软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2012,23(2):253−265 [doi: 10.3724/SP.J.1001.2012.04140]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

分布式搜索引擎系统效能建模与评价*

张伟哲+, 张宏莉, 许 笑, 何 慧

(哈尔滨工业大学 计算机科学与技术学院,黑龙江 哈尔滨 150001)

Distributed Search Engine System Productivity Modeling and Evaluation

ZHANG Wei-Zhe+, ZHANG Hong-Li, XU Xiao, HE Hui

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: E-mail: wzzhang@hit.edu.cn, http://pact518.hit.edu.cn

Zhang WZ, Zhang HL, Xu X, He H. Distributed search engine system productivity modeling and evaluation. *Journal of Software*, 2012,23(2):253−265. http://www.jos.org.cn/1000-9825/4140.htm

Abstract: This study extends the current productivity models for a typical Web search engine system, which consists of a Web crawling system and an indexing system. Five different design schemata are characterized according to this model and are compared through power consumption, networking cost, system scale, and query efficiency. The half-WAN scheme, which consists of a WAN-based crawling system and a multi-cluster indexing system, is proved to be the best choice for a large-scale highly-efficient Web search engine.

Key words: distributed search engine; productivity modeling; distributed crawling; distributed indexing; half-WAN-based search engine

摘 要: 针对分布式搜索引擎系统效能建模与评估问题,通过对当前分布式搜索引擎系统的建模与分类,扩展了能耗与网络开销的成本模型;对 5 种构建搜索引擎系统的设计方案,从系统成本、系统规模和查询响应时间等角度进行了详尽的理论分析与评价.由此发现,由广域网分布式采集系统和多机群索引系统组成的半广域网搜索引擎系统相对于其他系统具有相对较高的效能,同时能够较好地兼顾用户的服务质量.

关键词: 分布式搜索引擎;效能建模;分布式采集;分布式索引;半广域网搜索引擎

中图法分类号: TP311 **文献标识码**: A

近 20 年来,Web 信息持续增长以及其更新频度快速提高,使得 Web 搜索引擎系统不断扩大自身规模,其系统结构也从单机到多机,从集中式到分布式逐步演进.而今,分布式系统结构已经成为构建 Web 信息搜索引擎系统的优选方案之一[1,2].

传统搜索引擎系统的评价标准主要包括:(1) 采集吞吐量,即全系统单位时间内能够下载的总 Web 信息量,用于衡量采集系统的效率;(2) 用户查询吞吐量,即全系统单位时间内能处理的用户查询请求数,用于衡量索引系统的效率;(3) 单用户查询响应时间,即用户向搜索引擎系统发出查询请求与得到系统响应的时间间隔,用于保证用户服务质量.主流厂商多通过不断建立及扩大数据中心高性能机群规模,增加互联网基础设施的网速与

* 基金项目: 国家自然科学基金(61173145); 国家重点基础研究发展计划(973)(G2011CB302605); 国家高技术研究发展计划(863)(2010AA012504, 2011AA010705)

收稿时间: 2011-07-08; 修改时间: 2011-09-06; 定稿时间: 2011-11-14

带宽来提升上述性能指标.然而,该方法导致系统成本与能耗急剧增加,效能问题逐渐成为分布式搜索引擎面临的主要矛盾.如何使用更少的资金构建更大、更快的搜索系统,已经成为企业界重点关注的问题.

当前,分布式搜索引擎体系结构的设计从单纯追求高性能向成本与性能并重的高效能转变.人们往往认为,基于单数据中心的集中式系统比分布式系统更易维护,运营成本(包括能耗和网费)更低廉,因此到目前为止,大型的商用分布式搜索引擎并不多见.然而,如果放弃巨大的数据中心,转而采用多个较小的数据中心甚至单个计算机来组建上述系统,其成本降低将有可能抵消分布式结构所带来的种种开销[3].例如在采集系统方面,由于每个网页的采集时间与采集节点到网站的网络延迟或 RTT 存在线性关系[4],采集节点离网站距离越近,下载网页所需的时间越短.因此,单位时间内采集系统能够下载更多的网页,在系统成本不变的前提下增加吞吐量;此外,保持吞吐量不变,采集系统可以减少所需的节点机数量,从而降低系统成本.缩短采集节点与网站的距离,可以使数据中心和采集节点向 Web 而非企业方向偏移,能够优化利用网络局部资源.

成本较低但效率不高的系统同样也会被市场淘汰.例如,搜索引擎系统虽然可以通过减少网页下载量来节省成本,但会导致用户查询信息的新鲜程度下降.此外,用性能较弱的计算机替换性能较强的服务器,虽然能够降低系统成本,但是会导致系统规模的大幅增加.在众多个人计算机中查找信息的时间会比在服务器上查找信息的时间长,进而延长用户查询的响应时间,引发用户不满.因此,依托硬件基础设施以及软件体系结构,从效能的角度重新审视搜索引擎系统的开发和运行,追求系统的性能功耗比和数据处理能力均衡发展,保障系统资源的有效利用,具有重要的研究意义.

现有的分布式 Web 信息采集系统乃至分布式搜索引擎系统都缺乏对系统构建中的最根本要素——系统效能——的建模与评估.为此,本文致力于为各种搜索引擎架构(包括采集系统架构和索引系统架构)建立效能模型.在相同的采集和查询吞吐量前提下,在效能模型的基础上对各种架构进行成本比较,同时兼顾用户响应时间,最终得出表现最优的系统.本文借鉴分布式搜索引擎效能建模方法论,针对分布式采集与分布式索引系统,对 4 种搜索引擎系统的设计方案从系统成本、系统规模和查询响应时间的角度进行了详尽的理论分析与综合评价.

本文第 1 节分析分布式搜索引擎的研究现状及其在效能评估方面存在的局限性.第 2 节对当前分布式搜索引擎体系结构进行建模与分类.第 3 节给出分布式搜索引擎效能建模的方法论.第 4 节分别针对不同的分布式搜索引擎体系结构,从采集系统和索引系统进行效能建模与评价.第 5 节提出效能最优的半广域网搜索引擎系统体系结构.第 6 节对本文进行总结.

1 相关工作

分布式搜索引擎的研究主要集中在分布式采集与分布式索引两个层面.

分布式采集:Yahoo 研究院的 Baeza-Yates 等人[2]定义分布式信息采集系统为“原则上某些节点可以分布于不同的地理或网络位置”.文献[5]中提出分布式系统具有高可扩展性并能有效降低 Internet 负载,进一步提出了分布式信息采集系统的分类方法、评价指标等基本概念,为分布式采集的研究奠定了基础.UbiCrawer[6]设计并实现了完全分布式、容错并可扩展的广域网分布式信息采集平台,提出一致性哈希为其负载均衡核心方法.Dustin 等人[7]对多种分布式采集系统进行了比较,提出分布式方法是解决采集系统带宽瓶颈的有效方法.IPMicra[8,9]是第一个基于位置信息调度的分布式采集系统;SE4SEE[10]实现了基于网格[11]的分布式采集系统;Apoidea[12]实现了基于 P2P 协议的完全分布式采集系统.国内学术界以北京大学的天网搜索引擎[13]的采集系统和上海交通大学的 Igloo 系统(IglooG[14])为代表,实现了基于网格服务的分布式 Web 信息采集.

