



XDFullSearch 全文检索产品白皮书

V6.0

北京线点科技有限公司

二零一八年

目录

1 综述5

1.1 关于线点5

1.2 产品概述6

2 产品的特点7

2.1 系统体系架构7

2.2 产品特点8

3 技术优势10

3.1 强大的处理能力10

3.2 准确的分词处理10

3.3 丰富的数据格式支持10

3.4 跨数据源检索10

3.5 跨平台支持11

3.6 良好的扩展功能11

3.7 完善的安全机制11

3.8 搜索结果多样化展现方式12

3.9 易用性与用户体验12

4 功能概述12

4.1 数据采集12

4.2 数据处理（ETL）13

4.3 数据索引13

4.4 搜索服务接口14

4.5 全文检索功能14

4.5.1 关键字检索14

4.5.2	逻辑运算检索.....	15
4.5.3	同义词搜索	15
4.5.4	拼音和概念搜索.....	15
4.5.5	多语言检索	16
4.5.6	二次检索	16
4.5.7	智能检索	16
4.6	个性化搜索	16
4.6.1	相关度	16
4.6.2	干预结果排序.....	17
4.6.3	用户体验	17
4.7	用户搜索行为分析.....	18
4.8	搜索聚合	19
5	性能指标.....	19
5.1	系统压力负载.....	19
5.2	性能指标	20
6	软硬件要求.....	20
6.1	软件配置	20
6.2	硬件配置	21
7	售后服务.....	21
8	用户案例.....	23
8.1	上海市知识产权局.....	23
8.2	北京市知识产权局.....	24
8.3	中国移动研究院.....	24
8.4	广东电网数字图书馆	25

8.5 内蒙古人民政府网站内搜索 26

9 其他客户（部分）： 27

10 联系我们..... 28

1 综述

1.1 关于线点

线点科技是国内领先的企业大数据应用解决方案供应商，成立于2005年，为中关村高新技术和国家双软认证企业，并通过了ISO9001质量管理体系认证，在金融、电信、政府、IT、专利等众多行业拥有良好的声誉和品牌影响力。

线点科技拥有众多的全文检索与搜索引擎领域专业人才，成立了技术研发中心，致力于搜索引擎技术及相关产品的研究开发。线点科技以数据应用为导向，以全文检索、搜索引擎、中文语言知识处理三大核心技术，为客户提供品质优良的专业化产品、完善的服务和量身定做的应用解决方案，特别在对于海量数据检索、数据挖掘，中文数据分析，多源异构数据整合方面有着丰富的经验。

线点科技的公司理念是“人才为本，科技为先”。“精益求精，止于至善”是线点科技始终坚持的服务宗旨。线点科技以创新实干的姿态，高效而团结的精神，凭借多年累积的技术实力及丰富经验，为每一个客户提供专业、贴切、快捷、高效、规范的技术支持和服务。

1.2 产品概述

线点全文检索产品（XDFullSearch）是由北京线点科技公司采用国际领先的云计算架构和NoSQL检索技术，结合自己近十年的信息检索和中文信息处理研究成果而开发的。它是新一代企业搜索技术与资讯革命的中心，我们将改变组织营运的方法与策略，将企业搜索转变成IT架构里的重要组件，并让搜索成为每个重要应用的核心引擎。帮助客户以搜索科技领先竞争对手。

