

支持 Web 深层数据库网格的部分关键技术的研究*)

申德荣 聂铁铮 余恩运 寇月 于戈

(东北大学信息科学与工程学院 沈阳 110004)

摘要 当前,深层 Web 数据库数量急剧增加,然而其知识并没有得到有效的利用。本文提出将特定的深层 Web 数据库应用于网格环境中的思想,并针对支持深层 Web 数据库网格的部分关键技术进行研究,主要包括:(1)深层 Web 数据库元信息定义模型与模式抽取模型的研究;(2)多层次的模式匹配模型和自协调模型研究;(3)基于属性松弛的 Web 数据库查询与集成模型研究;(4)多目标函数代价模型和面向局部性的自适应优化调度模型研究。研究成果将为构建深层 Web 数据库网格提供良好的支持,就像网格的概念所定义的一样,为用户提供统一的接口,可按需为消费者提供集成的深层 Web 数据知识。其具有广阔的应用前景。

关键词 数据库网格,深层 Web,模式匹配,近似查询,数据集成优化

Study on Some Key Technologies Supporting Deep Web Database Grid

SHEN De-Rong NIE Tie-Zheng YU En-Yun KOU Yue YU Ge

(College of Information Science and Engineering, Northeastern University, Shenyang 110004)

Abstract Nowadays, the number of deep Web databases is increasing rapidly, while their knowledge cannot be utilized effectively. This paper provides the idea of applying deep Web databases to grid environment, and focuses on some key technologies supporting deep Web database grid, mainly including: (1) Study on meta information model definition of deep Web databases and schema extraction. (2) Study on multi-level schema matching model and self-adaptive model. (3) Study on Web database query and integration model based on attributes relaxes. (4) Study on multi-object cost function model and local-oriented adaptive optimization schedule model. The research results of the project are aiming to give better support for building deep Web database grid system, just as the definition of grid, and to provide integrated deep Web data knowledge to users by means of uniform portal on demands. These key technologies will play an important role in future applications.

Keywords Database grid, Deep Web, Schema matching, Similar query, Data integration optimization

1 引言

深层 Web(Deep Web)数据库(源)^[1,2]是指网络中的自治数据库,通过页面中的表单(form)给用户提供唯一的访问方式,数据库模式对用户透明。深层 Web 数据库网格就是应用网格技术实现深层 Web 数据库资源的共享,借用网格高效的并行处理能力,达到为用户高效地提供集成的深层数据库资源的目的。

随着 Web 的发展,Web 上的信息呈爆炸式增长,然而,虽然 Web 上的信息量大而丰富,但真正被应用的却很少,尤其是结构化的信息。Web 上已有众多的公共的数据库信息,据 2002 年统计^[1],Web 上有 330000 个在线数据库,而到 2004 年,Web 上已有 450000 个^[2]。这些数据库主要是商业数据库,如图书等销售信息库,以及科研领域的数据库如生命科学、天体物理等。随着技术的发展和社会的进步,Web 上深层数据源将会急剧增加。

然而,目前 Web 上的信息获取还主要基于 Google、yahoo、百度等搜索引擎,为人们提供导航的信息。而对于 Web

上深层的数据库知识,虽然我们可以通过逐一点击的方式获取需要的信息,但需要人们繁琐的信息收集过程,给人们带来许多不便。尽管如此,用户得到的还是零散的、不完备的且非系统的信息。

面对深层 Web 数据库知识利用不足,以及数据库网格概念的提出,本文提出深层 Web 数据库网格系统。其将特定的深层 Web 数据库应用于网格环境中。Web 上深层数据库资源的动态性迎合了数据库网格内资源的特点,同时,网格的高效和智能性也为深层 Web 数据库资源的获取和整合提供了可靠的环境保证。本文针对支持深层 Web 数据库网格的部分关键技术进行研究,主要包括模式抽取模型、模式匹配模型、近似查询与集成处理模型以及自适应的高效的调度模型等,目的是为构建深层 Web 数据库网格提供坚实的技术支持。

2 相关工作

本文研究内容将涉及数据库网格、异构数据库集成、信息检索、自动推理等相关理论和技术,重点针对支持深层 Web

*)基金项目:该课题得到国家自然科学基金(编号:60673139),国家 863 项目(编号:2003AA414210)资助。申德荣 教授,博士,CCF 会员,主要研究方向为分布式系统,数据网格以及 Deep Web。聂铁铮 博士研究生,主要研究方向为数据库网格,Deep Web。余恩运 研究生,主要研究方向为数据网格,Web 信息处理。寇月 博士研究生,主要研究方向为资源发现,近似查询。于戈 教授,博士生导师,CCF 会员,主要研究方向为数据库技术,数据挖掘,Web 技术等。