分布式索引:Melink 等人[15]首次引入了大规模分布式索引的概念,并对文件分配和存储的相关技术进行了讨论和优化.Tomasic 等人[16]研究了分布式索引中倒排表的组织方法,以对查询请求的处理过程进行优化.Moffat 等人[17]对分布式索引的两种文档划分方案:文档划分(document partition)和词表划分(term partition)进行了对比;文献[18,19]分别讨论了文档划分和词表划分中的负载均衡问题.Li 等人[20]从带宽和存储空间等方面研究了 P2P 索引,发现完成 P2P 搜索需要比传统搜索引擎更多的资源.Zhang 等人[21]采用 P2P 模型建立了基于词

表划分的分布式索引.还有一些相关工作[22−24]对搜索引擎查询过程中的 Cache 系统进行了深入研究.

上述文献大多没有考虑运营分布式采集系统和索引系统过程中所遇到的与效能直接相关的能耗、网费等成本问题.Craswell 等人[25]虽然对比了集中式搜索引擎与元搜索服务等系统结构的成本,但是其中的元搜索服务并非分布式搜索系统,仅为能够处理搜索请求的逻辑单元.Baeza-Yates 等人在文献[1]中首次以经济成本的角度审视分布式搜索引擎系统,并提出了分布式搜索引擎的成本模型.他们提出在特定条件下,分布式搜索引擎的成本可能低于集中式搜索引擎.但是其成本模型存在不足之处:首先,没有对当前分布式搜索引擎的分类覆盖全面,仅对单机群和多机群搜索引擎进行了分析,忽略了广域网搜索引擎;其次,该文重点对索引子系统进行了效能评价,没有对信息采集子系统中多机群采集系统与广域网采集系统进行详尽的效能分析.此外,在对索引子系统进行分析时,该文仅对比了多机群索引系统与单机群索引系统,没有对多机群索引系统与广域网索引系统进行效能剖析.

2 分布式搜索引擎系统建模与分类

本节对当前的分布式搜索引擎体系结构方案进行建模与分类.搜索引擎体系结构模型如图 1 所示,其主体由两大子系统构成:采集子系统和索引子系统.两者都依赖于分布式存储器.这里不单独列出存储器的架构,而是将其作为采集系统和索引系统的一部分.

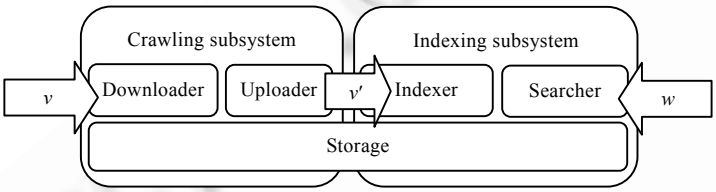


Fig.1 Architecture model of a typical search engine system

图 1 搜索引擎体系架构模型

采集系统输入为 Web 信息,由下载器(downloader)从 Web 下载.多数情况下,Web 信息以网页方式呈现(也有的以纯文本或二进制文件形式呈现,本文统称为网页或文档),量化此输入为 v,即单位时间进入采集系统的文档数量;采集系统的输出为文档快照,这一过程由上传器(uploader)完成,文档快照的输出速率等于 v',其数值应与 v 相近,否则会导致文档丢失.

索引系统有两种输入:(1) 采集系统的输出 v':文档进入索引系统后,首先经由索引器(indexer)写入索引,详细步骤为 Indexer 接收并存储文档快照,对文档进行分词,再将文档添入每个关键词的倒排表中;(2) 来自用户的查询请求,将其量化为 w,请求由检索器(searcher)处理,处理时需要在 Indexer 建立的索引中搜索.

搜索引擎的实现方案有以下几种(如图 2 所示):

(1) 单机方案(single-machine)

单台主机作为采集系统和索引系统.由于资源有限,因此吞吐量有限.

(2) 单机群方案(single-cluster)(后文用 S_0 表示)

系统由多台主机组合而成,且所有主机处于同一机群内部,通过高速局域网互联.这种机群通常需要上千台节点机,在管理、耗电、带宽、散热等方面消耗巨大,是目前主流商业搜索引擎的实现方案.

(3) 多机群方案(multi-cluster)(后文用 S_1 表示)

系统由多个相对较小的机群构成,每个机群分别具有一套单机群采集系统和一套单机群索引系统.各个机群之间通过租用线路互联,根据带宽使用情况缴纳租金.每个机群可以独立工作,也可以与其他机群协作.理论上,在相同吞吐量下,多机群方案与单机群方案相比能够在一定程度上降低成本,因此,随着云计算和大规模数据中心的不断涌现,该方案已经成为商业搜索引擎的演进趋势.

(4) 广域网方案(WAN-based or Internet-based)(后文用 S_2 表示)

系统由多台主机组合而成,但是各个主机分散于 Internet 各处,通常为 Internet 用户的个人主机,因此系统成本非常低.相应的采集系统和索引系统都是基于 P2P 方式实现,通过分布式路由调度采集任务和查找文档,通过冗余保证信息的可用性等.由于个人主机的能力有限,系统所需要的主机数量往往比多机群方案多出两个数量级.目前,已有若干公司建立了自己的完全分布式搜索引擎,但这些系统发展时间晚,因此用户规模不是很大.

此外,还存在一种半广域网系统,即采用广域网采集系统和多机群索引系统组成的分布式搜索引擎系统.虽然这种架构并没有在相关文献中被直接定义,但是在一些广域网采集系统的研究中[12,26,27],这样的系统结构被认为具有较大的发展前景和研究价值.

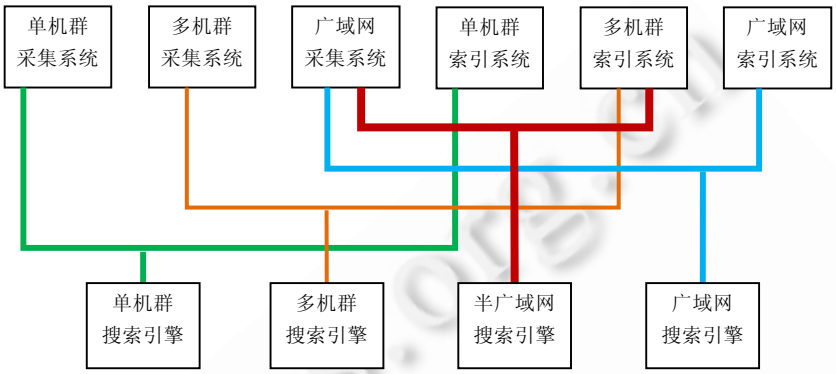


Fig.2 Classification of the multi-machine search engine systems with different crawling and indexing combination

图 2 多机搜索引擎系统中采集系统与索引系统的组合分类

3 分布式搜索引擎系统的效能建模

Baeza 等人在文献[1]中首次提出了分布式搜索引擎系统效能建模方法论.本节中多机群效能分析示例(公式(4)~公式(7))来源于文献[1].

