

# EKG 内网文档搜索引擎

**Enterprise Knowledge Gateway (EKG) System**

## 用户使用手册

**User Manual**



**灵玖中科软件（北京）有限公司**  
LING-JOIN Zhongke Software (Beijing) Co. Ltd.

## Document Information

Document ID	LINGJOIN-ELINT-WHITEPAPER	Version	V1.0
Security level	Confidential 秘密★	Status	Creation and first draft for comment
Author	灵玖中科软件（北京）有限公司	Date	2010/9/21
Publisher	/	Approved by	灵玖软件

## Version History

Note: The first version is "v0.1". Each subsequent version will add 0.1 to the exiting version. The version number should be updated only when there are significant changes, for example, changes made to reflect reviews. The first figure in the version 1.x denotes current review status by 1. x denotes review process has passed round 1 etc .Anyone who create, review or modify the document should describe his action.

Version	Author/Reviewer	Date	Description
V0.1	王毅伟	2010/9/15	Creation and first draft for comment.
V1.0	王毅伟	2010/9/21	first complete draft for comment.

## 目 录

一、EKG 简介 .....	4
二、EKG 优势 .....	5
三、用户操作介绍.....	6
3.1 搜索主界面介绍 .....	6
3.2 输入搜索关键词 .....	7
3.3 搜索结果界面介绍 .....	7
3.4 文章预览页面介绍 .....	8
3.5 右边栏导航条 .....	11
3.6 高级搜索 .....	14
3.7 用户权限介绍 .....	14
3.8 使用小技巧 .....	15

## 一、EKG 简介

内网文档搜索引擎系统是一款针对政府、企业或机构局域网内部文档管理、搜索与挖掘的软件产品，凝聚了灵玖中科软件（北京）有限公司多年自然语言理解与精准搜索引擎的技术积累。

随着信息化的高速发展，社会各界、政府及企业中的电子文档（word、excel、ppt、txt、pdf 等）也变得越来越。大量政府和企事业单位的文档是包含着客户关系、产品信息、市场情报、策划思想等软资产的载体。文档作为一种信息资源，作为企业生产、技术、科研和经营等活动的真实记录和一项基础性工作，同时作为与企业同步发展的无形资产，在企业管理等方面积极地发挥着重要作用。

政府/企事业单位的电子文档日益增多，文档格式的多样化、文档内容分散存储于各个不同的电脑上等问题，文档的管理存在三大隐患：

### 1、文档总量大，单篇文档的篇幅较长，检索与管理的效率极其低下

据统计，一台工作电脑上，大小、版本不同的电子文档一般都在数千到数万个以上，这样庞大的文档量，即使浏览一遍标题也需要花费大量的时间，靠人工有效管理几乎不可能。据统计，用资源管理器等常规手段查找文档，每篇文档平均需要 3 分钟。按用户每天需要查找 10 篇计算，一个人每年需花费 100 小时以上。而对于拥有成百上千电脑的单位，对这些文档的管理已经成为“食之无味弃之可惜”的鸡肋。

而内网文档搜索引擎在上前万文档中查找文件的速度是 0.1 秒钟以内。

### 2、文档分散存储，共享困难，协同工作往往靠上级或者个人协调

目前，除了少量文档通过共享目录、邮件、网络即时通讯软件被少部分员工共享外，大多数电子文档属于某个员工的“私产”，传统共享方式的效率低下使员工共享同事的文档比上互联网查找文档更为困难。

而通过内网文档搜索引擎，大家可以比 Google, Baidu 更快更便利。

### 3、分散存储，缺乏必要的安全保障措施，安全隐患极大

安全隐患主要表现在：

1) 大量文档分散成各台电脑上，一旦有员工离职，相关的大量工作文档被直接带走丢失，或沉落于电脑中，成为“不为人知”的秘密，由此带来大量文档资产的流失。

2) 由于大量文档处于不被管理的无序状态，文档被病毒感染、意外丢失、越权使用等情况屡屡发生，使文档的安全性无法得到保障。

3) 由于工作文档经常会被不断地修订，有些文档多达数十，乃至上百个不同的版本，大量的版本进一步增加了文档管理的难度。为此，很多用户会用新版本覆盖老版本，但是，这种情况下，一旦用户需要恢复老版本，已经很难找回来。

而通过内网文档搜索引擎，大家可以避免以上隐患。

## 二、EKG 功能

内网文档搜索引擎系统包含四大功能：

### 1 自动备份：

每天定时自动采集局域网内部共享的电子文档信息，存储在指定的服务器文件夹内，并保留同一文档的不同版本，基本信息采用数据库的形式进行存储。用户可以通过浏览器随时检索并提取任意版本的文档。

