# 网络搜索引擎发展现状的探讨

梁伟贤

■ 广东电信有限公司广州市天河分公司政企客户部

摘 要：本文在研究互联网中现代搜索引擎的发展现状的基础上，分析独立搜索引擎、元搜索引擎的优缺点，并重点分析了基于 Agent 技术的改进型元搜索引擎。此新型元搜索引擎具有自治性、能动性和协作性的特点，能帮助用户在众多的信息中，快速准确地搜索到符合自己需求的信息。最后提出新型网络搜索引擎发展趋势。
关键词：元搜索引擎 独立搜索引擎 Agent 技术

Abstract：On the base of development of Search Engine, this paper assays the advantages and shortcomings of Single Search Engine、Meta Search Engine，makes the Meta Search Engine as the research target，and assays a project of modified Meta Search Engine which is base on the Agent technology. The Agent technology has autonomy、variability and cooperation as the basic component. This modified Meta Search Engine can help the user grasp the necessary message quickly. At last, the paper summarizes the problems to be resolved of Search Engine.
Key words：Single Search Engine, Meta Search Engine, the Agent technology

## 1.引言

随着 Internet 应用的普及,Internet 已发展成为一个巨大的分布式信息空间。由于 Internet 面向社会和个人,信息的产生不受约束,人类的信息世界呈现出前所未有的复杂现象。Web 信息的大容量、异构性、分布性、动态性等特点造成了“信息过载”,如何有效地为用户提供 Web 信息检索已经成为一项重要的研究课题。

互联网中搜索引擎正是为了解决这个“迷航”问题而出现的技术。搜索引擎经历了从简单的机器人搜索软件、独立搜索引擎到专业搜索引擎、元搜索引擎的发展历程，成为人们在 Web 上寻找信息不可缺少的工具。

元搜索引擎被称为搜索引擎之上的搜索引擎。元搜索引擎具有查全率高、搜索范围更多更大，查准率较高等优点。但其同时亦具有检索性能、调用搜索引擎、检索结果的数量方面的局限性，需要进一步改进。

本文在研究现代搜索引擎的发展现状的基础上，分析独立搜索引擎、元搜索引擎的优缺点，并重点以元搜索引擎为研究对象，分析了一种基于 Agent 技术的改进型元搜索引擎的方案。

## 2. 搜索引擎简介以及发展现状分析

搜索引擎是一些在 Web 中主动搜索信息并对其自动建立索引的嵌入在 Web 网站中的具有搜索功能的程序。其索引内容存贮在可检索的大型数据库中，从而提供信息检索服务。当用户输入待搜索的主题词时，搜索引擎会告诉用户包含该主题词信息的所有网址，并提供该网址的链接，所以搜索引擎既是用于检索的软件又是可提供查询、检索的网站。。

### 2.1 搜索引擎发展现状概况

搜索引擎的分类方法有很多。其中，按包括搜索工具的数量可分为独立型和集成型。独立型搜索引擎自身具有独立的数据库。集成搜索引擎没有独立的数据库,只提供一个统一界面,把多个具有独立功能的引擎组合起来。

元搜索引擎属于集成搜索引擎,由于元搜索引擎具有许多优于传统独立型搜索引擎的特点，目前元搜索引擎已被广泛地应用于信息检索的众多领域。

### 2.2 独立搜索引擎发展现状

#### 2.2.1 独立搜索引擎的概念以及特点

独立搜索引擎是指自身有一套完整的搜索、整理、查询机制的搜索引擎。它曾经一度是搜索引擎的主流，即使是现在，也仍然是 Internet 上使用最广泛的搜索引擎。尤其是独立的专业搜索引擎，对于专业性较强的查询或者域很清晰的查询，往往能获得相对精确的结果，很受用户青睐。但是，它是以牺牲检索覆盖面为代价的，在一定程度上限制了其信息的全面性。随着 Web 信息量

飞速发展，独立搜索引擎已不可能囊括整个网络的内容，不得不寻求新的出路。

2.2.2 独立搜索引擎的基本组成

独立搜索引擎简单易用，并且构建相对容易，其中的搜索算法、组建机制、排序思想等都是建立其它搜索引擎的基础。一般的独立搜索引擎由“机器人”，索引软件，查询、排序软件这三部分组成。