分布式搜索引擎包含采集子系统和索引子系统,设Δt 时间内搜索系统总成本开销为 $Cost_{total}(\Delta t)$,采集子系统成本开销为 $Cost_{crawling}(\Delta t)$,索引子系统成本开销为 $Cost_{indexing}(\Delta t)$,则

$$Cost_{total}(\Delta t)=Cost_{crawling}(\Delta t)+Cost_{indexing}(\Delta t) \tag{1}$$

采集子系统成本开销主要包括信息采集节点的能耗成本 $E_{crawling}(\Delta t)$和采集过程中消耗的网络成本 $BW_{crawling}(\Delta t)$(见公式(2)).索引子系统承担索引与查询两项任务,因此其成本开销分别由索引过程中的能耗成本 $E_{indexing}(\Delta t)$、网络成本 $BW_{indexing}(\Delta t)$和查询过程中的能耗成本 $E_{query}(\Delta t)$、网络成本 $BW_{query}(\Delta t)$构成(见公式(3)).

$$Cost_{crawling}(\Delta t)=E_{crawling}(\Delta t)+BW_{crawling}(\Delta t) \tag{2}$$

$$Cost_{indexing}(\Delta t)=E_{indexing}(\Delta t)+BW_{indexing}(\Delta t)+E_{query}(\Delta t)+BW_{query}(\Delta t) \tag{3}$$

$E_{crawling}(\Delta t)$,$E_{indexing}(\Delta t)$和 $E_{query}(\Delta t)$通称为能耗开销 $E(\Delta t)$;$BW_{crawling}(\Delta t)$,$BW_{indexing}(\Delta t)$和 $BW_{query}(\Delta t)$通称为网络成本 $BW(\Delta t)$.

能耗开销模型:设机群 $S=\{S_1,\ldots,S_N\}$共 N 个,机群之间通过专用网络相连且分散于 Internet 多个位置.设 $C_w(\Delta t,i)$为机群 S_i 在Δt 时间内的总能耗,则Δt 时间内的总能耗 $E(\Delta t)$[1]等于所有机群在Δt 时间内的能耗之和,即

$$E(\Delta t)=\sum_{i=1}^{N}C_w(\Delta t,i) \tag{4}$$

设 $W(t,i)$为机群 S_i 在 t 时刻的能耗(单位为瓦特),单位功耗的成本为 u_w,则在$(t_1,t_1+\Delta t)$时段内,机群 S_i 的成本开销 $C_w(\Delta t,i)$[1]为

$$C_w(\Delta t,i)=\left(\int_{t_1}^{t_1+\Delta t} W(t,i)\cdot \mathrm{d}t\right)\cdot u_w \tag{5}$$

不失一般性,Web 采集、索引和查询等功能可同等抽象为通用操作 f.同一个节点机不会同时承担多种功能,则机群 S_i 的能耗 $W(t,i)$[1]实际上是由该机群的多种操作的能耗相加得到的,即

$$W(t,i)=\sum_f W_f(t,i) \tag{6}$$

设 $k_f(i)$为单位时间内系统分配给机群 S_i 的操作 f 的数量.设 $l_f(i)$为机群 S_i 执行操作 f 的平均时间开销,由于操作可以并发处理,所以实际 k 个操作的开销并不等于 $l_f(i)$与 k 的乘积.设同一主机上该操作的平均并发数为 $c_f(i)$,每台主机在单位时间内的能耗为 $e_f(i)$,则得到[1]:

$$W_f(t,i)=k_f(i)\cdot\frac{l_f(i)}{c_f(i)}\cdot e_f(i) \tag{7}$$

采集场景下,$k_c(i)$为单位时间内机群所要求下载的 Web 信息数(通常为网页数),$l_c(i)$为下载每个网页所需的的平均时间,$c_c(i)$为单机的并发连接数(设每个连接 1 次下载 1 个网页),则通过 $k_c(i)$,$l_c(i)$和 $c_c(i)$得到要在 t 时刻完成 $k_c(i)$所需的主机数,再乘以每台主机的能耗 $e_c(i)$,即得到机群 S_i 所需的能耗.

索引场景下,$k_q(i)$为单位时间内机群所要求处理的查询请求数,$l_q(i)$为处理每个查询请求所需的平均时间,$c_q(i)$为单机能够同时处理的查询请求数;同时,每台主机的能耗为 $e_q(i)$.

网络成本模型:多机群系统需要租用网络运营商(ISP)的线路进行网页下载,此外,多机群之间需要进行网络通信,以传递一些数据如跨域链接等.本文定义网络成本(所需传输数据总量与网费乘积)为所有机群网页下载的总成本 $\sum_{i=1}^{N}C_{web}(\Delta t,i)$ 与机群间通信成本 $C_{inter}(\Delta t)$之和.关于网络成本的更详细的量化过程见本文第 4 节.

$$BW(\Delta t)=\sum_{i=1}^{N}C_{web}(\Delta t,i)+C_{inter}(\Delta t) \tag{8}$$

以上模型没有考虑到机群散热问题,但是如果按照现在流行的集装箱式节点机放置方案,机群散热所需成本将与机群中的节点机数量成正比,可以认为是能耗中的一部分.

4 分布式搜索引擎体系结构效能评价

单机系统[28−32]吞吐量无法与多机系统相比.首先,由于存储空间以及接入带宽方面的限制,刀片服务器(小型机)甚至中型机,很难达到多机系统吞吐量;其次,虽然部分大型机和巨型机能够达到与大规模机群相当的硬件配置,但由于其价格与功耗都非常巨大,可扩展性较差,因此工业界一般不采用单机方案构建大型搜索引擎.本节着重量化多机搜索引擎不同体系结构的效能.

4.1 多机采集系统的效能评价

在相同吞吐量下,如果能够保证多机群系统下载网页的时间开销低于单机群系统,则多机群方案下系统能耗将低于单机群系统的能耗[1].然而,文献[1]中没有进一步剖析多机群与广域网采集系统的效能,本节致力于此.

4.1.1 多机群采集系统与广域网采集系统效能评价

本节分为两个部分:首先对多机群采集与广域网系统进行效能比较;然后,定性计算完全分布式方案下采集系统所需的系统规模,并讨论其实现的可行性.

在广域网采集系统下,由于采集节点并不在机群环境内,并且其维护成本由主机的贡献者承担,因此系统本身所需承担的成本仅消耗于接收经采集节点粗分析后得到的数据;同时,网络成本仅限于接收这些数据所造成的网络开销.

能耗开销:建立广域网采集系统 S_2 的效能模型.设 S_2 系统后有 N 个索引机群$\{S_{i2}:i\in\{1,2,\ldots,N\}\}$,每个机群在采集方面仅接收采集节点返回的数据,全系统中总控机群为 S_{21}.与多机群系统类似,Δt 时间内,S_{21} 机群需处理βP

个网页,其他机群分别处理 $P\cdot(1-\beta)/(N-1)$个网页.

在 S_1 系统中,开销 $l_c^1(i)$ 可以分解为下载网页和分析网页两者的开销之和,即 $l_c^1(i)=l_{c,d}^1(i)+l_{c,e}^1(i)$;在 S_2 系统中,$l_c^2(i)$ 为接收网页数据和解压缩两个过程的开销之和,即 $l_c^2(i)=l_{c,r}^2(i)+l_{c,u}^2(i)$.

在 S_2 系统中,采集节点获得网页后,需要先将其中的 HTML 结构化信息去除,提取出标题、链接、锚信息和正文等信息,压缩后再传输到机群式存储系统.设去结构化信息后原始信息与提取信息数据量之比为 $\omega_e=P_{raw}/P_e$,通常 $10\leqslant\omega_e\leqslant20$.网络传输数据时压缩比为 $\omega_c=P_e/P_c$,采用 gzip 压缩算法,通常 $2\leqslant\omega_c\leqslant10$.因此,网络传输所需的数据量为原始网页数据量的 $1/(\omega_e\cdot\omega_c)$.因此,$l_{c,d}^1(i)=\omega_e\cdot l_{c,r}^2(i)$.