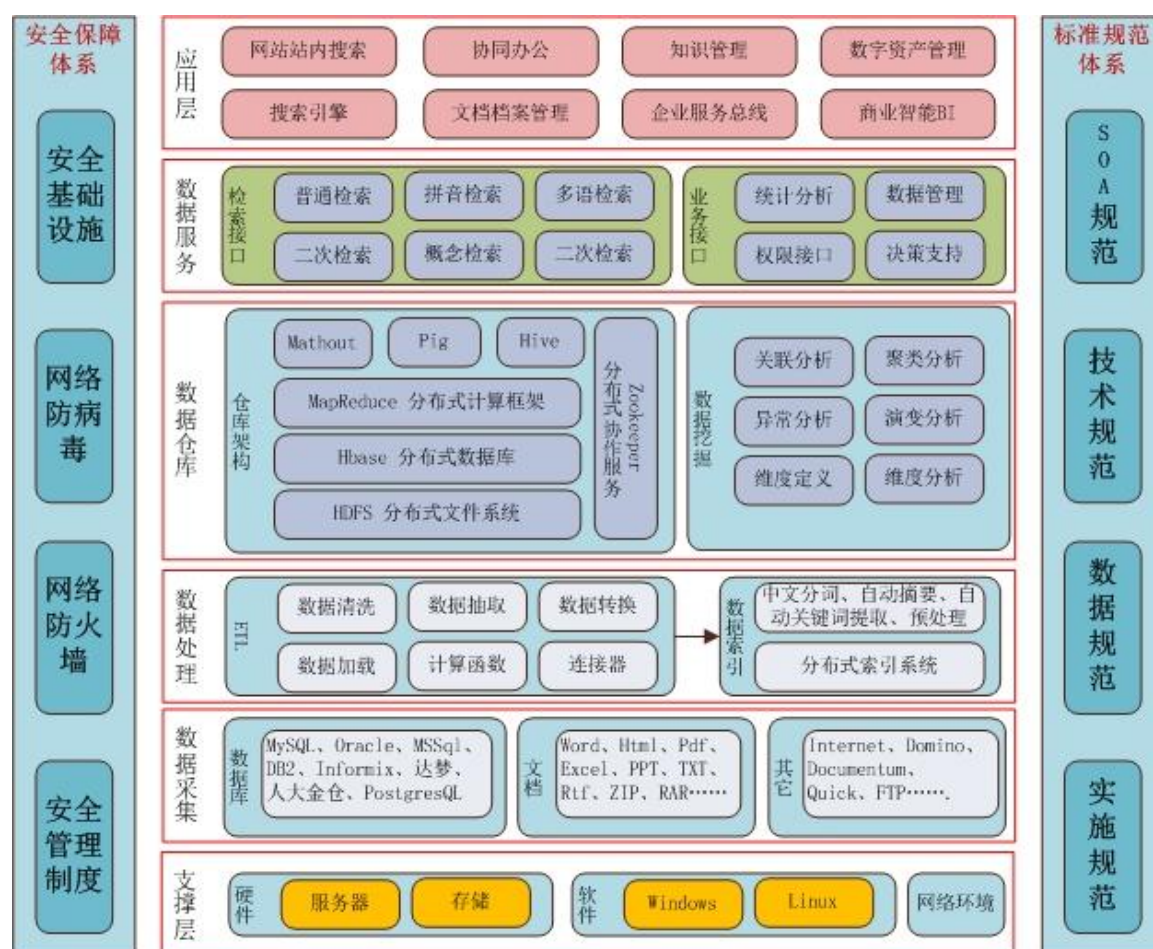
以线点全文检索产品为核心的企业搜索引擎解决方案，将不同来源的所有信息内容整合在一起并以易用的形式再现给用户，为企业解决大数据下的全文检索及数据分析和数据挖掘应用，并向用户提供高效的、准确的、安全的搜索体验。

线点全文检索以“聚合”为设计理念，不仅关注于大数据的整合与检索，同时在高效的并行计算架构基础上，通过业务驱动，实现内容的聚合与挖掘。“聚合”搜索，更关注在用户业务领域内的垂直化需求，以及基于对用户的特性标签和用户行为分析的个性化搜索展示和数据智能推送。从而使搜索引擎系统真正地具备行业特性和用户特性。

线点全文检索产品提供了对移动互联网的支持，支持多种云计算架构。并将搜索从传统电脑延伸到了手机，Pad等移动终端。

2 产品的特点

2.1 系统体系架构



在XDFullSearch的应用模型中，核心层为数据采集层、数据处理层、数据仓库和数据服务层。

数据采集层，通过各种不同的连接器，实现对各类数据库、文件系统、第三方应用的数据进行数据采集； 数据处理层，是对采集的数据进行清洗、抽取、处理等操作，并生成数据索引。数据仓库层，是通过Hbase、Zookeeper等分布式系统建立数据仓库，并提供数据挖掘功能；数据服务层，则是为其上的