### 2 精准搜索：

内网搜索引擎提供了多种精准搜索手段，其中包括：

- 1) 关键词搜索：搜索任意关键词；
- 2) 空间范围搜索：可以搜索特定员工或特定部门的文档资源；
- 3) 时间范围搜索：可以搜索任意时间段的文档资源；
- 4) 文档格式搜索：支持 Word, ppt, excel, pdf, rtx, html 以及程序代码等主流的文档格式；
- 5) 指定位置搜索：支持标题、正文等不同位置进行搜索；

### 3 知识分享：

内网搜索引擎提供了多种知识分享的方式，用户单位可以根据知识分享记录实行奖惩措施，其中包括：

- 1) 搜索并阅读或者下载任何同事的文献资料，系统自动记录每一次分享的时间、分享者与阅读者；
- 2) 部门统计：统计各部门的知识分享记录，鼓励部门间进行知识分享；
- 3) 人员统计：统计每个人分享出的文档总数，给出排行榜，鼓励个人分享文档；
- 4) 文档 Top10 推荐：推荐关注度最高 Top 10 的文档；
- 5) 贡献人员 Top10 推荐：推荐贡献度最高 Top 10 的同事；
- 6) 阅读 Top10 推荐：推荐阅读文档数量最大的 Top 10 同事；

### 4 安全防护：

内网搜索引擎提供的安全防护包括：

- 1) 针对每个分享的文件夹严格实行权限管理（由所有者自行设定），权限分为三级：个人（仅限于分享者个人阅读下载）、部门（限于同部门分享）、所有（局域网内所有人员分享）；
- 2) 内网隔离措施：内网搜索引擎只能由授权用户在内网进行访问，外网无法登录内网服务，真正物理隔离，确保文档资源不被窃取流失；
- 3) 文档全过程版本备份：文档修改过程中的版本完整备份，实时按需读取恢复数据；

## 三、用户操作介绍

### 3.1 搜索主界面介绍

打开浏览器，输入服务器端 IP 地址，即可访问文档搜索主界面



截图 1：主页面展示



注释 1：公司 LOGO，可在管理员后台更换，详情请看“管理员手册”



注释 2：文档搜索栏



截图 3：高级搜索按钮，详情请看“高级搜索”章节



注释 4：选择搜索文档类型，可在后台定制类型

### 3.2 输入搜索关键词

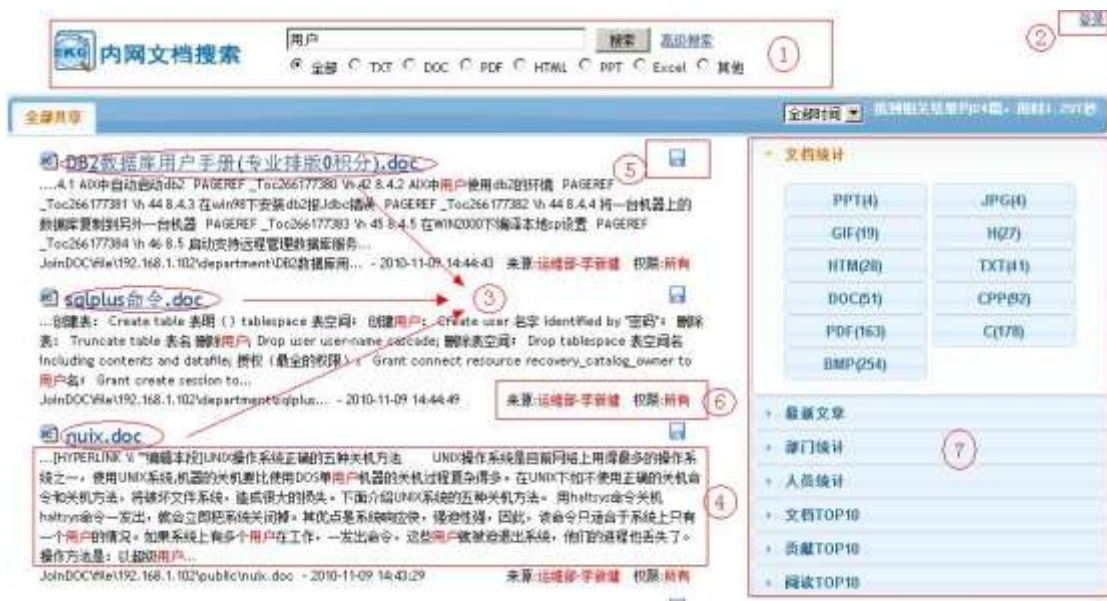
在文档搜索栏输入您需要查找的文档关键字，然后点击 **文档搜索** 按钮或者回车



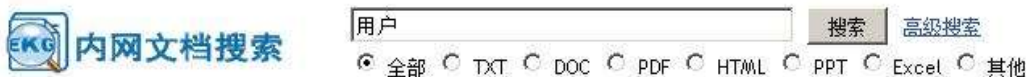
截图 2：输入查询关键字“灵玖软件”