在 S_1 系统中,分析网页开销 $l_{c,e}^1(i)$ 依赖于匹配算法对单位数据进行匹配的时间复杂度.HTML 匹配基于固定字符串集合,而基于自动机的匹配算法的时间复杂度一般为 $O(n)$,所以对于单个网页仅需对网页上的每个字符比较 1 遍.对于每个字节的比较操作,CPU 通常需要 4 个时钟周期,在没有流水线优化的情况下,使用目前常见的 2GHz 处理器,每个时钟周期为 10^{-9}s.静态网页文字内容平均大小为 10KB,则对一个网页匹配 1 遍所需的时间开销,即 $l_{c,e}^1(i)$ 应为 4ms;如果考虑流水线优化,则实际开销应小于 1ms.同时,在 2GHz 主频下,gunzip 的解压缩速度平均为 10MB/s,批量压缩的情况下,平均每个网页可以压缩到 2KB,则平均的 1 个网页的解压缩时间为 0.2ms< 1ms.由于 $l_{c,e}^1(i)$ 和 $l_{c,u}^2(i)$ 数量级相同,可近似认为 $l_{c,e}^1(i)\approx l_{c,u}^2(i)=l_c^{12}$.

在 S_1 系统中下载网页操作 $l_{c,d}^1(i)$ 和 S_2 系统中接收网页数据操作 $l_{c,r}^2(i)$ 为网络操作,其时间开销通常大于 10ms,所以 $l_{c,d}^1(i)=\omega_e\cdot l_{c,r}^2(i)>>l_c^{12}$, 则

$$l_c^1(i)\approx\omega_e\cdot l_c^2(i) \tag{9}$$

在 S_2 系统中,除总控机群之外的所有机群的 $l_c^2(i)$ 均约等于 l_0^2.同时,S_2 内机群间的分工与系统 S_1 相同,即 $\alpha=\beta$,则系统 S_2 在 t 时刻的能耗为

$$W^2(t)=v\cdot\alpha\cdot X\cdot l_c^2(1)+v\cdot(1-\alpha)\cdot X\cdot l_0^2 \tag{10}$$

直接对 $W^2(t)$和 $W^1(t)$进行比较,得到:

$$\frac{W^1(t)}{W^2(t)}=\frac{v\cdot\alpha\cdot X\cdot l_c^1(1)+v\cdot(1-\alpha)\cdot X\cdot l_0^1}{v\cdot\alpha\cdot X\cdot l_c^2(1)+v\cdot(1-\alpha)\cdot X\cdot l_0^2}\approx\frac{\omega_e}{1} \tag{11}$$

根据前面的分析,由于 $10\leqslant\omega_e\leqslant20$,所以 $W^1(t)>W^2(t)$,即系统 S_2 的能耗成本低于系统 S_1 的能耗成本.

网络成本:比较进入系统 S_1 和系统 S_2 的数据量的大小.由于系统从采集节点到机群所传输的数据是经过提取和压缩的,因此在两系统的机群具有相同网页到达速度的情况下,S_1 系统的第 i 个机群的带宽开销 BW_i^1 与 S_2 系统的第 i 个机群的带宽开销 BW_i^2 之间的关系为

$$BW_i^1=\omega_e\cdot\omega_c\cdot BW_i^2 \tag{12}$$

根据公式(9)和公式(10)可以计算出,如果要使系统 S_2 的能耗成本到达与系统 S_1 相等的程度,则系统 S_2 的吞吐量将达到 $\omega_e\cdot P/\Delta t$.同时,S_2 系统的带宽开销仅变为 $\omega_e\cdot BW_i^2$,即 $BW_i^2=BW_i^1/\omega_c$,仍然小于 S_1 系统的带宽开销.

综上所述,在相同的网页到达速率下,系统 S_2 的成本明显低于系统 S_1 的成本.但是,S_2 系统的优势依赖于采集节点机贡献者的多少,需要保障贡献者的群体足够大以至于可以使索引机群满足吞吐量 $P/\Delta t$.我们将在第 4.1.2 节分析 S_2 系统所需采集节点的规模.

讨论:本节的内容基于两个假设:(1) 在分析 S_2 系统采集模块的能耗开销时没有考虑贡献者资源的能耗.深入分析如下:S_2 系统需要先在分散的采集节点上下载网页,其时间开销包括采集节点下载网页 $l_{c,d}^2(i)$、分析网页 $l_{c,e}^2(i)$ 以及索引节点接收网页 $l_{c,r}^2(i)$、解压缩网页 $l_{c,u}^2(i)$,其能耗总和为 $l_{c,d}^2(i)+l_{c,e}^2(i)+l_{c,r}^2(i)+l_{c,u}^2(i)$.而 S_1 系统仅包括下载网页 $l_{c,d}^1(i)$ 和分析网页 $l_{c,e}^1(i)$ 两者开销之和,即 $l_c^1(i)=l_{c,d}^1(i)+l_{c,e}^1(i)$.根据第 4.1.1 节的分析,考虑了贡献者能耗的 S_2 系统效能显然要大于 S_1 系统的效能.然而,当前广域网信息采集系统在实现过程中一般均基于志愿者计算模型[33],目的是汇集全球个人计算机提供面向 SETI@home 的分布式计算服务.一方面,利用个人计算机

空闲资源;另一方面,广域网采集系统的贡献者也可获得利润或更快捷的索引服务.例如,Faroo 为加入者反馈公司利润;Majestic 设立全球加入者贡献排名;YaCy 只允许加入者享有查询索引的权利,并提供个性化定制搜索.从完整的 S_2 系统角度考量贡献者能耗确实有所增加;但从搜索引擎构建者的视角,由于仅维护索引机群,能耗反而有所下降(即 S_2 系统能耗小于 S_1 系统的能耗).(2) 没有考虑文件系统对小文件的开销.补充分析如下:以存储 10KB 小文件为例,主流硬盘内部数据传输率(internal transfer rate)约为 100MB/s,因此单个网页存储时间约为 0.1ms;即使考虑多级缓存及缓存失效问题,文件访问开销仍然远低于 10ms 数量级的网络访问开销.而 S_1,S_2 均涉及到网页采集与网页接收,因此在第 4.1.1 节中没有考虑文件访问开销.

4.1.2 广域网采集系统所需采集节点规模分析

设 S_1 和 S_2 系统分别达到吞吐量 $v_i=P_i/\Delta t$ 所需机器数为 $n_1(i)$和 $n_2(i)$.S_1 系统直接将网页下载至机群,其时间开销为 $l_c^1(i)=l_{c,d}^1(i)+l_{c,e}^1(i)$,则 $n_1(i)$为

$$n_1(i)=v_i\cdot\frac{l_c^1(i)}{c_c^1(i)}=v_i\cdot\frac{l_{c,d}^1(i)+l_{c,e}^1(i)}{c_c^1(i)} \tag{13}$$

S_2 系统需要先在分散的采集节点上下载网页,而后转发至索引节点,则采集节点时间开销为 $l_{c,d}^2(i)+l_{c,e}^2(i)+l_{c,r}^2(i)+l_{c,u}^2(i)$;而索引节点接收并解压缩网页,则索引节点时间开销为 $l_c^2(i)=l_{c,r}^2(i)+l_{c,u}^2(i)$,则 $n_2(i)$为

$$n_2(i)=n_c^2(i)+n_r^2(i)=v_i\cdot\frac{l_{c,d}^2(i)+l_{c,e}^2(i)+l_{c,r}^2(i)}{c_c^2(i)}+v_i\cdot\frac{l_{c,r}^2(i)+l_{c,u}^2(i)}{c_r^2(i)}=v_i\cdot\frac{l_{c,d}^2(i)+l_{c,e}^2(i)+l_{c,r}^2(i)}{c_c^2(i)}+v_i\cdot\frac{l_c^2(i)}{c_r^2(i)} \tag{14}$$

其中,n_c^2 为采集节点数量,n_r^2 为机群上接收节点的数量,$c_c^2(i)$ 为采集节点的平均并发连接数.