数据库网格的模式抽取模型、模式匹配模型、近似查询与集成处理模型以及自适应的高效的调度模型进行研究。目前,国内外相关工作介绍如下。

随着深层 Web(Deep Web)数据源数量的增加,其重要性日趋突现。其研究也在基于 spider 或 crawler 引擎获取数据库的基础上提出了自动集成其数据的新需求。已有的深层 Web 数据源模式抽取^[3-5]主要以输入模式为重点,基于查询接口的输入表单信息以及表单中的标签元素的结构与类型定义接口模式。但是,要实现深层 Web 数据自动集成,只提供输入接口模式元信息是不够的,还需要考虑抽取接口的输出模式。

针对 Web 数据库的近似查询的研究典型的是 Ullas Nambiar^[6,7],其是针对单数据库的近似查询进行研究。但有关网格环境下的深层 Web 数据库的近似查询还没有见到。

已有 Web 数据库的模式匹配研究^[8,9]典型遵循以下规则:先对用户查询接口进行模式匹配处理,而后进行聚类分析,接着将用户查询分解和重写,并映射到查询接口上进行查询处理,最后将所得查询结果按照查询逻辑进行集成。但研究重点在接口模式匹配与映射上。

有关数据库网格的研究,国内外还处于初级阶段,典型的研究是 Globus DAI^[10]和 DQP^[11]。DAI 将数据库包装为 Web 服务实现数据库服务调用。DQP 是构建在 OGSA-DAI 之上的静态查询,强调并行执行以提高效率。这些典型研究致力于为网格环境下的数据库集成处理提供标准规范,但对网格环境下的资源动态性方面的支持没有讨论。

典型的自适应查询处理研究典型有:执行过程中根据运行反馈动态调整执行计划^[12]的 Ginga 模型和 Telegraph 项目^[13];通过逐步修正部分查询结果以达到获得最终完整的查询结果的目的 Niagara Internet 查询系统^[12];采用“事件-条件-动作”规则协调适应性行为的华盛顿大学的 Tukwila 数据集成系统^[14];为流式数据监控应用设计的 Aurora 系统^[15]等等。但针对动态 Web 数据库查询的自适应机制还没有见到。

面对与日剧增的深层 Web 数据库以及其信息利用不足的特点,本文结合网格技术与数据库技术,研究支持构建深层 Web 数据库网格的关键技术,目的是为用户提供一个具有一致接口的共享深层数据库资源的数据库网格系统,弥补深层数据库资源信息利用不足的缺憾。

3 深层 Web 数据库网格层次结构

目前,有关数据库网格技术和深层 Web 数据库访问技术的研究在国内外还都处在初级阶段,可见的报道不多。另外,由于深层 Web 数据库资源的元信息需要自捕获和自组织,无法达到 100%的正确,以及 Web 需要支持近似查询等需求,都给网格内的数据处理增加了难度。本文提出的深层 Web 数据库网格的层次结构如图 1 所示。层次说明如下:

资源层是 Web 环境下的深层数据库资源,Web 表单是访问资源的唯一方式。

知识层负责获取和组织深层 Web 数据库资源元信息。主要包括:抽取元信息;基于元信息将深层 Web 数据库资源包装为 Web 服务并自动注册于元信息仓中;基于元信息构建模式匹配知识仓。

服务层是支持深层 Web 数据库网格服务的核心服务组件。主要包括:资源发现、资源调度、查询处理、数据集成与服务质量等,为按需集成深层 Web 数据库资源提供支持。