数据应用层提供各类数据接口。

2.2 产品特点

(1) 提供全方位检索手段，包括外部特征与正文内容的各种逻辑组合检索、位置检索、二次检索、渐进检索、模糊检索、历史检索、词根检索、大小写敏感检索、概念检索、对检索结果按与检索表达式的相关性程度排序等。实现精确计算，报告准确的检索记录数。

(2) 支持对检索结果的各种排序：基于文章内容与检索表达式的相关度计算的相关性排序；基于一个或多个特征属性的字段排序。

(3) 支持检索结果的统计和浏览

(4) 提供动态索引，能够实时提供检索服务，原子更新，可添加、删除、修改和增加已存在文档的字段值，无需再次传送整个文档。分布式索引设计，提供近乎实时的操作和 NoSQL 特性，如 realtime-get，乐观锁和持续更新。

(5) 支持按词索引、按字索引和字词混合索引。即适应不同应用环境的需求。

(6) 完善的非结构化数据管理。

(7) 支持多种数据源数据的全文索引。可自动将关系型数据库中的数据导入全文数据库，如DB2、Oracle、SQLServer、MySQL、sybase、Informix、Postgresql、Access、达梦数据库、人大金仓数据库，并自动保持同步更新。

(8) 支持多种格式文档信息全文检索，文档格式二百余种，支持ftp采集。

(9) 集成先进的中文自然语言处理技术，实现自然语言及智能扩展检索，如同义词、概念检索、拼音检索检索。支持拼写检查(DirectSpellChecker) 可直接从自动字典检索校正；可基于关联度来选择一组文档；

(10) 提供按栏目、网站、网站群的全文检索功能。

(11) 支持中英文的混合检索，支持日、韩、法、藏、维多语系。

(12) 允许使用文中的任意字、词、短语、句和片段进行检索，支持 RegexpQuery 正则表达式查询

(13) 支持结构化数据和非结构化数据的混合检索

(14) 支持对domino、notes、documentum、Quickr等多个系统的集成

(15) 支持个人电脑，手机，Pad等多终端

3 技术优势

3.1 强大的处理能力

线点科技全文检索产品不仅提供了快速检索功能（在千万级数据下，全文检索能在0.5秒钟之内完成），同时也有着高效的海量数据处理能力。每小时能处理百万的数据库数据，或是几万篇的文档。包括中文的分词和建立数据索引。我们的产品多个大型项目中经过了实际的检验，成功实施过多个超过5千万数据规模的项目。

3.2 准确的分词处理

能准确分析信息内容，进行分词处理，有效保证中文检索的准确性；系统应具有开放接口，可快速集成第三方分词产品；同时应支持用户自定义专业词汇，支持中文、英文等多种语言；

3.3 丰富的数据格式支持

除了基于数据库的搜索以外，线点科技全文检索产品还支持文件检索方式，可以通过指定文件所在目录进行高效检索，并内置文件解析器，可对Text、HTML、XML、RTF、MS OFFICE文档（Word/Excel/Powerpoint）和PDF文件自动解析。同时根据用户需要产品也支持从其它各类数据源获取要检索的数据内容，例如：XML文件、exchange邮件服务器等等其它数据池。

3.4 跨数据源检索

可以从文件系统、网站、关系型数据库系统等系统中采集数据，并提供二次开发包，以实现组织对特有数据源的处理；支持多种数据源数据的全文索引。可自动将关系型数据库中的数据导入全文数据库，如DB2、Oracle、SQLServer、