根据公式(9)可知,$l_c^1(i)\approx\omega_e\cdot l_c^2(i)$,因此,

$$n_1(i)\approx\omega_e\cdot n_r^2(i) \tag{15}$$

根据网络延迟的三角不等性[34],S_2 系统的采集节点上由于转发造成的时间开销与系统 S_1 直接下载的时间开销满足三角不等性:

$$l_{c,d}^1(i)\leqslant l_{c,d}^2(i)+l_{c,r}^2(i)<\theta\cdot l_{c,d}^1(i) \tag{16}$$

其中,$\theta\cdot l_{c,d}^1(i)$ 为时间开销上限.如果两系统的采集节点在字符串匹配方面具有相同的性能,则可以得到:

$$n_c^2(i)=v_i\cdot\frac{l_{c,d}^2(i)+l_{c,e}^2(i)+l_{c,r}^2(i)}{c_c^2(i)}\geqslant v_i\cdot\frac{l_{c,d}^1(i)+l_{c,e}^1(i)}{c_c^2(i)}=v_i\cdot\frac{l_c^1(i)}{c_c^2(i)}=n_1(i)\cdot\frac{c_c^1(i)}{c_c^2(i)} \tag{17}$$

$$n_c^2(i)=v_i\cdot\frac{l_{c,d}^2(i)+l_{c,e}^2(i)+l_{c,r}^2(i)}{c_c^2(i)}<v_i\cdot\frac{\theta\cdot l_{c,d}^1(i)+l_{c,e}^1(i)}{c_c^2(i)}\approx v_i\cdot\frac{\theta\cdot l_c^1(i)}{c_c^2(i)}=n_1(i)\cdot\theta\cdot\frac{c_c^1(i)}{c_c^2(i)} \tag{18}$$

考虑到系统 S_2 中采集节点的不稳定性,增加一个成功率因子 η_2,$0<\eta_2<1$,假设系统 S_1 的采集节点不存在不稳定性问题.综合公式(14)、公式(15)、公式(17)、公式(18)可得:

$$n_1(i)\cdot\left(\frac{c_c^1(i)}{\eta_2\cdot c_c^2(i)}+\frac{1}{\omega_e}\right)\leqslant n_2(i)<n_1(i)\cdot\left(\theta\cdot\frac{c_c^1(i)}{\eta_2\cdot c_c^2(i)}+\frac{1}{\omega_e}\right) \tag{19}$$

以下就 S_2 系统的采集节点数 $n_c^2(i)$ 进行量化分析.设 S_1 系统的采集节点的并发能力与 S_2 系统采集节点的并发能力之比为 $c_c^1(i)/c_c^2(i)=10$;η_2=0.3 即 S_2 系统中采集节点执行采集任务的成功率为 30%;$n_1(i)$=100,即 S_1 系统一个机群中具有 100 台采集节机.则根据公式(17)得 n_c^2 的取值范围约为[3333,θ×3333).由于系统同样也对网络距离进行了优化,假设θ<2,则 n_c^2 的取值范围约为[3333,6667].能否在这个取值范围内尽量减少采集节点的数量,取决于采集节点能否最大限度地利用网络距离优势:一方面,缩短采集节点与目标网站之间的网络距离;另一方面,缩短采集节点与索引机群之间的网络距离.

4.1.3 多机采集系统效能比较结果

(1) 如果能够保证多机群系统下载网页的时间开销低于单机群系统,则多机群系统中采集系统的成本将低于单机群系统中的采集系统的成本.而通过降低网络距离,实现降低网页下载时间的目标是切实可行的;

(2) 如果从工业界角度,采集系统遵循志愿者模型,厂商仅需维护索引系统能耗成本,则广域网采集系统的成本低于多机群采集系统的成本,且广域网采集系统理论上所需达到的规模和吞吐量都是实际可行的.

综上所述,在相同吞吐量的前提下,广域网采集系统是构建多机采集系统的最优选择.

4.2 多机索引系统的效能评价

文献[1]中深入剖析了多机群索引相对于单机群索引系统优势,强调在特定条件下可能带来较低的成本.条件包括:需要较高的查询本地化率;采集系统的下载吞吐量和索引系统的查询吞吐量之间的比值应尽量接近;共享文档集合不能太大等.然而文献[1]中没有进一步剖析多机群索引与广域网索引系统的效能,本节致力于此.

4.2.1 多机群索引系统效能分析

文献[1]中的多机群索引成本仅探究了用户查询请求涉及的系统成本,而忽略了索引系统获得新文档后索引重建的成本,下面补充分析.

多机群索引重建成本:设每个文档加入倒排表的时间开销为常数 l_{in},单位时间内的能耗为常数 $e_{in}(i)=e$,主机性能相同,则 $c_q(i)=c$,设全系统索引重建时文档加入速度等于采集系统接收文档的速度 $v=P/\Delta t$;机群 S_{j1} 采集的非共享文档在其采集的所有文档中的比重(称为本地化率 locality)平均为 y,则索引建立能耗为

$$W_{in}(t)=\sum_{i=1}^{N}W_{in}(t,i)=\sum_{i=1}^{N}\frac{l_{in}}{N}\cdot\frac{k_{in}(i)+\sum_{j:j\neq i}k_{in}(j)\cdot(1-y)}{c}\cdot e=\frac{l_{in}\cdot e\cdot v}{N\cdot c}\cdot(1+(1-y)\cdot(N-1)) \tag{20}$$

机群之间需要交换处于 G 集合中的共享文档;同时,为了对所有文档进行打分,需要在机群间传递关键词统计信息,主要是关键词在文档集合中的词频统计.因此,重建索引所需的网络成本包含共享文档交换和打分参数交换两个部分.设 u_{bw} 为维持单位带宽的包月费,σ为单位时间内各机群间交换打分参数的次数,函数 $Score(T)$表示全部 term 打分参数的大小(字节数),S_d 为平均文档大小(字节数),则重建索引所需的网络成本为

$$\begin{aligned}BW_{indexing}(\Delta t)&=\sigma\cdot\sum_{i=1}^{N}\sum_{j:j\neq i}Score(T)\cdot u_{bw}+\sum_{i=1}^{N}\sum_{j:j\neq i}k_{in}(i)\cdot(1-y)\cdot u_{bw}\\&=N\cdot(N-1)\cdot\sigma\cdot Score(T)\cdot u_{bw}+v\cdot(1-y)\cdot(N-1)\cdot S_d\cdot u_{bw}\end{aligned} \tag{21}$$

4.2.2 广域网索引系统所需索引节点规模分析

假设 S_2 系统采用词表划分方案,因为如果在完全分布式情况下采用文档划分方案,那么每完成一次查询将需要进行与系统中的索引节点数目相当的网络通信,对于每时每刻都在接收成千上万个查询请求的索引系统来说,这种广播式通信显然是不可行的.设 S_1 系统下一个索引机群所需的索引节点数量为

$$n_q^1(i)=v_i\cdot(1+(1-y)\cdot(N-1))\cdot\frac{l_{in}^1(i)}{c_{in}^1(i)}+w_i\cdot(1+(1-x)\cdot(N-1))\cdot\frac{l_q^1(i)}{c_q^1(i)} \tag{22}$$

