



专业搜索白皮书

企业级信息引擎

管理海量信息的新方法

北京联索科技有限公司

www.lexotech.com

©2009 版权所有

目录

前言	1
理解专业搜索系统	1
专业搜索和信息检索	1
专业搜索和通用搜索	2
如何管理和利用海量信息?	2
实现搜索系统的 3 种方式	2
什么是好的专业搜索系统?	3
总结	3

互联时代的信息引擎

前言

社会朝着数字化和网络化的方向演进，面对互联网和组织内外飞速增长的信息，不仅要实时获取最有价值的信息，还要及时响应对信息的需求，提供最优质的信息及服务，我们的政府、企业和网站面临着与日俱增的信息挑战。

无论是作为信息服务的需求方还是提供方，都希望能够高效管控信息洪流，从中分离并整合关键数据、精纯信息，做到信息完整、精确、可靠、及时和统一，并且使其安全共享，为越来越多的信息访问者和工作协同方提供个性化的信息，从而提高工作效率，有效整合企业资源、降低运营成本，创造更多更好的社会价值。



Figure 1 专业搜索的应用

信息管理、电子商务、门户搜索等应用，越来越需要专业的信息技术平台进行支持，专业搜索正是为满足这样的需求应运而生的。它重点解决了目前各类搜索引擎的可管理性缺失的问题，将个性化搜索从理念变成现实，不仅符合当前，更是迎接未来，充分满足搭建企业搜索

（Enterprise Search）和互联网搜索引擎（Web Search Engine）的应用要求。

理解专业搜索系统

专业搜索是指对网络化信息和 Web 数据以及用户互动行为进行统一管理的搜索引擎软件系统。它针对行业信息或者主题信息，结合领域知

识，以任意信息单元为获取目标和处理对象，是智能化的企业级信息搜集与检索的通用软件平台。专业搜索是综合性信息管理系统的重要组成部分，能够辨别和检索来自文档管理、内容管理系统和关系型数据库等系统的信息，也能够作为搜索中间件，与其它企业应用集成。

与其它搜索引擎不同，专业搜索做为完整的网络信息处理和统一信息管理平台，尤其重视信息质量，前台提供搜索门户式的信息访问体验，后台承担持续性自动化信息处理中心的任务。从杂乱信息到有序信息，专业搜索独具特色的信息处理的三个阶段：

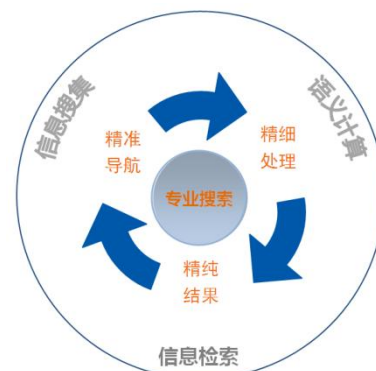


Figure 2 智能信息处理过程

专业搜索和信息检索

专业搜索与现存的信息检索系统不同，既不同于关系数据库检索系统，也不同于全文检索系统。

海量而复杂多样的多媒体信息使得传统关系数据库无能为力；简单地将信息检索技术转移到动态、实时的网络环境中，同样难以满足企业搜索的要求。

专业搜索系统将非结构化数据管理为特长的搜索引擎技术作为基础技术，利用目录、信息提取、分类法管理和信息可视化等手段，支持自动分类、聚类等语义计算技术，同时灵活集成企业现有业务流程和应用系统，并与适合结构化数据管理的数据库系统进行无缝集成。

与用户的 Web 搜索习惯相适应，专业搜索让用户在访问企业级信息时也能得到快捷的搜索体验。专业搜索系统能够快速而及时地为信息访问者有效提供特定信息，尤其是整合了内部信息资源和互联网信息资源的个性化信息，使得用户可以在自己控制的信息库中定位和调用网页、文档、数据和多媒体内容。

专业搜索和通用搜索

与通用搜索引擎不同，专业搜索是专为企业级用户研发的垂直搜索引擎平台。

专业搜索立足服务于各行各业的特定需要，满足细分化的市场应用特点，与通用搜索引擎相比，具有差异化的应用目标和技术方向。

目前的搜索引擎没能做到“博大精深”，因为客观上二者不易兼得。大型搜索引擎系统强导航作用、弱精准信息服务，部分网站现在提供的搜索功能是导入该类搜索引擎所提供的搜索结果服务，获得搜索功能的成本虽低，但不适合信息量大、信息更新频繁、信息安全要求高的网站使用，实际作用非常小，甚至成为网站健康发展的不利因素。

