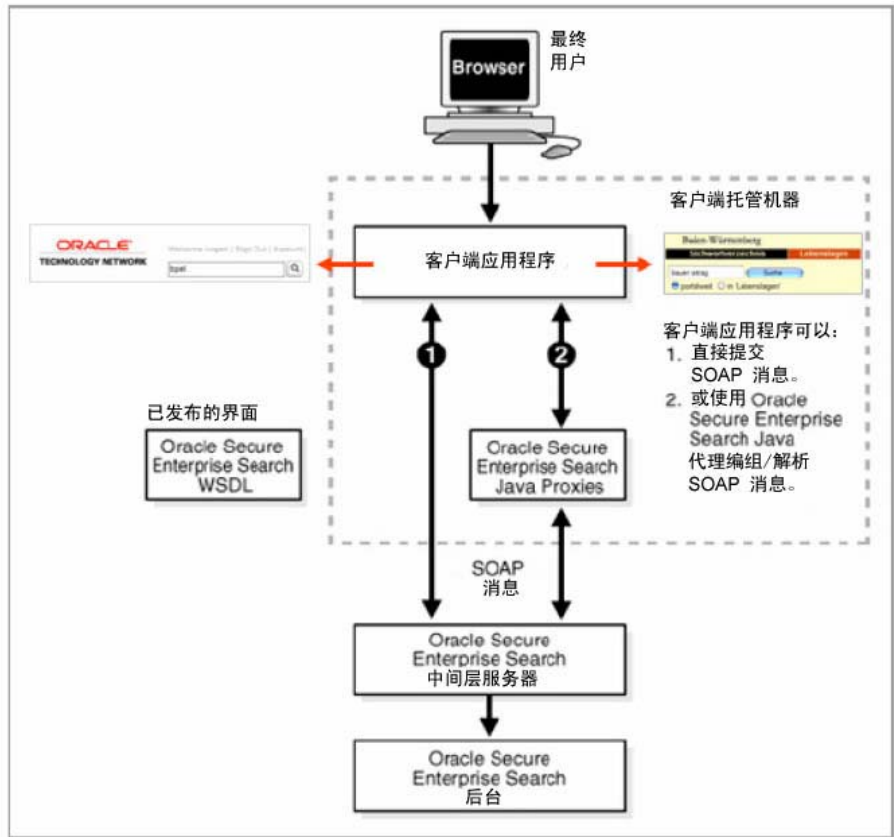


图 3: SES Web 服务体系结构

在上述示例中，最终用户在输入域（“搜索框”）中输入了搜索项。搜索请求将从 CMS 应用服务器直接发送到 SES web 服务。SES 将执行搜索，并通过 SOAP 以 XML 的形式将结果返回给调用应用程序。结果将以嵌入方式在应用程序中显示。

SES 没有使用 UDDI 信息库 — WSDL 描述可以直接从服务器获取。

以下代码片段演示了使用 API 是多么简单：

```
import oracle.soap.transport.http.OracleSOAPHTTPConnection;
import oracle.soap.encoding.soapenc.EncUtils;
import oracle.search.query.webservice.client.*;

public class TestWS
{
    public static void main (String[] argv)
    {
```

```

try
{
    OracleSearchService search = new OracleSearchService();

    // Add your own code here, for example to populate
    // the query string.

    // Set SOAP URL.The URL is
    http://<host>:<port>/search/query/OracleSearch
    stub.setSoapURL("http://oes-serv-
    example:7777/search/query/OracleSearch");

    String queryString = "oracle";

    //
    // Do a simple search for the queryString we set up above
    //
    OracleSearchResult result = stub.doOracleSimpleSearch(
        queryString, // query
        new Integer(1), // startIndex
        new Integer(3), //
docsRequested
        Boolean.FALSE, // dupRemoved
        Boolean.FALSE, // dupMarked
        Boolean.TRUE); //
        // Get the result set
        returnCount

    ResultElement[] resElemArray = result.getResultElements();

    // Loop through the results displaying the document title
    for (int i=0; i<resElemArray.length; i++)
    {
        System.out.println("Document Title:
        "+resElemArray[i].getTitle());
    }
}
catch(Exception ex)
{
    ex.printStackTrace();
}
}
}