2.2.3 独立搜索引擎的不足

独立搜索引擎的不足主要表现在以下方面：（1）查全率比较低：每个引擎的覆盖面都相当有限。据统计，没有一个搜索引擎的索引量超过整个网络网页总数的1/6。（2）查准率比较低：目前通过搜索引擎检索的网络信息资源相关性非常差，浪费了用户大量的相关判断时间。（3）每一个搜索引擎都有自己的检索规则，用户利用不同的搜索引擎需要进行不同的适应过程，增加了用户的负担。（4）多数搜索引擎采用关键词检索，并提供高级检索功能，但用户很难通过组配关键词来准确表达自己的信息需求，导致检索效率低下。（5）更新速度比较慢：搜索引擎机器人只能在由系统管理员确定的一定时间间隔内跟踪特定信息，不能保证信息的及时更新，导致产生错链和死链，导致检索速度变慢。

2.3 元搜索引擎的发展现状

元搜索引擎又被称为搜索引擎之上的搜索引擎，用户只需递交一次检索请求，由元搜索引擎负责转换处理后提交给多个预先选定的独立搜索引擎,并将所有查询结果集中起来以整体统一的格式呈现到用户面前。

2.3.1 元搜索引擎的组成

一般的元搜索引擎由三部分组成，即：检索请求提交机制、检索接口代理机制、检索结果显示机制。"请求提交"负责实现用户"个性化"的检索设置要求，包括调用哪些搜索引擎、检索时间限制、结果数量限制等。"接口代理"负责将用户的检索请求"翻译"成满足不同搜索引擎"本地化"要求的格式。"结果显示"负责所有源搜索引擎检索结果的去重、合并、输出处理等。使用元搜索引擎同时对几个搜索引擎进行检索，获得分级编排的检索结果。

2.3.2 元搜索引擎的工作原理

用户通过 WWW 服务访问元搜索引擎，向 Web 服务器提交检索式。当 Web 服务器收到查询请求时，先访问结果数据库，查看近期是否有相同的检索，如果有则直接返回保存的结果，完成查询；如果没有相同的检索，就分析检索式并转化成与所要查找各搜索引擎相应的检索式格式，然后送至 Web 处理接口模块。Web 处理接口通过并行的方式同时查询多个搜索引擎，把所有的结果集中到一起。根据各搜索引擎的重要性，以及所得结果的相关度，对结果进行抽取并排序，生成最终结果返回给用户。同时，把结果存到自己的数据库里，以备下次查询参考使用。

2.3.3 元搜索引擎的分类

不同的元搜索引擎在可以检索的目标搜索引擎、检索提问的处理方式以及如何编译和显示结果方面，有着很大的差异。有些元引擎一个接一个地搜索目标搜索引擎，另一些则同时进行搜索，有些搜索引擎将检索提问转变成目标搜索引擎的提问语言按功能划分。元搜索引擎包括多线索式搜索引擎和 All-in-One 式搜索引擎；按运行方式的差异可分为在线搜索引擎和桌面搜索引擎。

2.4 元搜索与独立搜索引擎比较的优势

元搜索引擎区别于独立搜索引擎，最主要的是一般没有自己独立的索引数据库，可以投入更多力量提供统一检索界面，形成一个由多个分布的、具有独立功能的搜索引擎构成的虚拟整体，用户通过元搜索引擎的功能实现对这个虚拟整体中各独立搜索引擎数据库的查询、显示等操作。

2.5 发展元搜索引擎的原因

要开发元搜索引擎的主要理由是：Web 数据量太大，而且增长迅猛，单个引擎的容量、处理 Web 能力难以扩展到很大的规模，所以每个引擎只能包含一部分文 Web 档。元搜索能够分散处理负载，增加检索的范围。

同时，元搜索具有较好的扩展性，可以加入多个成员引擎。它使得各个成员引擎规模变小，性能更好，这样成员引擎的检索响应时间短，还可以使得检索的内容保持最新。

## 3 新型 Agent 技术简介

3.1 Agent 技术

Agent 是一个应用范围极广的术语，一般被用来指具有感知能力、问题求解能力及与外界进行通讯能力的一个实体。有权威专家给出的 Agent 定义如下：Agent 是一定环境下的计算机系统，它能够对所在的环境进行灵活的自治动作，以满足其设计的目标。具有自主性、社会性、反应性和能动性的计算机硬件或软件系统可以均可称为 Agent。一般说来 Agent 的基本必备特性如下：自治（主）性、通信能力、感知能力和反应能力、能（主）动性、持续性、协作性。