MySQL、sybase、Informix、Postgresql、Access、达梦数据库、人大金仓数据库，并自动保持同步更新。

3.5 跨平台支持

支持各种主流操作系统和主流开发平台，如Windows、Linux；并支持主流开发平台，如Java、.Net、 PHP、asp、flex；，系统支持SAN、NAS、LAN-free和DAS等多种存储设备；

线点科技全文检索产品既可以搭建在传统的网络和服务端环境下，同时也提供了对私有云 (Private Clouds) 的支持。可以与vSphere、Hyper-V和Xen集成，从而充分利用现有硬件资源和软件资源，降低企业的成本。

3.6 良好的扩展功能

线点科技XDFullSearch产品支持分布式，并且在各个功能模块均可以实现分布。通过管理控制台强大的索引管理功能，管理人员可以针对数据域制定优化方案，在存储空间和响应速度以及内存占用之间调整平衡，以达到最优性能。XDFullSearch使用多级缓存检索结果的策略实现海量数据的快速响应，如搭配高性能服务器，则效果更加明显。

XDFullSearch 支持与Hadoop、Hbase、HDFS等分布式系统集成部署。

同时，系统采用了插件式的开发方式，可以通过开发多种插件进行功能扩展，满足客户的个性化需求。

3.7 完善的安全机制

提供完善灵活的安全访问控制机制，包括权限控制、用户认证和用户授权，以保证系统内信息内容安全，提供系统级、索引库级、记录集和字段级的安全控制，只有用户有权访问的内容才显示在搜索结果列表和分类视图中。

XDFullSearch 支持 LDAP (Lightweight Directory Access Protocol)

协议，LDAP提供了一种国际通用的目录访问标准，可以在任何计算机平台上，用很容易获得的而且数目不断增加的LDAP的客户端程序访问LDAP目录。而且也很容易定制应用程序为它加上LDAP的支持。意味着XDFullSearch可以与任何支持LDAP协议的软件或系统进行数据通信。

支持“索引时安全性”，即搜索引擎的爬行器在相应数据源的爬取过程中提取数据源端的安全控制信息（ACL），并构建于索引中。用户提交搜索请求时，搜索引擎利用索引中的安全控制信息对用户搜索结果进行安全过滤，从而达到用户只能看到有权限访问的信息的目的。

3.8 搜索结果多样化展现方式

无论是数据来源于数据库中，或者是各类文件，对于搜索的展示结果，用户可以通过接口或委托线点科技开发进行个性化、多样化的展现，也可以类似数据库那样，对于数据字段进行自由调用查询和展现。

3.9 易用性与用户体验

系统主要功能均提供可视化界面维护，采用B/S架构，界面操作友好，功能清晰，便于用户使用和维护。

4 功能概述

4.1 数据采集

XDFullSearch支持对SqlServer、Oracle、Mysql、DB2、PostgreSQL、Informix、人大金仓、达梦等常见的十余种数据库进行数据采集，采集对象可以为单表、多表或视图，并且可以进行增量采集和数据自动更新。支持blob、clob等大字段内容；支持一个字段指向外部文件联合采集建索引；除了关系数据库，还支持Native-XML数据库。

线点全文检索产品支持对本地文件、网络文件目录、FTP目录、NFS文件目录进行遍历、采集和更新。支持的文件类型近300余种，包括常见的Txt、HTML、Xml、Office文档、Pdf、Rar文件，包括JPEG、BMP、GIF、TIFF、PCX、AVI、WAV等二十多种图像声音动画格式。此外也可以自定义文件扩展名和解析器。用户也可以通过开发插件方式来解析其他的文件格式内容。

线点全文检索产品可支持多国语系（英文、繁体中文、日文、西欧语系等）混合的文件；可支持藏文、维文等少数民族语言，文件编码支持Big5、GBK、Unicode、UTF-8等。

全文检索产品还以插件方式提供对多种第三方数据的采集连接，这些数据来源包括MS Exchange、Lotus Notes、Domino、SAP、MS SharePoint、EMC Documentum等。线点科技在多年的项目实践中，与第三方的数据连接超过一百种，例如万方、知网、维普、聚合等等。