为了方便比较,令 $y=1$,$x=0$,将 S_1 系统退化为最简单的没有任何优化的分布式索引系统.即

$$n_q^1(i)=v_i\cdot\frac{l_{in}^1(i)}{c_{in}^1(i)}+w_i\cdot N\cdot\frac{l_q^1(i)}{c_q^1(i)} \tag{23}$$

为了代替原本 S_1 系统下的一个索引机群,设 S_2 系统需要的索引节点数量为

$$n_q^2(i)=\gamma_{doc}\cdot v_i\cdot u_{term}\cdot\frac{l_{in}^2(i)}{c_{in}^2(i)}+\gamma_{term}\cdot w_i(term)\cdot\frac{l_q^2(i)}{c_q^2(i)} \tag{24}$$

其中,$l_{in}^2(i)$ 为 S_2 系统内每个索引节点索引一个网页所需的时间开销;u_{term} 为平均每个文档中包含的关键词数量;$w_i(term)$为对所有关键词进行关键词查询(term query)的流量,可简化为 $w_i(term)=w_i\cdot u'_{term}$,$u'_{term}$ 为平均每条查询语句所包含的关键词数;$l_q^2(i)$ 为该操作的时间开销;γ_{doc} 为每个文档的平均冗余度($\gamma_{doc}>1$);γ_{term} 为每个关键词的倒排表的平均冗余度($\gamma_{term}>1$).为了尽量保证索引数据的完整性,对每个关键词的倒排表及倒排表中的文档进行冗余是必须的实现环节.

观察公式(23)和公式(24),对于处于分母部分的参数,S_2 系统显然小于 S_1 系统;同时,对于处于分子部分的参

数,S_2 系统显然大于 S_1 系统,也就是说 $n_q^2(i)>n_q^1(i)$.并且两式右侧中最右边的加数,也就是用于处理查询请求的资源数,高于用于重建索引所需的资源数(左边的加数),即右边的加数决定了超过 50%的索引节点数(注意,索引重建和倒排表查询功能实际上处于同一节点上,因此我们说加数仅表示资源数,并不是实际节点数).基于这一假设,我们以下仅通过分析用于处理查询请求的资源数量定性的判断 S_2 系统的规模.在单机性能上, $l_{in}^2(i)=l_{in}^1(i)$, $l_q^2(i)=l_q^1(i)$, 那么对于右边的加数,S_2 系统与 S_1 系统的差异仅在于系数,将两个右加数比较,得到:

$$\frac{right\quad n_q^2(i)}{right\quad n_q^1(i)}=\frac{\gamma_{term}\cdot w_i(term)\cdot\dfrac{l_q^2(i)}{c_q^2(i)}}{w_i\cdot N\cdot\dfrac{l_q^1(i)}{c_q^1(i)}}=\frac{\gamma_{term}\cdot u'_{term}\cdot c_q^1(i)}{N\cdot c_q^2(i)} \tag{25}$$

如果 $c_q^1(i)/c_q^2(i)=10, u'_{term}=3$, γ_{term}=10,则上式变为 300/N.根据前面的分析,通常 2<N<5,因此,上式的比值取值范围为(60,150).

4.2.3 多机群索引系统与广域网索引系统效能评价

S_2 系统虽然能够在吞吐量上达到 S_1 系统的水平,但是完全分布式的系统结构并不能完美地将 S_1 系统再现,这将导致用户查询满意度的下降.这主要体现在:

(1) 系统复杂性极大地增加.将索引系统变为完全分布式结构在设计和开发上都将引入较高的复杂性,其复杂程度将付出远远超过将采集系统变为完全分布式结构的代价.

(2) 分布式查询的时间开销较大.虽然根据前面的分析,S_2 系统在足够的规模下能够达到与 S_1 系统相同数量的单位时间处理能力,但对于发出查询请求用户,单条查询语句处理时间仍然高于 S_1 系统.理论分析如下:

完全分布式索引系统为了保证存储、索引、检索、排序等操作的全局一致性,主要采用结构化 P2P 协议实现.这里采用最流行的 DHT 理论路由开销 $O(\log M)$进行衡量,设索引节点数为 M.

首先给出 S_1 系统对单条查询的时间开销,这里忽略用户主机到索引节点的延迟时间.

$$t_1=\max_i(t_{inter,i}+t_{search,i}) \tag{26}$$

$t_{inter,i}$ 为一次机群间通信的平均延迟时间,由于机群之间可以通过专用网络连接,所以延迟时间可以限制得比较短;$t_{search,i}$ 为机群本地查询时间,由于机群内部采用高速局域网连接,我们认为节点间通信的时间开销可以忽略不计,则 $t_{search,i}$ 也等于单节点的本地查询时间.由于查询是同时向多个机群并发发出的,因此最终时间开销由查询时间最长的机群决定,即 $\max_i(t_{inter,i}+t_{search,i})$.

说明一下 S_0 系统的响应时间.由于 S_0 系统所有节点机都处于同一个机群内,因此其查询时间应直接等于 $t_{search,i}$,但是可能由于机群规模过大,略大于 S_1 系统的 $t_{search,i}$,因此,这两个系统的查询响应时间主要区别在于 S_1 系统具有 $t_{inter,i}$.根据经验,由于多机群之间也是专用网络,$t_{inter,i}$ 平均值应与 $t_{search,i}$ 相当,则可近似地认为 S_1 的响应时间 t_1 比 S_0 系统的 t_0 增长了 1 倍.

然后给出 S_2 系统对单条查询的时间开销,同样忽略用户主机到索引节点的延迟时间.假设节点本地查询的时间开销与 S_1 系统相同.

$$t_2=\max_j(t_{route,j}+t_{search,j})\approx\max_j(t'_{inter}\cdot\log M+t_{search,j}) \tag{27}$$

查询的时间开销由查询时间最长的索引节点决定.由于使用结构化 P2P 组织节点,因此节点间查询通信时的延迟时间并不等于直接端到端通信的延迟时间,而是等于在P2P覆盖网上路由的时间.根据 $O(\log M)$的理论数值,将路由时间开销近似为 $t_{route,j}=t'_{inter}\cdot\log M$.如果 S_2 系统采取此表划分方式,那么其结构化 P2P 的实现必然是根据 1 个或若干个关键词的摘要值作为节点 ID.而这种以摘要值作为 ID 的路由方式,由于存在与网络实际拓扑不相匹配的问题,因此路由的每一跳的延迟时间 t'_{inter} 不可控制,设其等于互联网上任意两主机间延迟时间的期望值.设 $t'_{inter}=t_{inter,i}$. 根据前面的推导,假设 S_2 系统的索引节点数为 S_1 系统的 100 倍,S_1 系统具有 1 000 个索引节点,则 M 等于 100 000,于是 $t_{route,j}=t_{inter}\cdot 5$.如果 $t_{route,j}$ 最长等于 100ms,同时 $t_{search,i}$ 最长等于 100ms,则 t_1 等于 200ms, t_0 等于 100ms,而 t_2 等于 600ms,明显长于 t_1 和 t_0.并且考虑到以上假设的节点规模和路由跳数都较为保守,对于

大型搜索系统,t_2 的数值将会进一步增大,但是相比之下,t_1 和 t_0 的数值并不会有显著的变化.

综上所述,由于 S_2 系统上单条查询的时间开销引入了与系统规模有关的参数,从而造成其可扩展性受到影响,无法到达与 S_1 和 S_0 系统相同的较短的查询时间.