```

管理

管理工具是一个 web 应用程序，它允许管理员：

定义和搜索数据源。

定义搜索器参数，如 URL 边界规则、搜索深度、语言和代理设置等。

创建和修改搜索器的计划。

设置查询选项 — 查询选项允许用户限制他们的搜索。可以将搜索限制到文档属性（如 **title**、**author**）和数据组。数据源组是公开给搜索引擎用户的逻辑实体。在输入查询后，系统会要求搜索引擎用户选择一个或多个数据组来进行搜索。每个数据组都包含一个或多个数据源。

调整搜索结果列表的相关性排序 — SES 允许管理员改变文档在搜索结果列表中的排序。利用这一特性提高重要文档的匹配度，使它们更易于找到。

为特定搜索项定义建议链接。

为特定搜索项定义备选字。

为特定数据源设置身份验证机制。

管理搜索元数据的备份和恢复。

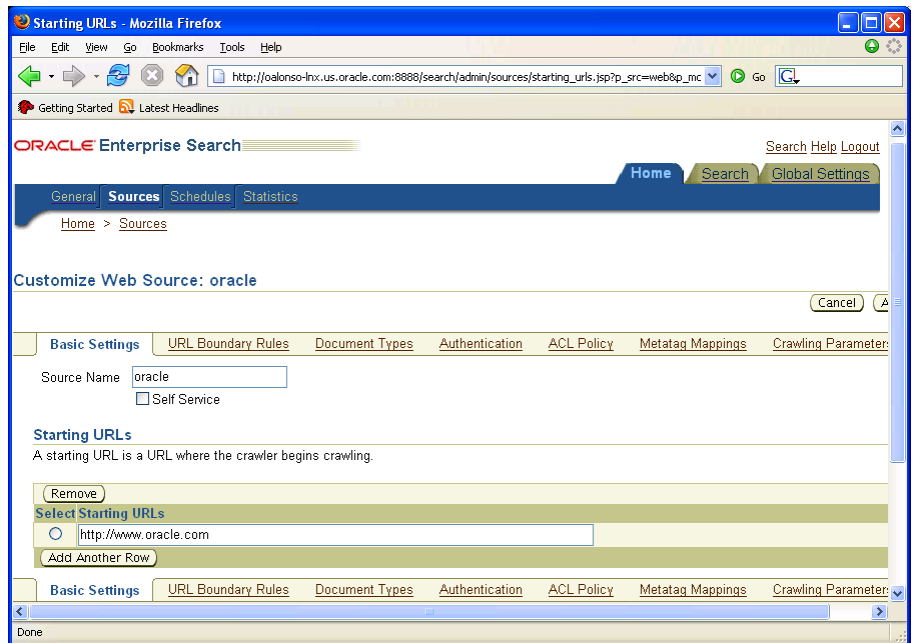


图 4: SES 管理界面

搜索质量

搜索质量或查找相关信息的能力是所有搜索引擎最重要的特性之一。SES 使用了大量技术来提供优秀的搜索质量。

SES 将以下技术用于搜索和索引处理的各个阶段，以增强整体搜索质量：

元数据处理。识别页面和文档中的元数据（如标题、作者、描述、标题行、电子邮件和链接锚文本）非常重要。

重复消除。企业内联网中有许多重复内容。同一幻灯片、web 页和文本文档的副本遍布各处。有时是人为生成多个文件，有时是服务器复制内容以生成镜像。其他重复问题还包括不同的版本、格式、HTML 样式、特定于站点的链接、联络信息等。在任何情况下，用户在搜索时应该只能看到文档或 web 页的一个副本。

完全消除重复有助于在搜索阶段、甚至在索引文档之前识别和移除重复内容。

链接分析。一个用于提高相关性的最广泛使用的技术就是链接分析。简单来说，其思想就是通过分析 web 集合的链接结构来发现权威页面。被许多页面链接的页面很重要。由高链接匹配度页面链接的页面也很重要。目前有许多算法，如 HITS 和 PageRank。但 SES 拥有自己的算法实现。