4.2 数据处理（ETL）

ETL即数据抽取（Extract）、转换（Transform）、装载（Load）的过程。它是构建数据仓库的重要环节。数据处理模块，主要是对数据的清洗、过滤、转换和处理。

XDFullSearch可以条目（记录）为单位，也可以字段为最小颗粒度进行数据处理。全文检索产品以流水线模式进行多阶段的数据处理，并支持插件方式，自定义流水线，或是在流水线中插入、去除、修改处理流程。

通过数据处理（ETL），可以实现实体提取、自动关键词提取、自动摘要、引擎分类等功能。也可以通过编写扩展插件，满足用户的个性化需求。

4.3 数据索引

XDFullSearch支持对多Unicode、GB2312等多种编码，多国语言以及藏维文等少数民族语言的内容索引。支持按词索引、按字索引和字词混合索引。

即适应不同应用环境的需求。索引结构支持多种数据类型(如日期、字符串、短语、文档和二进制多媒体类型)等等;

支持动态增量索引,能够实时提供检索服务。即数据增删改时快速同步更新索引,而不需重建整个索引。同时高效的数据和索引压缩,实现了低空间膨胀率。

线点全文检索产品能负载的索引大小,理论上仅受磁盘空间存储的限制,通过分布式功能,索引可以存储在不同的磁盘上或是不同服务器上,同时联机的服务器可以达到上千台。

整个索引系统会被当作一个统一的资源来考虑,系统将会根据不同节点或是不同服务器的负载情况,来对新增的索引自动进行分发,从而更有效更合理地利用服务器资源。

4.4 搜索服务接口

全文检索产品开发接口支持主流开发平台,支持的语言包括.NET, JAVA, PHP等。可与各类WEB应用进行无缝集成。同时提供完善的二次开发API手册帮助及应用示例。

4.5 全文检索功能

线点全文检索产品的全文检索功能采用了优化的查询算法,包括分布式索引技术、多线程并行运算技术、Bigram技术等等,并且使用了多级缓存机制,支持大并发的数据访问,也能够在上百G数据下,达到亚秒级的检索速度。以下对云搜索产品支持的检索方法进行介绍:

4.5.1 关键字检索

关键字检索是最普通的检索方式,可以划分为两种,一种是单词检索方

式，另一种是短语检索方式。如果按字段因素考虑还可以划分为加入字段限制的关键字检索与不加入字段限制的关键字检索。如果用户只输入关键词而不设定检索字段进行检索，则系统按照设置的默认检索字段进行检索。

支持大小写无关检索、全半角无关检索、中文简繁体扩展检索，对数值、日期等特征字段可以进行比较和范围检索；

4.5.2 逻辑运算检索

狭义的逻辑运算符号包括AND、OR、NOT，广义的逻辑运算符除了包括上述三个运算符外还包括括号（），括号必须成对出现，而且括号允许嵌套。逻辑运算符的优先级别为NOT最大，其次AND，最后是OR，括号可以改变逻辑运算的顺序。此外产品还支持*? %通配符，帮助用户更准确命中结果。

4.5.3 同义词搜索

同义词搜索是指搜索中对于义同字不同的语言处理，例如中国石油，使用了多种不同的名称：“中国石油”、“中石油”、“中国石油天然气股份有限公司”等。如果我们仅用其中的某一个或某几个名称检索，而不知道或忽略了一些其他的叫法，就会产生漏检。