4.2.4 多机索引系统效能比较结果

(1) 多机群索引系统的成本在一定条件下能够低于单机群索引系统.首先,多机群体系结构更适用于采集吞吐量和查询吞吐量都较高的需求;其次,多机群索引系统必须提高查询处理的本地化率,以尽量降低成本.

(2) 广域网索引系统与前两者相比虽然能够大幅度降低系统成本,但是由于广域网索引系统的查询时间与系统规模相关,其查询时间远长于多机系统和单机群系统;同时,P2P 结构也对数据完整性构成困难.这两个缺陷将导致广域网索引系统对用户的服务质量下降,进而可能失去市场竞争力.

综上所述,多机群索引系统提供了降低成本可行性,同时具有可接受的查询响应时间,是构建多机索引系统的最优选择.

5 半广域网搜索引擎系统

经过上述效能评价我们发现,最优的系统结构是广域网采集系统与多机群索引系统的组合.该组合相对于其他系统具有较低的成本,同时能够较好地兼顾对用户查询的响应时间,其所需的系统规模在实践上也是可以实现的.我们将采用这一组合的搜索引擎系统称为半广域网搜索引擎系统.

在如图 3 所示的系统中,采集节点由安装在贡献计算机上的采集程序包括下载器和上传器的一部分组成,采集节点不仅要下载网页,而且要对网页内容进行提取,即还承担了一部分计算任务,之后交由本地上传器上传至机群.机群端主要包括采集系统上传器的一部分和索引系统.机群端的上传器(图中标为 Receip 的模块)接收来自于采集节点的网页信息,并提交给本地的索引机群(图中标为 Index Cluster 的模块).由于索引系统是分布式的,因此全局存在若干个这样的机群系统,各个机群通过专用网络连接;采集节点可根据其所在网络位置决定将网页信息上传给哪一个机群,从而进一步降低每篇文档的处理时间.

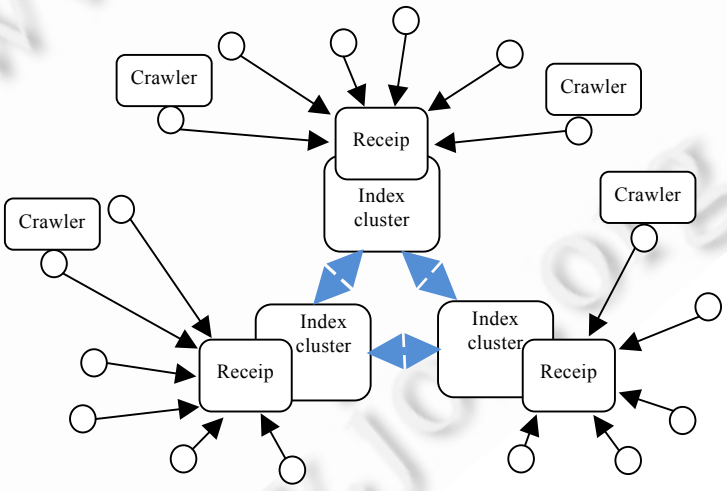


Fig.3 Architecture of the half-WAN-based search engine system

图 3 半广域网搜索引擎体系结构

将所有比较过的采集系统和索引系统进行组合,与半广域网搜索引擎系统放在一起画出变化趋势,如图 4 所示.从图中可以看出,半广域网系统正好处于成本、响应时间和系统规模的折中点:在其之前,基于机群的系统虽然规模相对较大,响应时间相对较短,但是成本偏高;在其之后,广域网系统虽然成本大为降低,但是响应时间和系统规模都呈现大幅增长.

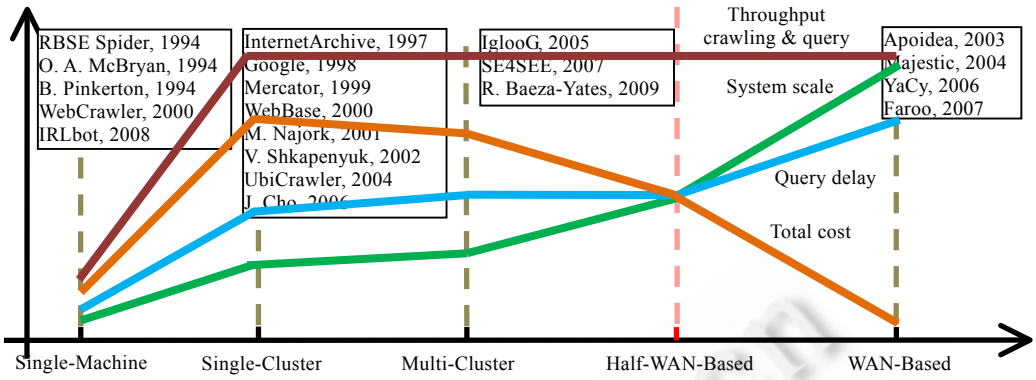


Fig.4 Productivity comparisons of different search engine design schemes

图 4 搜索引擎系统架构方案效能比较

6 结束语

本文通过对分布式搜索引擎成本、系统规模和响应时间这些效能指标的综合考量,对分布式搜索引擎体系结构设计中的效能建模与评估进行了深入研究.具体贡献如下:(1) 对当前主流的分布式搜索引擎体系结构模型进行建模与分类,根据单机群方案、多机群方案、广域网方案在分布式信息采集与分布式索引方面的不同组合,提出了单机群搜索引擎、多机群搜索引擎、半广域网搜索引擎和广域网搜索引擎这 4 种体系结构.(2) 从分布式搜索引擎采集子系统与索引子系统两个层面,分别建模并量化评估了多机群与广域网系统间的效能关系,发现广域网采集系统具有最低的系统成本,是构建多机采集系统的最优选择;而多机群索引系统提供了降低成本的可能,同时具有可接受的查询响应时间,是构建多机索引系统的最优选择.(3) 提出了具有广域网采集系统与多机群索引系统的半广域网搜索引擎系统,在相同的采集和查询吞吐量下,与其他已有系统相比具有较低的成本,同时能够较好地兼顾对用户查询的响应时间,是构建多机搜索引擎系统的最优选择,对未来的分布式搜索引擎体系结构研究具有一定的借鉴意义.

致谢 在此,我们向对本文的工作给予支持和建议的审稿人表示由衷的感谢.

References:

[1] Baeza-Yates R, Gionis A, Junqueira F, Plachouras V, Telloli L. On the feasibility of multi-site Web search engines. In: Proc. of the 18th ACM Conf. on Information and Knowledge Management (CIKM 2009). ACM Press, 2009. 425–434. [doi: 10.1145/1645953.1646009]

[2] Baeza-Yates R, Castillo C, Junqueira F, Plachouras V, Silvestri F. Challenges in distributed information retrieval. In: Proc. of the Int'l Conf. on Data Engineering (ICDE). Istanbul: IEEE Computer Society Press, 2007. 6–20. [doi: 10.1109/ICDE.2007.367846]

[3] Church K, Greenberg A, Hamilton J. On delivering embarrassingly distributed cloud services. In: Proc. of the 7th ACM Workshop on Hot Topics in Networks Hotnets. New York: ACM Press, 2008. 1–6.