用户层为网格提供给用户的一致访问门户。

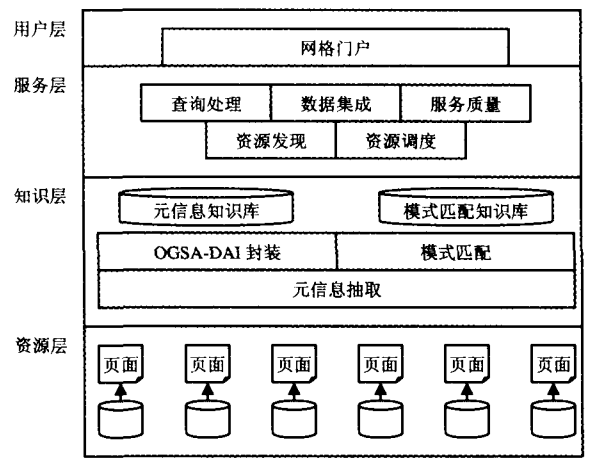


图 1 深层 Web 数据库网格

4 部分关键研究内容

当前,深层 Web 数据源信息利用严重不足,本文提出深层 Web 数据库网格概念,目的是利用网格高性能实现深层 Web 数据源信息有效集成和共享。该网格将为用户提供统一的门户,并为用户高效智能地获取集成的数据信息提供使能的环境。在研究过程中,我们将基于多数据库网格技术,借鉴分布式数据库理论,结合 IR 技术、不确定理论、智能推理理论、Web 服务和数据挖掘等理论知识。本文重点针对以下几个关键问题进行探究。(1)深层 Web 数据源元信息抽取模型;(2)基于抽取的元信息研究模式匹配模型,支持按需准确地发现深层 Web 数据库资源;(3)研究网格门户支持的近似查询模型和近似集成模型,支持按需为用户提供相关的近似匹配结果;(4)研究多目标函数代价模型和动态自适应的优化调度模型,达到提高网格获取数据信息的效率和满足个性化需求的目的。

4.1 深层 Web 数据库元信息定义模型与模式抽取模型的研究

通常,数据库网格内的数据库资源服务元信息(模式信息和访问信息等)是由用户遵照标准规范描述并自主注册于资源仓的,网格根据服务注册的元信息可方便实现数据库资源服务的访问。然而,深层 Web 数据库的元信息只能从其提供的表单抽取获得,而对深层 Web 数据库的访问也需要将服务请求映射到 Web 表单上实现,因此,标准的数据库网格的元信息定义规范如 DAI 不适合于深层 Web 数据库的元信息描述。为此,研究深层数据库资源的元信息定义模型以及模式抽取模型。

(1)深层 Web 数据库元信息定义模型(MetaM)研究。深层 Web 数据库作为网格中的服务资源,需要“提供”访问输入接口和结果输出接口。由于深层 Web 数据库资源的元信息是通过分析其提供的 Web 页面而获得的,并且深层 Web 数据库的访问也需要映射到 Form 表单上实现。因此,在输入接口中应描述输入接口到 Web 表单的映射,同理,输出接口也是如此。因此,深层 Web 数据库元信息的定义模型可描述为: $MetaM = (IS, OS)$, 其中 $IS = (A, AR, MI)$ 为输入接口描述, $OS = (A, AR, MO)$ 为输出接口描述; A 为接口模式属性集, AR 为属性间关系集合, MI 和 MO 分别是表单与输入接口和输出接口模式的映射。

(2)模式抽取模型的研究。面对只提供和只能获取有限

知识的深层 Web 数据库资源,要正确获取数据源的模式信息是非常困难的。为此考虑借用 Web 上相关的知识,如数据源的分类信息、领域信息、查询接口可提供的实例信息、同类接口中的属性实例信息等。通过综合分析、处理这些可获得的有限信息,从而确定模式抽取规则。重点研究基于正则表达推演机制的模式抽取推理规则,以提高模式抽取的准确度。

4.2 多层次的深层模式匹配模型的研究

已抽取的深层 Web 数据库的有限元信息是网格内资源定位依据的基础数据。然而在 Web 动态环境中,只是简单地依赖两两模式匹配很难获得有效而准确的结果,若按需实时推理出最大的集成模式集时间代价又太大。为此,本项目提出通过重复利用已有模式匹配知识组成一个语义上下文的模式匹配信息库,简称模式匹配知识库。为网格按需发现资源提供稳定、快速且准确的模式匹配知识。同时通过自协调机制保证知识库实时反映网格内的数据资源模式的语义关系。主要研究内容:

(1) 多层次的模式匹配模型(MMatchM)研究。MMatchM 是构建模式匹配知识库所依赖的模式匹配模型。首先根据相似距离匹配和结构匹配等进行第一层匹配,获得初始的匹配集 $InitMthSet = \{\langle S_i, S_j \rangle | SimaValue_{\langle S_i, S_j \rangle} > 0\}$, $SimValue_{\langle S_i, S_j \rangle}$, 表示模式 S_i 和 S_j 的相似匹配度;之后,利用候选模式匹配对间的关系以及历史模式匹配知识库,借鉴神经元影响模型和路由自适应算法 RIP 的迭代思想,重点研究自迭代地发掘模式元素间深层语义匹配关系的匹配模型;同时研究结合启发式规则和历史匹配知识上下文的合理的阈值 ϵ 的确定策略,最终获得确定性匹配集 $SFinaMthSet = \{\langle S_i, S_j \rangle | SemSimValeu_{\langle S_i, S_j \rangle} > \epsilon\}$, $SemSimValue_{\langle S_i, S_j \rangle}$ 为模式 S_i 和 S_j 间基于语义的相似匹配度值。