3.2 多 Agent 技术

基于多 Agent 技术的系统是指多个 Agent 相互通讯、彼此协调，共同完成作业任务的系统，它不仅具备一般

分布式系统所具有的资源共享、易于扩张、可靠性强、灵活性强、实时性好的特点，而且各 Agent 能够通过相互协调解决大规模的复杂问题，使系统具有很强的鲁棒性、可靠性和自组织能力。在多 Agent 系统中，单个 Agent 是一个物理的或抽象的实体，能作用于自身和环境，操纵环境的部分表示，并与其他 Agent 通讯，具有感知、通讯、行动及控制和推理能力等基本功能。多 Agent 技术的这些特点，使得其在处理基于互联网的知识问题方面，具有广阔的应用前景。

## 4 基于 Agent 技术的新型元搜索引擎的探讨

### 4.1 系统体系结构框架

该系统逻辑上可分为三层体系结构，最上面一层是用户接口 Agent，中间一层是信息检索 Agent，最下面一层是检索结果处理 Agent。用户接口 Agent 负责用户与系统的交互：用户在这一层上提出检索要求，系统通过本层的 Agent 判断出用户的可能实际需求，再将查询要求主动或半主动地推送给用户，获得用户同意后再进行下步查询工作。信息检索 Agent 负责信息检索：本层 Agent 通过分析用户接口 Agent 层送来的查询要求，找到适合查询要求的若干搜索引擎进行检索。检索结果处理 Agent 负责将检索结果进行处理：本层 Agent 将上层的检索结果进行一系列的处理，最后转化成统一的形式返回到用户界面。

### 4.2 实现新型元搜索引擎的运行环境

元搜索引擎可以是CGI,ASP,ISAPI程序,但是这些实现在性能或软件移植上有一定的缺陷。由于 Java 在性能和移植上都具有突出的特点,引擎的实现可以采用 JavaServlet 技术。它的基本原理从客户端接受请求,并对请求作出反应,然后把运行的结果返回给客户,就可以与 FORM 或 Applet 相结合完成浏览器与服务器应用；它只需要加载一次,只有当 Servlet 发生变化时,才需要重新加载。用 Servlet 制作的搜索引擎在服务器端运行,多个用户会同时向一个 Servlet 发请求,这就要保证软件有很好的并发性。尤其要注意数据变量的同步访问、共享处理的问题。

### 4.3 新型 Agent 技术元搜索引擎的优势

目前存在的元搜索引擎虽然增强了定位和收集信息的能力，但由于其检索结果数量大增，没有针对用户个性需求对信息予以优化重组，用户每一次信息检索被孤立地对待，割裂了其需求在一定时段内的相关性。

基于多 Agent 的元搜索引擎技术可以解决这一问题。Agent 技术具有能够进行高级问题求解，可随环境变化修改自己的目标、学习知识并提高能力等智能特性，通过 Agent 的逐步学习，了解用户对信息需求范围，进行有针对性的搜索，并且根据用户检索次数的增加，检索的准确性也会逐渐地提高。用户能够通过一个统一的搜索引擎界面快速找到自己所需的资源。利用一个结果输出优化 Agent 对元搜索引擎的检索结果进行过滤、合成和排序，这样经过优化处理后的检索结果必然能够更好地满足用户的科研和学习要求。新型 Agent 技术元搜索引擎变得更加简单、高效、可行。

## 5 结论

独立搜索引擎、元搜索引擎作为目前应用于搜索引擎研究领域的两大引擎各具有各自的优缺点。基于 Agent 技术的新型元搜索引擎，将 Agent 技术与元搜索引擎相结合，具有自治性、通信能力、能动性、协作性强等特点，将对现有的搜索引擎的改进中起到积极的作用。可以预见，未来的网络搜索引擎发展中，基于 Agent 技术的新型元搜索将是不可忽视的一个新趋势。

## 参考文献

1.邱诚. 搜索引擎及其发展浅析[J]. 图书馆研究与工作, 2002,(2):8—10.