而专业搜索引擎具备有效平衡信息处理深度和广度的专业信息处理能力，核心功能是对各种类型的信息进行高质量的聚合、再分和挖掘。所以，专业搜索引擎能够按照用户指定的对内容的特殊要求和分类标准，从互联网上搜集和处理多种信息来源的相关信息，提供精准的信息。

如何管理和利用海量信息？

搜索管理包含：用户管理、信息源管理、内容管理、信息发布管理以及搜索统计分析等。搜索管理系统对专题、信息源，信息库、系统配置、专业词库等，以及搜索系统的运营维护进行全面管理。通过智能信息处理技术，对海量信息精准自动采集，通过搜索管理，可以在无人值守情况下的自动发布符合信息质量标准的合规的信息。

搜索的管理不仅是对信息的管理，还能够对用户的行为进行监测分析，更重要的是对人和信息之间的互动的管理，这也是搜索引擎的生命力所在。关键词和用户点击的搜索结果相关联，结合语义分析算法以及用户行为分析的算法，根据数据统计分析，自动淘汰低质量信息内容，始终维护一个高质量的信息库，使用时间越长、次数越多，信息库质量就越好，搜索结果也越准确。

查询关键词因为反映用户的意图也更有分析价值和利用价值。

实现搜索系统的 3 种方式

国际上，搜索引擎市场通常只计算传统安装方式的搜索产品，而排除了出售与搜索引擎相关的单项技术的公司。著名市场研究公司 Gartner 也把以应用服务方式（Application Service Provider, SaaS）提供服务的搜索产品排除在企业搜索引擎市场以外。

传统的信息搜集与检索程序能够满足数据量小的情况下的需要，结果准确性差，全面性差，无法实现高级搜索，用户的搜索体验不如人意，运行效率低。

企业级用户实现搜索引擎系统可以分为三种主要方式：

1. 购买专业厂商提供的搜索软件产品；
2. 通过数据库内置检索系统的查询语言进行二次开发；
3. 利用开放源代码的搜索技术，组建开发队伍，进行自行开发。

	数据库内置检索	专业搜索系统	修改开源代码
技术			
纯互联网架构	否	是	随意性强
亚秒级的查询响应	否	是	未验证
全面的搜索核心技术	否	是	否
信息自动提取和分类	否	是	否
中文信息处理	无	有	无
搜索管理			
信息质量控制	否	是	否
集群功能	是	是	未验证
用户管理	是	是	否
行为分析	否	是	否
实施和维护			
不间断稳定运行	是	是	否
简易的应用集成	否	是	否
不用依赖开发者	否	是	否
按需扩展系统	否	是	否
持续运行效能优化	是	是	否
总结			
成本	低	低	高
风险	低	低	高
价值	低	高	中

Figure 3 搜索引擎的主要实现方式

用户购买搜索软件产品可以获得功能强大的统一信息处理和管理的成熟平台，获得最佳实践经验，降低技术风险和人力成本。

什么是好的专业搜索系统？

如何评价专业搜索系统？好的搜索系统应该从4个方面进行评估，分别是应用专业化、服务个性化、技术智能化、体验人性化。

对于不同的商业应用，商业模型各不相同，企业级搜索引擎并不存在惟一的评判标准，也没有相同的相关性评价模型，专业搜索产品技术的优劣可以通过功能和性能来进行评价，在用户体验方面，重点在于搜索结果相关度的评价。

专业搜索将个性化机制引入到搜索引擎产品中，以应用目标为导向，平衡业务流程和技术实效。搜索引擎产品中重要的个性化调节参数包括新鲜度、全面性、权威性、统计分析、内容质量、性能调优等。就像视听音响系统多种多样的调节器，可以分别独立进行可视化地调谐操作。

搜索系统的使用者要得到相关的搜索结果不仅有赖于搜索系统对上下文的理解，同时要了解搜索用户的特点。用户用关键词表达的查询要求与用户所需的相关性结果之间存在着巨大的技术鸿沟，因此搜索引擎产品必须提供复杂的信息检

索后台功能和可控的参数，利用语言计算工具改善查准率和查全率，以超越关键词检索所能提供的准确度。

总结

中国的搜索引擎产品和国外的搜索引擎产品尽管在应用上具有比较大的差距，但是在核心技术方面难分上下，在文化和中文语言上更是具有先天优势。

专业搜索可以对接各种数据类型的信息源，对搜索引擎进行精确导航，对信息内容进行标记，自动获取并管理海量信息中的精华，在经过智能化信息处理的专业存储库中随意搜索和呈现个性化内容。

现在，国内的专业搜索系统产品已经适应专业信息服务的要求，不仅能够高效率地满足信息服务和信息利用的需要，而且能够系统地、全面地做到个性化的定制和个性化信息服务。