(2) 自协调模式匹配模型研究。Web 上的数据是不断更新和变化的,因此,需要动态协调模式匹配知识库内的信息。我们把模式及模式间的匹配关系看作一张无向图,图中的点为模式,图中的线为模式间的匹配关系矩阵,当任何一条线发生变化时,要影响相邻的点和线。本部分重点研究基于图的协调模型。

4.3 基于属性松弛的深层 Web 数据库查询与集成模型的研究

具有强大的查询功能是网格的典型特征,如网格不仅支持精确查询也支持不确定查询,并且具有对结果集成、排序和整理的能力。深层 Web 数据库网格也是如此,希望在不需要附加信息的情况下能为用户提供尽量满意的结果。然而,深层 Web 数据库的模式信息是在 Web 上自动捕捉的,其准确性取决于抽取模式算法的正确性;另外,数据库只支持精确查询,且从 Web 上获得的知识有限,也可能存在许多不确定的数据信息如 null 值、or-set、unknown 和 range 值等。因此,研究网格内支持不确定数据的近似查询与集成更具有挑战性。

(1) 基于属性松弛的深层 Web 数据库查询模型(RIQueryM)研究。由于 Web 上的数据库只支持精确查询,因此,研究如何依据精确的结果集获得近似结果信息的查询模型。研究思路是:获取有效的精确的样板数据集 $Q\{a_1, a_2, \dots, a_m, \dots, a_n\}$;分析、挖掘数据集中属性之间的关系,得到关系偏序集 $R\{(a_i \xrightarrow{w} a_j) / a_i, a_j \in Q, w \in [0, 1]\}$;依据关系偏序集 R 和精确的结果集 Q 研究松弛查询规则,得到松弛查询序列集 $q' = \{q_1, q_2, \dots, q_r\}$ 。RIQueryM 可描述为 $(q_r,$

$Q')$, q_r 为查询请求。

(2) 支持不确定数据的近似集成模型(IMInteM)研究。为用户提供集成的数据信息是网格的基本特征。本部分研究支持不确定数据的近似集成模型,主要研究思路:基于典型的不确定数据定义研究不确定数据的相似匹配规则;融合确定数据、不确定数据、精确匹配和近似匹配,研究支持不确定数据的近似匹配模型 $SIM(I_i, I'_i) = SIM_c(I_i, I'_i) * SIM_s(I_i, I'_i)$, 其中, SIM_c 和 SIM_s 分别描述数据信息的完备度和数据间的相似度, I_i 为确定的元组, I'_i 为 I_i 相匹配的元组。进而 IMInteM 描述为 $IMInteM = \{agg(I'_i) | SIM(I_i, I'_i) > \epsilon\}$, $agg(I'_i)$ 为集成的数据, ϵ 为相似阈值。

4.4 自适应的优化调度模型的研究

查询优化是支持网格高效、智能查询处理的基本保证。由于 Web 环境下的深层数据源数量大,具有分布、自治、异构、动态等特点,人们无法预言其假设条件和预知一些信息,这都使查询优化调度成为了研究的重点和难点。本部分针对多目标函数代价模型和动态自适应优化调度模型进行研究。

(1) 多目标函数代价模型(MfCostM)研究。优化器的目的是寻找代价最小的查询执行计划,如: $\min_{e \in E} (e)$ 。 E 为查询执行计划, $c(e)$ 为执行 e 的代价。由于深层 Web 数据库网格支持强大的查询功能,查询范围广,可以说是泛查询,并且网格强调按需获得相关的结果信息,因此,支持网格的代价模型需要在原分布式数据处理的代价模型基础上考虑更多的质量因素,如数据源质量、有效性、费用、安全等因素,以按需支持不同类用户或领域应用的查询需求。为此,研究多目标函数代价模型 $MfCostM = \sum w_i f_i(p_i)$, 其中 $f_i(p_i)$ 为针对参量 p_i 的代价函数, w_i 为 p_i 参数的权重。通用的代价评估目标函数描述为 $S(P, F)$, 其中 P 为参数集合, F 为对应参数的代价函数,以此适应多种参数要求。