在点击 **文档搜索** 前还可以选择您想搜索的文档类型

### 3.3 搜索结果界面介绍



截图 3：搜索结果页面展示



注释 1：顶部搜索栏，功能同主页

注释 2：用户登录按钮

**DB2数据库用户手册(专业排版0积分).doc**

注释 3：文章标题和文档类型。单击可预览全文，详见“预览文章”章节

...创建表: Create table 表明 ( ) tablespace 表空间; 创建用户: Create user 名字 identified by "密码"; 删除表: Truncate table 表名 删除用户; Drop user user-name cascade; 删除表空间: Drop tablespace 表空间名 including contents and datafile; 授权 (最全的权限): Grant connect resource recovery\_catalog\_owner to 用户名; Grant create session to...

注释 4: 每条搜索文章的标题和局部内容。红色标注为匹配的搜索关键词



注释 4: 文章下载按钮, 点击可下载到本地阅读编辑

来源:运维部-李新健 权限:所有

注释 5: “来源”指由谁分享, “权限”详见“用户权限”章节



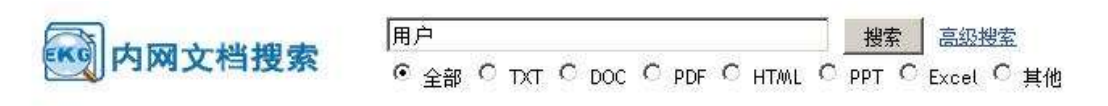
注释 6: 右边栏导航树, 详见“右边栏导航”章节

### 3.4 文章预览页面介绍

点击文章标题, 会进入文章预览页面



截图 4: 文章预览页面展示



注释 1: 顶部搜索栏, 功能同主页



注释 2: 文档模式

1. 文本模式: 用 text 形式显示文章
2. 文档模式: 用 PDF 形式显示文章



注释 3: 下载当前文档到本地



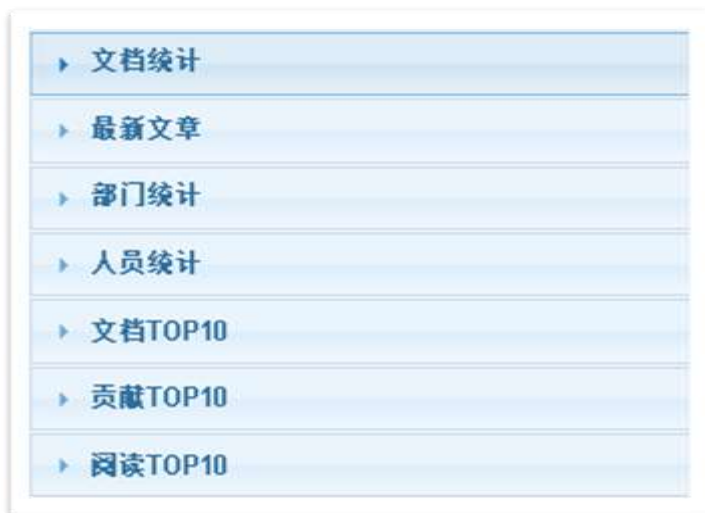


注释 4：阅读器工具条和文档正文

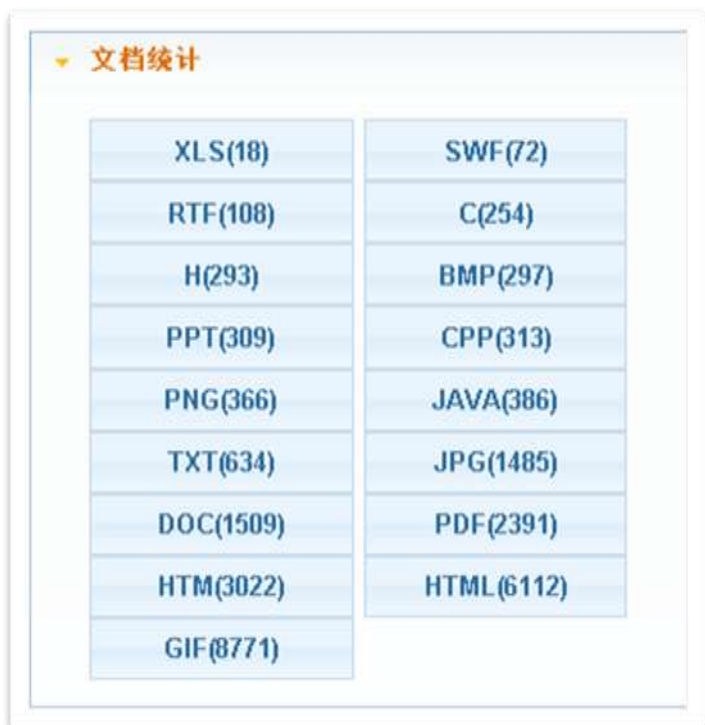


注释 5：文档简单属性表格

### 3.5 右边栏导航条



截图 5: 右边栏导航条展示



文档统计导航条统计了当前系统中储存的各类型文档数

▼ 最新文章

稿件投稿声明.doc
研究与发展.txt
基于层词隐马模型的汉语词法分析 0126.pdf
基于多层隐马模型的汉语词法分析研究 Release.pdf
基于多层隐马模型的汉语词法分析研究 0825-Yhk2.doc
评审结果通知.doc
1460.jpg
1459.gif
1458.gif
1457.gif

最新文章导航条显示了系统最新收录的 TOP10 文章

▼ 部门统计

人事部(0)	财务部(0)
市场部(0)	行政部(0)
销售部(0)	运维部(0)
开发部(26340)	

部门统计导航条是对部门共享文件总数的一个统计

▼ 人员统计

ces(0)	孙健(0)
Kevin(0)	曾飞(0)
王毅伟(0)	赵小宝(0)
高健(0)	刘志华(1173)
于漪泉(25167)	李新健(0)

人员统计导航条是对员工共享文件总数的一个统计

▼ 文档TOP10

1.txt
l_1_1.gif
IBM_中国研究中心_申请流程.htm