采用同义词搜索，系统会将搜索词语进行同义延伸，在搜索时同时对该词语的同义词进行搜索，从而提高搜索的查全率。

4.5.4 拼音和概念搜索

拼音检索指的是在系统中输入拼音即可找到对应的中文词语内容，这其中包含了拼音纠错、多音字处理的功能。

概念检索体系中，词之间的关系将以概念进行来划分，通俗地说是以词义或语义来进行相关度判断，而不是词语的中文文字。通过概念检索更能够

准确地把握用户检索意图，提高用户的满意度，例如 用户检索“手机”，那么将提示“移动电话、Mobile”等多个概念。

线点云搜素产品包含近200万的概念词库，并且允许用户自定义和扩展概念词库。

4.5.5 多语言检索

除了中文简体以外，产品还支持中文的简繁体检索、中英文混合双语检索、支持德语，法语，日语等多国语言、支持藏文，维吾尔文，蒙文等少数民族语言检索。

4.5.6 二次检索

二次检索的确切定义为在已有的检索结果中进行检索，即在某次（不局限于第一次）检索中输入新的检索条件进行检索。二次检索可以提高用户检索结果的准确率，通过漏斗过滤机制，引导用户检索到目标数据。

4.5.7 智能检索

线点科技独创的智能检索技术，在查全和查准方面比一般检索系统高得多。智能检索是线点科技多年的检索技术和经验的总结，也是通过对用户搜索行为和搜索意图分析而得出的检索最优路径。智能检索融合了智能中文分词、同义词检索、拼音检索、主题词检索、非法、停止词处理、英文词根检索等多种检索技术，符合大多数用户的检索习惯。

此外，结合数据分析和用户行为分析，可以实现智能语义搜索功能，通过用户搜索词自动判断用户搜索意图，直接推送数据结果。

4.6 个性化搜索

4.6.1 相关度

为了使检索结果与用户的搜索意图更相关，有两种方式对相关性排序算法进行调整。一种是调整索引项的相关性，例如新闻的相关文章，网页的相关链接等等。另一种是动态的属性，例如网页点击率，词频，近义词等等。

结合以上两种相关联的大小数值（应用中生成的一个分数值），即得到“相关度”。应用人员可以创建一个或几个权重。在一次检索的基础上，用户可以指定哪些相关属性用于计算排序的相关分数，并能设定具体的数值以控制相关性排序结果。

用户也可以选择忽略相关性排序，使结果只按其他的属性排序，如创建日期，标题或作者等。

4.6.2 干预结果排序

字段加权：对文档内的某些字段进行加权，使这些字段的权重高于其他字段。举例来说，点击一篇文档的标题要比点击内容重要10倍。字段的权重值可设定任何数值或合并到已经存在的索引字段中。

内容置顶：此功能可以控制哪些索引内容最优先排序，通过将该文档置于搜索结果排序的最前端以体现其重要性并推荐给用户，它一般的应用是基于某种环境中某一特定网页被认为是某种条件检索的最权威页面。例如，当用户输入某个公司名字搜索时，应将该公司官网链接置在首位。

智能推荐：线点全文检索产品除了根据内容排序，还能够与用户系统集成，通过分析用户的喜好和用户行为，选择性地推荐其感兴趣的内容进行优先展示。

4.6.3 用户体验

搜索提示：搜索输入框可以根据用户的键入词进行自动补全，搜索提示，搜索提示的字典来源，既可以是用户的搜索历史，也可以是自定义行业的字

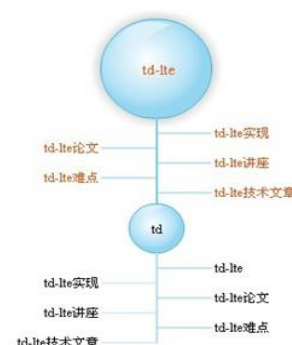
典。



启发式搜索罗盘：根据用户输入，推测用户意图，以罗盘方式提示用户进行进一步地搜索。

拼写检查：系统可以对用户输入的搜索词进行拼写检查，对错误词进行自动纠正。

概念搜索：通过概念词库，在搜索结果中，按照词语概念进行检索和推荐，提高数据查全率。



动态导航：

- 对搜索结果进行归类和细分
- 搜索结果的多维度动态导航
- 搜索结果排名控制，特色内容
- 搜索分析和调整
- 帮助用户发现信息/知识

序号	公开号	名称
1	CN1058493	电磁感应断路器
2	CN1058675	电磁带调亮时空滤波阵列天线
3	CN1058690	电磁灶防电泳电路
4	CN86100764	多功能节能电磁接触器继电器
5	CN1036486	无整流器脉冲直流电机及派生电
6	CN86106942	用于密封式电动压缩机的电动机
7	CN1058120	抗磁超导耦合旋转电机