(2) 面向局部性的自适应优化调度模型(LASchM)研究。本项目借鉴已有分布式数据库的查询优化策略,重点针对可扩展性和动态协调性进行研究。针对可扩展性,重点研究融合任务并行执行和数据分片并行执行的调度模型;针对动态协调性,重点从协调规则、协调时机以及将协调限制在最小范围内进行研究。避免由于不可预见的事件引起丢失优化的查询执行策略。因此,模型可描述为 $LASchM(E((T, F), R, A)) | \min_{e \in E} MfCostM(e)$, T 为子任务集合, F 为数据片段集合, R 为调度规则定义, A 为自适应协调规则。

结论 本项目的研究将涉及数据库技术、网格技术、WebIR、数据挖掘以及智能推理等理论。本文的研发成果将为构建深层 Web 数据库网格提供良好的支持,为用户提供一个按需获取深层 Web 数据库集成知识的使能环境。就像网格的概念所定义的一样,为用户提供统一的接口,可按需为消费者提供集成的数据知识。深层 Web 数据库网格不仅达到了提升 Web 服务的目的,也可起到推动科技发展、给商家带来更多商机、促进经济繁荣以及推进社会进步的作用。该研究成果具有重要实用价值。

参考文献

- 1 He B. Discovering Complex Matching across Web Query Interfaces: A Correlation Mining Approach. KDD2004, Seattle, Washington, USA, 2004
- 2 He B, Zhang Z. MetaQuerier: Querying Structured Web Sources on-the-fly. SIGMOD 2005, Baltimore, Maryland, USA, 2005
- 3 He B, Chang K C-C. Statistical Schema Matching across Web Query Interfaces. SIGMOD2003, San Diego, California, 2003
(下转第 154 页)

表2 三种不同方法计算结果比较

需求量	路段	H-J	EDO	GA
100	y1	1.25	1.31	1.33
	y2	1.20	1.19	1.22
	y3	0.00	0.06	0.00
	y4	0.95	0.94	0.97
	y5	1.10	1.06	1.09
	目标值	1200.61	1200.64	1200.58
150	分配次数	11	8	5400
	y1	5.95	5.98	6.04
	y2	5.65	5.52	5.47
	y3	0.00	0.02	0.00
	y4	4.60	4.61	4.65
	y5	5.20	5.27	5.26
200	目标值	3156.38	3156.24	3156.21
	分配次数	14	11	7500
	y1	13.00	12.85	13.03
	y2	11.75	12.02	11.68
	y3	0.00	0.02	0.01
	y4	10.25	10.33	10.30
300	y5	11.75	11.77	11.75
	目标值	7086.21	7086.45	7086.15
	分配次数	20	12	4200
	y1	28.44	28.11	28.44
	y2	25.75	26.03	25.74
	y3	0.00	0.01	0.00
300	y4	23.44	23.39	23.44
	y5	26.56	26.58	26.54
	目标值	21209.91	21210.54	21209.91
	分配次数	28	14	7710

(上接第 57 页)

以运行时间作为性能评价的标准。由于本文所给方法和模拟运行的方法的运行时间都与在超周期中所有任务的请求个数相关,因此以任务集中的任务实例个数进行分类。实验中共生成 20 个集合类。集合类 1 包括所有任务实例个数小于等于 10 的集合;集合类 2 包括所有任务实例个数大于 10 且小于等于 100 的集合;集合类 3 包括所有任务实例个数大于 100 且小于等于 200 的集合;…;集合类 20 包括所有任务实例个数大于 1900 且小于等于 2000 的集合。每个集合类中有 20 个任务集。

图 2 与图 3 是模拟运行方法与采用本文提出新方法的运行时间比值分布图,图中比值表示模拟运行方法的运行时间除以新方法的运行时间。从图 2、图 3 中可以看出在任务实例个数的变化过程中,新方法的运行时间始终小于模拟运行方法的运行时间。

小结 本文针对抢占阈值调度策略,给出了具有释放抖动和特定释放偏移的周期任务最大响应时间的计算方法。本文提出的新方法是针对强实时任务的可调度性分析方法的补充和完善,对判断强实时周期任务的可调度性具有重要意义。

参考文献

- 1 宾雪莲,杨玉海,金士尧. 基于 EDF 抢占式调度的周期任务最小

(上接第 125 页)