管理员还可以使用两个额外特性来控制相关性：备用链接和建议字。在注册众所周知的权威页面时，备用链接比关键字有用。当用户搜索这些关键字时，这些链接将显示在搜索结果页面的顶端。建议字可以将用户搜索项映射到同义词。例如，将“移动电话”映射到“手机”或“无线电话”。

如果用户在拼写查询项时遇到困难，拼写检查特性将根据字典和搜索数据中的数据建议正确的拼写。

除了 SES 的所有搜索特性以外，还可以同时组合浏览和搜索特性。您可以单击浏览链接来浏览 SES 在搜索后自动创建的所有目录。如果要尝试浏览可以搜索的所有内容，这是一个很好的切入点。当然，您随时都可以在搜索框中搜索目录。

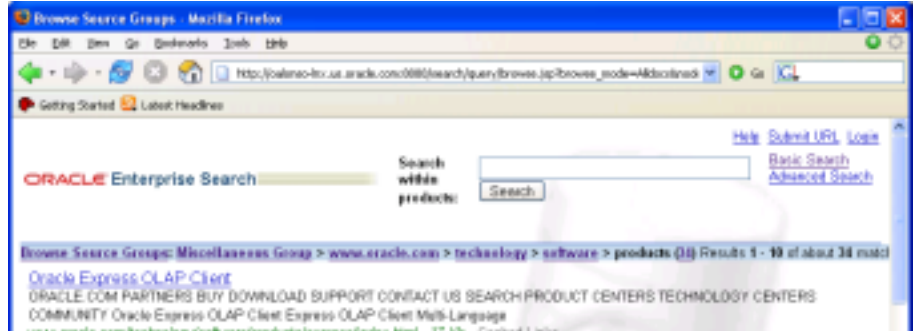


图 5: 浏览内容目录

安全搜索

SES 的特性是 **安全搜索**— 用户登录并查找不公开文档的能力。为此，SES 应该具有安全搜索功能，以及在数据源以外存储访问控制列表 (ACL) 信息的能力。

SES 可以与 Oracle Internet Directory (OID) 集成 — OID 是一个轻型目录访问协议 (LDAP) 服务器，可提供网络身份验证功能。用户登录到 OID 后，SES 将仅允许他们查找有权查看的文档。

搜索器可以通过各种方式处理安全的数据源。如果数据源由同一个（或不同的）OID 服务器保护，则不需要这么做。任何数据源都能够以安全方式搜索，只要它由以下其中一个方式保护：

1. OID 身份验证
2. HTTP 基本身份验证
3. 基于表单的身份验证

为安全搜索器提供访问凭证的方式有多种。

基于管理的身份验证

定义数据源之后，管理员可以输入授权口令（用户口令或“超级用户”口令）。这将用于收集数据源的信息。可以为数据源定义一个 ACL，决定谁可以搜索信息。

自助式身份验证

定义数据源之后，管理员可以设置数据源，但不提供任何用户名或口令。随后，用户可以登录并提供自己的访问凭证。之后，系统将创建一个特定于该用户的数据源，并且他们是可搜索该信息的唯一用户。

自定义代理或搜索器

自定义代理是一个 JAVA 模块，可用于搜索用户指定的任何数据。代理可传回要索引信息的指针 (URL)，并有选择地为每个文档指定 ACL。这就为访问控制带来了极大的灵活性。

SES 还提供了 QTA (查询时身份验证) API，以允许客户在查询时更好地控制搜索结果。SES 使用该技术作为主接口，从而根据身份验证访问过滤文档。

SES 方法

使用 SES 需要遵循哪些步骤？SES 搜索引擎遵循四个逻辑步骤来提供全局搜索 — 收集、分析、使之可查询以及维护。这些步骤不是虚构的，您确实能够在大多数组织的业务流程中看到它们。

收集步骤

收集查询结构化关系数据库和非结构化文件中的信息、WORD 处理文档、电子表格、演示文稿、电子邮件、新闻信源、ADOBE ACROBAT 文件以及 WEB 页。SES 通过“搜索”企业内联网并查看公司各种信息库 (数据库、WEB 页、IMAP 邮件服务器等) 中的所有信息，来收集该信息。