2.刘丽,孙燕唐.智能型元搜索引擎的设计与实现[J]. 计算机工程, 2003, 19(16): 118-121.

3.李广建. 元搜索引擎及其主要技术[J]. 情报学报, 2002, 20(2):175-179.

4.李振东,费翔林. 基于概念的信息检索模型研究[J]. 南京大学学报,2002,38(1):108-109.

5.凌美秀. 关于搜索引擎当前存在的主要问题及其发展趋势探讨[J]. 高校图书馆工作, 2001, (5):29—33.

6.唐铭杰. 论搜索引擎的发展概况及发展趋势[J]. 情报杂志, 2001, (5):70—71.

7.Lawrence S., Gilesc L. Accessibility of information on the Web[J]. Nature, 2003, 400(7):107-109.

8.楮亚萍,张华等. 搜索引擎的现状及分析[J]. 计算机与现代化, 2001,

9.唐春生，金以慧. 基于代理机制的 Internet 信息自动提取[J]. 计算机工

10.程与应用，2001;37(10):38~41.

11.曾春，邢春晓，周立柱. 个性化服务技术综述[J]. 软件学报, 2003;23(5):21~26.

12.赖茂生. 计算机情报检索[M]. 北京:北京大学出版社,1993.

13.彭一中等. 网络信息资源检索[M]. 长沙: 湖南大学出版社, 2002.

14.郭家义. 网络信息检索效率研究[J]. 图书与情报. 2003,10(2):51~54.

15.钟涛，陈新明，万均，张世勇. 中文文本 WEB 搜索引擎的设计与实现.

16.计算机工程与应用; 2001, 17(10):149-151.

17.欧洁. 基于相关术语集的搜索引擎选择. 计算机科学; 2003，7（1）: 42~45.

## 作者简介

梁伟贤，1972 年 9 月出生，男，汉族，广东南海人，硕士，助理工程师 . 毕业院校: 华南理工大学计算机学院计算机软件工程专业. 从 1996 年起至今，在广州电信局长期从事宽带网络（包括光纤城域网、ATM、ADSL、会议电视、远程教育、远程医疗等系统）等网络运营、维护、工程、调试、调单调度等重要工作，对互联网现状和发展有独特的见解和认识. 自 90 年代中至今，见证了广州市电信 INTERNET 互联网发展的整个过程. 现任职广州市电信天河分公司转型业务拓展部，具体负责信息模式转型工作和楼宇业务、系统集成项目拓展及营销支撑等工作，对目前中国电信互联网宽带业务、转型业务及营销策划等各项工作具有深刻了解和认识.

# 番禺电信网络智能化改造

余海荣

■ 中国电信南沙分公司

**摘　要:** 本文从从网络管理的角度，对番禺电信本地网提出智能化改造和优化方案，借鉴移动网络的用户数据属性管理方式，引入了集中数据库，既节省了全网的中继资源，又节省了大量的信令资源，优化了汇接局的资源配置，规范了本地交换网络的数据资源。

**关键字:** 电信 网络 NGN 智能化 改造

番禺电信固定电话网目前是一个由长途局、网关局，汇接局（TM、MLS）和端局（LS）组成的三个层面二级汇接网络结构，三个层面分别是长途（含互联互通）层、汇接层、端局层。目前番禺区内的所有端局均开高效直达路由经 PYTM1、PYTM2、东城 MSL、南城 MSL、北城 MSL 汇接至 DC1 长途局和网间接口局，部分话务量大的端局也开设高效直达路由至 DC1 长途局和网间接口局。近几年来，随着用户数和话务量的剧增，本地网汇接局的处理能力以及业务功能均跟不上发展需要，主要表现在以下几点。

A. 番禺电信汇接局大多数机型较旧，交换机路由功能差，智能化程度不高，新业务开放能力弱，新功能实施速度慢，无法适应多运营商环境下新的技术规范要求。若进行升版改造需要较大投入，而且部分设备供应商由于生产战略上的转变，不能提供后续技术支持，不利于新业务的开发和推广。

B. 番禺电信现网的汇接局所使用的交换机寿命大部分已超过或接近 10 年，设备老化，故障率较高。

C. 由于设备性能的局限性，汇接局多处于满负荷工作状态，很多局点已无扩容空间，只有通过不断增加