4.7 用户搜索行为分析

通过记录用户的搜索行为和点击行为，可以分析用户的行为特征，进而判断出数据内容的质量、受欢迎程度，从而对数据排序和权重进行优化。通过对搜索轨迹特征的分析，可以更快地猜测用户意图，帮助其找到目标数据。

通过对用户行为数据的挖掘，还可以结合客户的业务需求进行建模和定制和开发。

4.8 搜索聚合

“搜索聚合”是线点全文检索产品的特色功能，它将离散型的数据片段，与用户的业务模型进行碰撞，使业务属性黏着在数据上，从而实现用户业务角度上个性化的聚合展现。

例如：当用户搜索词为行业领域里的专有词时，系统可以根据这个词相关的时间，地点，人物进行聚合展示。

5 性能指标

5.1 系统压力负载

单节点负载能力：

项目	指标	说明
设计容量	1000 万条数据或 200G 文档	总索引数量
访问并发量	200	与服务器配置相关

分布式系统负载能力：

项目	指标	说明
设计容量	10 亿级数据或百 T 级文档	总索引数量

访问并发量	2000	与服务器配置相关
-------	------	----------

5.2 性能指标

项目	指标	说明
数据库采集速度	100 万-300 万/小时	可通过分布式扩展
文档采集速度	10000 篇/小时	可通过分布式扩展
文档解析成功率	> 96%	
访问并发量	单机可支持 200 以上并发	可通过分布式扩展
检索响应时间	简单检索用时小于 0.5 秒， 复杂检索小于 1 秒；	

6 软硬件要求

6.1 软件配置

项目	配置	说明
操作系统软件	Linux (ES/AS) Windows	windows/linux 平台可选
JDK	JDK1.6	Jdk1.6 以上
应用软件	XDFullSearch	1 台机器可部署多个节点

WEB 中间件	Tomcat5.5 / Apache2.0	支持其它中间件
数据库服务器	XDFullSearch 支持主流数据库	

6.2 硬件配置

项目	配置	说明
CPU	>2.0GHz 四核	
硬盘	>160G	可安装磁盘阵列
内存	>2GB	
网络	100Mbit 以太网卡	

7 售后服务

为了帮助客户切实地解决问题，特别是在产品应用中遇到的问题，线点科技建立了完善的售后服务体系，通过售后服务团队，我们可以提供离岸（Off-Site）和在线（On-Site）的售后服务支持，以满足不同客户对于售后服务的要求。

➤ 服务项目

服务项目	√
一年内免费技术支持	√
一年内免费产品升级	√

三年内产品 bug 免费维护	√
MSN/QQ 在线支持	√
Email 在线支持	√
工作时间电话支持	√
系统 Bug 反馈	√
系统部署与培训	√
二次开发技术支持	√
系统集成支持	√
分布式部署支持	√
集群、负载均衡部署支持	√
现场技术支持	√
定制开发支持	√

注：以上有些项目为收费服务内容

➤ 响应时间

- (1) 对于使用、配置问题在 2 小时内给予答复，24 小时内解决问题；
- (2) 对于需要修改的 bug，包括接口，功能点，脚本错误等等，要求在 4 小时内给予答复，48 小时内解决问题；
- (3) 对于客户要求的定制内容，根据双方协商时间完成。

8 用户案例

8.1 上海市知识产权局

项目介绍：

目前全世界的专利数据有4000-5000万，对于专利检索，不仅要求速度，而且对于查准率，查全率都有很高的要求，专利检索的语法不仅要求支持逻辑检索、布尔检索、概念检索等，还要求支持检索通配符、前方截词、后方截词、中间截词等多种语法。

