**参考文献**

【1】相似重复记录清理方法研究综述。叶焕倬 吴 迪 (中南财经政法大学信息与安全工程学院信息系 武汉 430073) ：56-57

【2】Galhardas H, F lorescu D, Sh asha D, et a.l Declarative Data C lean ing: Language , M odel and A lgorithm s[ C ]. In: P roceedings of the 27th International Conference on Very Large Data Bases, R om e, Ita ly. 2001: 371- 380

【3】B itton D, D e W itt D J . Duplicate Record E li m in ation in Large Data Files[ J]. ACM Transactions on Database System s, 1983, 8 ( 2 ): 255- 265.

【4】H ernandezM, Stolfo S. TheMerge /Pu rge Prob lem for LargeDatabases [ C]. In: Proceedings of theACM SIGMOD InternationalConference on Management of Data, S an Jose, California . 1995: 127- 138.

【5】H ernandezM A, Stolfo S J. Real- W orld Data is D irty : Data C leansing and th e M erge /Purge Prob lem [ J]. D ata M ining and K now ledge D iscovery, 1998, 2( 1): 9- 37.

【6】梁雪，任剑锋，景丽． 基于 ＱＰＳＯＬＳＳＶＭ 的数据库相似重复记录检测算法［Ｊ］． 计算机科学，２０１２，３９（１ １ ）：１５７ １５９．

【7】马翔． 粒子群优化 ＢＰ 神经网络用于重复记录检测［ Ｊ］． 辽宁工程技术大学学报：自然科学版，２０１０，２９（５）：９５９ ９６３．

【8】BiRon D，De Witt D J．Duplicate Record Elimination in Large Data Files [J]. ACM

Transactions on Database Systems，1 983，8(2)：255-265．

【9】Elma garmid A K，Ipeirotis P G, Verykios V S．Duplicate Record Detection：A Survey [J]. IEEE Transactions on Knowledge and Data Engineering，2007，19(1)：1-16．

【10】Jin L，Li C，Mehrotra S．Efficient record linkage in large data sets[C]．Eighth InternationalConference on DatabaseSystems forAdvancedApplications．Kyoto，2003：137-148．

【11】Wangy Y R，Madnick S E．The inter-database instance identification problem in integrating autonomous systems[A]．In：Wang Y R, Madrtick S E，eds．Proceedings of the Fifth International Conference on Data Engineering[C]．IEEE Computer Societh，1989：46-55．

【12】Galhardas H, Florescu D. An Extensible Framework for Data Cleaning[C]. In: Proceedings of the 16th IEEE International Conferenceon Data Engineering. San Diego, California, 2000: 312.

【13】陈伟 数据清理关键技术及其软件平台的研究与应用 [D]. 南京 :南京航空航天大学 ,2004:15-24.

【14】Flanagan T.,Safdie E. Apractical Guide to Achieving Enterprise Data Quality [DＢ /OL] .http://www.techguide.com,1999.

【15】Monge AE, Elkan C. An efficient domain-independent algorithm for detecting approximately

duplicate database records[C].In:Workshop on Research Issues on Data Mining and Knowledge

Discovery,1997:211-217.

【16】C. L. Forgy. 0PS5 User’s Manual. Technical Report CMU-CS-81-135, Carnegie Mellon University, July 1981.

【17】Newcombe HB,Kennedy JM,Axford SJ,etal.Automatic linkage of vital records[J].

Science,1959(130):954-959.

【18】Qiu Y F,Tian Z P,Ji W Y,et al.An efficient approach for detecting approximately duplicate

database records[J].Chinese J of Computers, 2001 (1):69-77.

【19】Gianni Costa, Giuseppe Manco, Riccardo Ortale.An incremental clustering scheme for data

de-duplication [C]. Data Mining and Knowledge Discovery, 2010(20):152-187.

【20】Alfredo Ferro, Rosalba Giugno, Piera Laura Puglisi, Alfredo Pulvirenti.An efficient duplicate

record detection using q-grams array inverted index [C].In: 2010 Proceedings of the 12th

international conference on Data warehousing and knowledge discovery,2010:309-323.

【21】郭志懋 周傲英

.数据质量和数据清洗研究综述

[J].软件学报

,2002 (11):2076-2082.

【22】邱越峰

,

田增平等

.

一种高效的检测相似重复记录的方法

[J].

计算机学报

,2001 (1):69-77.

【23】韩京宇

,

徐立臻

,

董逸生

.

一种大数据量的相似记录检测方法

[J].

计算机研究与发展，

2005,

(12):2206-2212.

【24】陈 伟

,

王 昊

,

朱 文 明

.

一 种 提 高 相 似 重 复 记 录 检 测 精 度 的 方 法

[J].

计 算 机 应 用 与 软

件

,2006,23(10):29-30.

【25】李 星 毅

,

包 从 剑

,

施 化 吉

.

数 据 仓 库 中 的 相 似 重 复 记 录 检 测 方 法

[J].

电 子 科 技 大 学 学

报

,2007,36(6):1273-1277.

【26】庞雄文

,

姚占林

,

李拥军

.

大数据量的高效重复记录检测方法

[J].

华中科技大学学报

(

自然科学

版

),2010,2(38):8-11.

【27】周丽娟

,

肖满生

.

基于数据分组匹配的相似重复记录检测

[J].

计算机工程

,2010:104-106.

【28】Raman, V., Hellerstein,J. Potter’s wheel:an interactive data cleaning system. Peoceedings of the 27th International Conference on Very Large Data Bases. Roma:Morgan Kaufmann, 2001:381-390.

【29】Lee, M.L., Ling, T.W., Low, W. L. IntelliClean: a knowledge-based intelligent data cleaner. Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Boston:ACM Press, 2000:290:294.