- 4 He H. Automatic Integration of Web Search Interfaces with WISE-Integrator. WWW2004, New York, NY, 2004
- 5 Wu W S, Yu C. An Interactive Clustering-based Approach to Integrating Source Query Interfaces on the Deep Web. SIGMOD2004, Paris, France, 2004
- 6 Nambiar U, Kambhampati S. Answering Imprecise Queries over Autonomous Web Databases. ICDE, 2006
- 7 Zhang W, Li J Z. Processing Probabilistic Range Query over Imprecise Data Based on Quality of Result. LNCS3842, 2006. 441~449
- 8 Bilke A. Schema Matching using Duplicates. ICDE2005, Tokyo, Japan, 2005
- 9 Wang J Y. Instance-based Schema Matching for Web Databases

结束语 笔者提出的算法具有良好的搜索性能,对于处理有界连续变量的规划问题具有普遍适用性。对于求解连续交通网络设计问题,算法也具有一定实用价值。但从算例可以看出,算法需要较多的交通分配次数,对于大型路网来说,势必消耗大量的 CPU 时间。因此,有必要对算法进行更进一步的改进。可以从两个方面着手解决交通分配的问题,一是寻求更快速的交通分配算法;二是在遗传算法中尽量减少交通分配的次数,可以通过构造一个标准(例如采用类似隐枚举法的标准),仅对满足一定条件的方案进行交通分配,不满足条件的方案的目标函数依照“满足一定条件的方案比不满足条件的方案的目标值好”的原则,参照给出目标值。由于篇幅限制,笔者将另文给出解决办法。

参考文献

- 1 刘灿齐. 现代交通规划学[M]. 北京:人民交通出版社, 2001
- 2 袁亚湘. 非线性规划数值方法[M]. 上海:上海科技出版社, 1993
- 3 Suwansirikul C, Friesz, Tobin R L. Equilibrium Decomposed Optimization: A Heuristic for the Continuous Equilibrium Network Design Problem. Transportation Science, 1987, 21:254~263
- 响应时间分析. 计算机科学, 2004, 31(9):114~116
- 2 Redell O, Törmgren M. Calculating exact worst-case response times for static priority scheduled tasks with offsets and jitter. In: Proceedings of the 8th real-time and embedded technology and applications symposium. RTAS'02, San Jose, CA, USA. IEEE computer society press, 2002. 164~172
- 3 Bate I, Burns A. Schedulability analysis of fixed priority real-time systems with offsets. PEUROMICRO-RTS'97. IEEE computer society press, 97
- 4 Redell O. Response time analysis for implementation of distributed control systems; [Phd thesis]. Royal institute of technology, KTH, Sweden, 2003
- 5 Goosesens J, Devillers R. The non-optimality of the monotonic priority assignments for hard real-time offset free systems. Real-time System, 1997, 13(2):107~126
- 6 杨玉海,宾雪莲,金士尧. 基于抢占阈值调度的周期任务最小响应时间分析. 计算机应用研究, 2004, 21(11):41~44
- 7 宾雪莲,金士尧,杨玉海. 周期多帧任务模型的响应时间分析. 计算机工程与科学, 2003, 25(6):104~108
- 8 宾雪莲. 实时任务调度分析: [博士论文]. 长沙:国防科技大学研究生院, 2004
- 9 宾雪莲,杨玉海,金士尧. 一种基于分组与适当选取策略的实时多处理器的动态调度算法. 计算机学报, 2006, 29(1):81~91

by Domain-specific Query Probing. VLDB2004, Toronto, Canada, 2004

- 10 He B. Statistical Schema Matching across Web Query Interfaces. SIGMOD2003, San Diego, California, 2003
- 11 Anjomshoaa A. The Design and Implementation of Grid Database Services in OGSA-DAI. In: Proceedings of UK e-Science All Hands Meeting, Nottingham, EPSRC, 2003
- 12 Paques H, Liu L, Pu C. Distributed Query Adaptation and Its Trade-offs. SAC 2003, Melbourne, Florida, USA, 2003
- 13 Urhan T. Dynamic Pipeline Scheduling for Improving Interactive Query Performance. VLDB2001, Rome, Italy, 2001
- 14 Naughton J, DeWitt D, Maier D. The Niagara Internet Query System. VLDB2000, Cairo, Egypt, 2000
- 15 Ives Z. Adaptive Query Processing for Internet Applications. IEEE Data Engineering Bulletin, 2000, 23(2):19~26