[4] Cardwell N, Savage S, Anderson T. Modeling TCP latency. In: Proc. of the 19th Annual IEEE Int'l Conf. on Computer Communications 2000. IEEE Press, 2000. 1742–1751. [doi: 10.1109/INFCOM.2000.832574]

[5] Cho J, Garcia-Molina H. Parallel crawlers. In: Proc. of the 11th Int'l Conf. on World Wide Web. 2002. 124–135. [doi: 10.1145/511446.511464]

[6] Boldi P, Codenotti B, Santini M, Vigna S. Ubicrawler: A scalable fully distributed Web crawler. Software-Practice and Experience, 2004,34(8):711–726. [doi: 10.1002/spe.587]

[7] Dustin B. Distributed high-performance Web crawlers: A survey of the state of the art. 2003. http://www.cs.ucsd.edu/~dboswell/PastWork/WebCrawlingSurvey.pdf

[8] Papapetrou O, Samaras G. Ipmicra: An IP-address based location aware distributed Web crawler. In: Proc. of the 5th Int'l Conf. on Internet Computing (IC 2004). Las Vegas, 2004. 694−699.

[9] Papapetrou O, Samaras G. Ipmicra: Toward a distributed and adaptable location aware Web crawler. In: Proc. of the 8th East European Conf. (ADBIS 2004). 2004. http://www.sztaki.hu/conferences/ADBIS/12-Samaras.pdf

[10] Cambazoglu BB, Karaca E, Kucukyilmaz T, Turk A, Aykanat C. Architecture of a grid-enabled Web search engine. Information Processing and Management, 2007,43(3):609−623. [doi: 10.1016/j.ipm.2006.10.011]

[11] Hafri Y, Djeraba C. High performance crawling system. In: Proc. of the 6th ACM SIGMM Int'l Workshop on Multimedia Information Retrieval. ACM Press, 2004. 299−306. [doi: 10.1145/1026711.1026760]

[12] Singh A, Srivatsa M, Liu L, Miller T. Apoidea: A decentralized peer-to-peer architecture for crawling the World Wide Web. In: Proc. of the SIGIR 2003 Workshop on Distributed Information Retrieval. 2004. 126−142. [doi: 10.1007/978-3-540-24610-7_10]

[13] Li XM, Yan HF, Wang JM. Search Engine: Principle, Technology and System. Beijing: Science Press, 2005 (in Chinese).

[14] Liu F, Ma FY, Ye YM, Li ML, Yu JD. IGLOOG: A distributed Web crawler based on grid service. In: Proc. of the APWeb 2005. 2005. 207−216. [doi: 10.1007/978-3-540-31849-1_21]

[15] Melink S, Raghavan S, Yang B, Garcia-Molina H. Building a distributed full-text index for the Web. ACM Trans. on Information System, 2001,19(3):217−241. [doi: 10.1145/502115.502116]

[16] Tomasic A, Garcia-Molina H. Query processing and inverted indices in shared-nothing text document information retrieval systems. The VLDB Journal, 1993,2(3):243−276.

[17] Moffat A, Webber W, Zobel J, Baeza-Yates R. A pipelined architecture for distributed text query evaluation. Information Retrieval, 2007,10(3):205−231. [doi: 10.1007/s10791-006-9014-4]

[18] Badue CS, Baeza-Yates R, Ribeiro-Neto B, Ziviani A, Ziviani N. Analyzing imbalance among homogeneous index servers in a Web search system. Information Process Management, 2007,43(3):592−608. [doi: 10.1016/j.ipm.2006.09.002]

[19] Moffat A, Webber W, Zobel J. Load balancing for term-distributed parallel retrieval. In: Proc. of the 29th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2006. 348−355. [doi: 10.1145/1148170.1148232]

[20] Li JY, Loo BT, Hellerstein JM, Kaashoek F, Karger DR, Morris R. On the feasibility of peer-to-peer Web indexing and search. Lecture Notes in Computer Science, 2003,2735:207−215. [doi: 10.1007/978-3-540-45172-3_19]

[21] Zhang J, Suel T. Efficient query evaluation on large textual collections in a peer-to-peer environment. In: Proc. of the 5th IEEE Int'l Conf. on Peer-to-Peer Computing. 2005. 225−233. [doi: 10.1109/P2P.2005.7]

[22] Fagni T, Perego R, Silvestri F, Orlando S. Boosting the performance of Web search engines: Caching and prefetching query results by expoiting historical usage data. ACM Trans. on Information System, 2006,24(1):51−78. [doi: 10.1145/1125857.1125859]

[23] Saraiva PC, de Moura ES, Ziviani N, Meira W, Fonseca R, Riberio-Neto B. Ranking-Preserving two-level caching for scalable search engines. In: Proc. of the 24th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2001. 51−58. [doi: 10.1145/383952.383959]

[24] Baeza-Yates R, Gionis A, Junqueira F, Murdock V, Plachouras V, Silvestri F. The impact of caching on search engines. In: Proc. of the 30th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2007. 183−190. [doi: 10.1145/1277741.1277775]

[25] Craswell N, Crimmins F, Hawking D, Moffat A. Performance and cost tradeoffs in Web search. In: Proc. of the 15th Australasian Database Conf. (ADC). 2004. 161−169.

[26] Majestic-12: Distributed Web search. http://www.majestic12.co.uk/

[27] Loo BT, Cooper O, Krishnamurthy S. Distributed Web crawling over DHTs. Technical Report, CSD-4-1305, Berkeley: Technical Department of Electrical Engineering and Computer Sciences, University of California, 2004. http://repository.upenn.edu/cgi/viewcontent.cgi?article=1345&context=cis_papers

[28] Mcbryan OA. Genvl and WWWW: Tools for taming the Web. In: Proc. of the 1st Int'l World Wide Web Conf. 1994. 79−90.

[29] Pinkerton B. Finding what people want: Experiences with the webcrawler. In: Proc. of the 1st World Wide Web Conf. Geneva, 1994. http://thinkpink.com/bp/WebCrawler/WWW94.html

[30] Pinkerton B. Webcrawler: Finding what people want [Ph.D. Thesis]. Washington: University of Washington, 2000.

[31] Lee HT, Leonard D, Wang XM, Loguinov D. Irlbot: Scaling to 6 billion pages and beyond. In: Proc. of the Int'l World Wide Web Conf. ACM Press, 2008. 427−436. [doi: 10.1145/1541822.1541823]

[32] Eichmann D. The rbse spider—Balancing effective search against Web load. In: Proc. of the 1st Int'l World Wide Web Conf. 1994. 113−120.

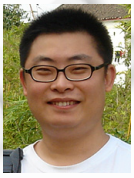
[33] Anderson DP, Cobb J, Korpela E, Lebofsky M, Werthimer D. SETI@home: An experiment in public-resource computing. Communications of the ACM, 2002,45(11):56−61. [doi: 10.1145/581571.581573]

[34] Francis P, Jamin S, Jin C, Jin YX, Raz D, Shavitt Y, Zhang L. IDMAPS: A global Internet host distance estimation service. IEEE/ACM Trans. on Networking, 2001,9(5):525−540. [doi: 10.1109/90.958323]

附中文参考文献:

[13] 李晓明,闫宏飞,王继民.搜索引擎:原理、技术与系统.北京:科学出版社,2005.

张伟哲(1976－),男,黑龙江哈尔滨人,博士,副教授,主要研究领域为网络计算,网络安全.



许笑(1983－),男,博士生,主要研究领域为网格计算,分布式系统.

张宏莉(1973－),女,博士,教授,博士生导师,主要研究领域为网络安全,网络计算.

何慧(1974－),女,博士,副教授,主要研究领域为网络计算,网络安全.