之前的专利数据都存放在Sql Server 2005里面，而且由于专利数据繁多，都是以国家、分类等方式建立了多个数据库，多个表，在多个机器中存放，要达到统一的检索目的，必须需要借助一个优秀的搜索引擎系统来实现。

不仅如此，项目对于软件提供商也提出了很严格的要求，包括产品规范，产品的稳定性，对于服务的响应要求时间，现场的技术支持等等，还设立了项目监理厂商，负责对项目的进度、质量进行监理和验收测试。

此项目由中软中标，然后再经过对多家厂商的全文检索产品进行比较、测试、试用、验证后选择了线点科技的产品作为数据引擎中间件的软件提供商。

由于专利检索的检索式比较复杂，为了保障高效的检索效率，线点科技为此系统部署了近20个节点的分布式检索系统，平均每个节点负载200-250万数据，压力设计在并发100人以上，为达到性能的要求，系统共使用了1台web服务器，2台检索服务器（小型机），1台数据库服务器。

目前项目已经验收完成，并且通过了客户、监理方的验收。



8.2 北京市知识产权局

北京市知识产权局专利检索系统囊括了大约6000万的专利数据，提供简单检索、表格检索、高级检索、二次检索等多种检索方式。

该项目于2010年开始，2011年完成验收。项目有紫光公司负责集成总包。线点科技承担了专利检索相关的模块和功能开发。



8.3 中国移动研究院

项目介绍

中国移动研究院需要对其的内部门户搭建一个站内检索系统。门户系统包含了办公门户、协同平台、文档管理等多个模块，其中涉及Portal、Documentum/Lotus、DB2数据库、论坛等多个应用系统的整合，包括单点登录、权限整合等功能，以及各种格式的文档解析等等。

中国移动研究院同样在经过对几家搜索引擎厂商的招投标后，线点科技凭借其优秀的软件产品、强劲的开发实力和多个类似系统的实施经验，中标此项目，进行门户搜索系统的搭建、整合工作。



8.4 广东电网数字图书馆

项目介绍

广东电网公司亟需搭建一个数字图书馆系统，除了电网自身的资料数据外，从万方、维普、聚合等数据厂商又购买了一些数字数据，数据分类包括期刊、论文、专利、成果等等。数据的存储有文档文件、Excel表格、数据库数据等多种形式，数据总量达到6千万。

为了解决这些多源异构数据的整合问题，广东电网公司选择了线点科技的全文检索产品作为数据整合软件，进行这些相关数据的整合检索，并在此基础上提供相关的业务功能。

系统设计支撑容量为一亿，按照每年增长500万计算，可以至少支持五年内的应用，为此系统部署了近10个节点。



8.5 内蒙古人民政府网站内搜索



9 其他客户（部分）：

神华集团企业统一搜索平台	中共中央办公厅统一检索系统
船级社内容采集与检索平台	国家质检总局
中国石化企业统一搜索门户	北京市经信委统一搜索门户
中国电建知识搜索平台	北京市人力社保热线检索系统
浙江大学知识资源搜索门户	南京市税收政策垂直搜索引擎
海南航空业务数据检索平台	河北省科技厅网站群搜索系统
宝钢集团协同平台统一检索系统	教育部学位中心搜索引擎系统
中电集团 36 所内网搜索引擎	金华市公安局综合信息检索系统
CETC 重点库资源检索系统	温岭市公安局信息检索系统
韩亚银行内部统一检索平台	北京市科委
国家电网统一搜索平台	青海海南州藏语委
中海油集团统一搜索平台	重庆市知识产权局专利检索系统
中国港湾知识共享搜索平台	长沙市知识产权局专利检索系统
首都机场集团统一搜索门户	广东省食药监局
中国黄金集团统一搜索门户	中科院文献情报中心

10 联系我们

欢迎致电

公司地址：北京市海淀区上地信息路2号国际科技创业园1号楼26层

联系电话：400-605-6088

010-82896088 - 8875

传真：010-82896055-8899

邮编：100085

网址：<http://www.xd-tech.com>