在收集过程中，会分析链接关系，以防止循环操作和错误返回。因此，SES 管理员可以更轻松地保持搜索结果完整并且最新。

分析步骤

在分析阶段，SES 会查看已收集信息的内容和结构。要搜索信息，必须对其进行索引。在分析阶段，SES 使用 ORACLE TEXT 引擎、通过创建一个集成索引从已收集信息中提取内容和结构，从而有效地“规范了”结构化和非结构化数据。ORACLE TEXT 索引包含完整的字列表和其他信息。

在索引期间，第三方过滤软件将从文档中提取文本和元数据。该过滤技术可自动识别文档类型，调用正确的过滤器，并生成可索引的文本和数据。还支持几个预先定义的元数据域，包括 AUTHOR、DATE 和 TITLE。过滤器支持最常见的文件类型，如 MS OFFICE 和 PDF。

与某些文档管理系统不同，SES 的收集和分析不会带来任何影响。所分析的信息和文档驻留在自己名称下的原始位置，而非以物理方式移动文档。

在典型的 WEB 搜索技术中，会返回几百个结果。随着信息库数量的增加，文档的相关性排序功能降低。SES 使用获奖的相关性排序工具 ORACLE TEXT 来确保用户始终能够实现大海捞针的神话。

使搜索结果可搜索

“使之可搜索”的功能可以通过编程方式提供对已索引的所有信息的访问。ORACLE SES 为该功能提供了一个 WEB 服务 API。将搜索项传递给该查询 API 可找到所有相关文档，无论它们是存储在 WEB 服务器、数据库还是应用程序中。客户可以使用 SES API 将全局搜索集成到自己的 WEB 页或应用程序中。

维护步骤

维护步骤可确保搜索结果不断更新。SES 可让您根据不同的计划安排从多个 WEB 站点和信息库中收集信息。例如，IMAP 消息处理服务器可以不断更新，但不需要频繁收集已发布的内容。SES 通过提供简单、直观的实用程序来维护内容，这些实用程序可让管理员轻松地跟踪由于增长或并购而添加的新内容。

SES 版本 10.1.8 中的新内容

健壮的连接器框架

用户搜索引擎（如 GOOGLE 和 YAHOO）在 WEB 服务器上索引和搜索的大多是 HTML 页面。企业搜索引擎还必须索引门户、文档管理系统、自定义应用程序以及其他软件应用程序和系统。ORACLE SES 提供了一系列内置“连接器”（连接器是基于 SES 插件 API 的 JAVA 类），用于访问目前最常用系统中存储的内容。

新的连接器系列提供了对驻留在以下专用系统和应用程序中的文档的访问：

WINDOWS NT 文件系统 (NTFS) — NT 文件共享可以通过网络连接索引，而且不必位于 SES 主机上。SES 通过读取组和用户的访问信息并将其存储在搜索引擎索引中，来提供强大的访问控制。

对于安装在 UNIX 操作系统上的 SES，在要索引的 NT 文件系统所在的 AD 域中安装一个较小的代理进程。代理会将内容、元数据和访问控制信息发送给 SES 计算机中的连接器（代理协议基于 HTTP，并且可以使用 HTTPS 加密）。必须启用 MICROSOFT IIS，代理才能工作。

EMC DOCUMENTUM CONTENT SERVER — 索引

CONTENTSERVER DOCBASES 文件柜和文件夹中的文件。原生身份插件允许 SES 根据 DOCUMENTUM 中的权限设置仅显示用户有权访问的文档。支持高效的重新搜索 — 如果 DOCUMENTUM 中的文档更改或移动，则只需重新索引。

IBM LOTUS NOTES — IBM LOTUS NOTES DOMINO 服务器实例上的 NOTES 数据库（NOTES MAIL 和自定义应用程序计划在未来版本中推出）。连接器可以自动浏览 NOTES 服务器实例上的所有 NOTES 数据库。SES 提供了一个 NOTES 身份识别插件，以使用 NOTES 目录进行 NOTES 本地用户和组的身份验证。