NEC(中国)研究院招聘信息.txt
PICSRules_1.1.htm

文档 TOP10 导航条是共享文件被查看和下载量的前 10 位排序

▼ 贡献TOP10

于漪泉
-----

共享 TOP10 导航条是员工共享文件共享量的前 10 位排序

▼ 阅读TOP10

孙健
----

阅读 TOP10 导航条是员工阅读文件共享的前 10 名排序

## 3.6 高级搜索

点击搜索条的  即可进入高级搜索页面



高级搜索

按时间检索：从  到

搜索条件	①	包含以下全部的关键词	<input type="text"/> (必须填写)
搜索条件	②	包含以下任意一个关键词	<input type="text"/>
搜索条件	③	不包括以下关键词	<input type="text"/>
搜索结果显示条数	④	选择搜索结果显示的条数	每页显示10条
时间	⑤	限定要搜索的网页的时间是	全部时间
排序方式	⑥	排序的方式是	<input checked="" type="radio"/> 按相关度排序 <input type="radio"/> 按时间排序
文档格式	⑦	搜索网页格式是	所有网页和文件
部门客户筛选	⑧	部门客户筛选	选择部门
关键词位置	⑨	查询关键词位于	<input checked="" type="radio"/> 文档的任何地方 <input type="radio"/> 仅文档的标题中 <input type="radio"/> 仅在文档的路径中

注释0：选择精确的文档时间段

注释1：输入文档包含的关键词（组），用空格隔开

注释2：输入文档包含的任意一个关键词（组），用空格隔开

注释3：输入文档不包含的关键词（组），用空格隔开

注释4：每页显示的文章数

注释5：限定时间范围

注释6：选择文档的排序方式（按相关度或者按时间）

注释7：选择文档格式

注释8：选择搜索的文档出处（部门、人员）

注释9：选择关键词出现的位置（全文、标题或正文）

## 3.7 用户权限介绍

EKG 内网文档搜索引擎设置了三种用户：管理员，高级用户，和普通用户。

EKG 内网文档搜索引擎设置了三种权限：领导和我，部门，公共

权限\用户	外网用户	未登录用户	普通用户	高级用户	管理员
搜索文档	X	√	√	√	√
公共	标题可见	X	√	√	√
	摘要可见	X	√	√	√
	下载	X	√	√	√
	预览	X	√	√	√

部门	标题可见	X	√	√	√	√
	摘要可见	X	X	○	√	√
	下载	X	X	○	√	√
	预览	X	X	○	√	√
领导和我	标题可见	X	√	√	√	√
	摘要可见	X	X	◎	√	√
	下载	X	X	◎	√	√
	预览	X	X	◎	√	√
后台管理		X	X	X	X	√

注释 X: 限制

注释 ○: 只能查看自己部门文章, 不能查看其他部门文章

注释 ◎: 只能查看自己文章, 不能查看他人文章

注释 √: 非限制

### 3.8 使用小技巧

#### 一. 关键词输入技巧:

1. 表述准确

2. 查询词的主题关联与简练

3. 根据文档特征选择查询词(如要搜索分词方面的 WORD 文档可输入 “分词 双数组 DOC” )

#### 二. 文档的范围检索

在对文档目标比较明确时, 建议使用高级搜索, 对文档的搜索范围进行过滤, 达到最高的命中率.