FileNet Content Engine — 索引内容引擎实例中的对象存储和文档。可以自动浏览对象存储以检索所有文档。

FileNet Image Services — Image Services 库中包含文本（时间戳、透明文本和便笺）的文件夹、文档和图像批注。原生身份插件。

Open Text Livelink — LiveLink Enterprise Server 工作区中的文件夹、文件、项目和任务列表。提供了针对 Livelink 的身份插件。

Microsoft Exchange — 索引 Exchange 2000 和 2003 存储中的电子邮件、附件、日历项和相关元数据属性。支持高效的增量重新搜索。需要 Microsoft IIS 以及将与 Exchange 服务器安装在同一个 Windows 域中的 Oracle “代理”软件（版本 10.1.8 随附的代理可以在 Exchange 主机和 SES 主机间传送内容和元数据）。

所有连接器都是预先配置的（上面的某些连接器可能需要额外许可），SES 与旧式信息库之间的访问控制集成均由连接器提供。

Oracle SES 还可以搜索许多 Oracle 内部数据源：

OracleAS Portal 页面组、页面和项。

Oracle Collaboration Suite Content Services 和 Calendar

Oracle ContentDB — 文件夹、文档和目录。支持高效的重新搜索：在增量搜索期间，只会重新索引内容更改的文档、更改的元数据/目录元数据以及移动的文档。

Oracle 电子商务套件 11i — 允许搜索 Oracle 数据库基础 11i 中的视图或查询。视图或查询中的每个记录都被视为一个单独的文档。

Oracle Siebel 8.0 — RSS 信源。

安全插件体系结构

现在，安全企业搜索可以直接集成到第三方访问控制和身份管理解决方案中，包括 Microsoft 的 Active Directory。不需要将用户或组与 Oracle Internet Directory 同步。SES 可以通过新的授权 API 和身份“插件”体系结构直接访问 Active Directory（不需要额外编写代码）。此外，SES 还为 Oracle 的 Internet Directory 和 Microsoft 的 Active Directory 提供了插件。新的体系结构甚至允许客户构建自己的“身份插件”（提供用户和组信息），以使用专用（非 LDAP）安全模式搜索数据源。

其他改进

一个新的建议内容特性可让您在搜索结果屏幕中索引并显示实时内容。在搜索结果列表中显示内容之前，可以应用一个样式表。

一个用于通过 OracleAS 一次性登录配置安全搜索的简化方法；一个“标题覆盖 (title fallback)”特性，可以使用更有意义的标题覆盖在搜索期间搜索到的默认文档标题；一个更简单的联合数据源配置，以及不区分大小写的相关性增强（在输入“oracle”时，也会关注包含“Oracle”的文档）。

支持从 SES 第 1 版 (10.1.6) 升级。

总结

企业内联网与互联网不同 — 其中的信息来自多种不同类型的数据源；搜索需要访问口令保护的内容；确定内联网文档重要性的技术与互联网使用的不同，并且有效的内容通常不在结果列表中。安全企业搜索旨在为内联网实现用户从互联网上获得的信息增加。通过部署 Oracle 的搜索解决方案，您不仅可以安全有效地查找信息、缓解信息过载，还可以发掘深层内联网中难以触及的隐藏信息。

参考资源

[1] SES 主页：<http://www.oracle.com/technology/global/cn/products/ses/index.html>

[2] SES 业务白皮书

[3] SES 特性概述



Oracle 企业搜索白皮书

2007 年 7 月

作者: Stefan Buchta

合作者:

Oracle Corporation

全球总部

500 Oracle Parkway
Redwood Shores, CA 94065

U.S.A.

全球咨询热线:

电话: +1.650.506.7000

传真: +1.650.506.7200

www.oracle.com

Oracle Corporation

www.oracle.com

Oracle Corporation 提供的软件支持互联网。Oracle 是 Oracle Corporation 的注册商标。本文提及的各种产品和服务的名称可能是 Oracle Corporation 的商标。其他所有提及的产品和服务名称可能是其各自所有者的商标。

版权 © 2007 Oracle Corporation

保留所有权利。