



wdilab

Objekt-Matching, Dublettenbehandlung

Motivation

Unternehmen führen häufig Daten aus verschiedenen Datenquellen zusammen, z.B. in Business-Intelligence-Projekten, bei Unternehmensfusionen oder Kundentransaktionen.

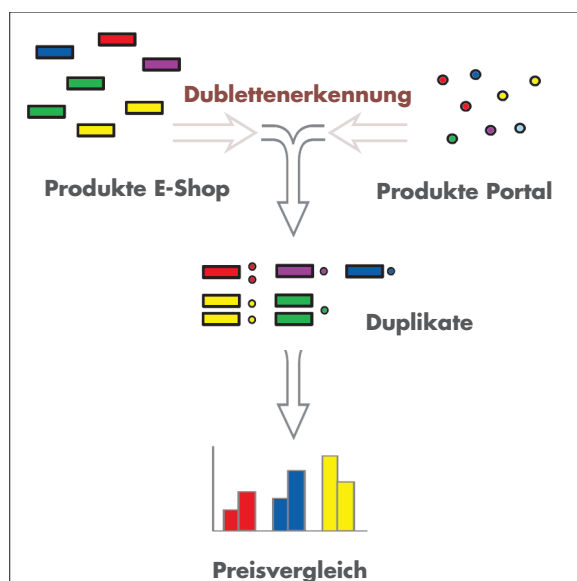
- Erhöhung der Datenqualität
- Aktualisierung von Datenbeständen und Datenanalyse

Anwendung

Integration und Aufbereitung unternehmensinterner bzw. unternehmensexterner Daten (Webdaten) z.B. zur

- Bereinigung der Kundenbestände
- Erstellung von Konkurrenzanalysen (Produkt, Preis, Anbieter, Zielgruppen)

Anwendungsbeispiel Preisvergleich



Ziele

- Kombination verschiedener Verfahren zur Sicherung einer sehr guten Ergebnisqualität
- Wiederverwendung von Ergebnissen zur Performance- und Qualitätssteigerung
- Automatisches Tuning

Innovation

- Flexible Rahmenarchitektur zur optimierten und (semi-) automatischen Dublettenerkennung
- Generischer Ansatz—erreicht qualitativ hochwertige Ergebnisse

Prototyp Fever: Screenshot der GUI

name	ε(M)	O(M)
Canon Cyan Photo Ink Cartridge - Cyan - CLI8PC	0.945	1
Canon CLI-8PC Photo Cyan Ink Cartridge - 0624B002	0.945	1
Canon Magenta Photo Ink Cartridge - Magenta - CLI8PM	0.945	1
Canon CLI-8PM Photo Magenta Ink Cartridge - 0625B002	0.945	1
Sony 2GB Memory Stick Micro (M2) - MSA2GU2	0.945	1
Sony 2GB Memory Stick Micro (M2) - MSA2GD	0.945	1
SIRIUS STILETTO 2 Portable Satellite Radio - SL2PK1	0.945	1
SIRIUS SL2PK1 SIRIUS Stiletto 2 Portable Radio	0.945	1
Sony 1GB Memory Stick PRO Duo Mark 2 Media Card - MSMT1G	0.945	1
Sony 1GB Memory Stick PRO Duo Card - MSMT1G	0.945	1
Sony 2GB Memory Stick PRO Duo Mark 2 Media Card - MSMT2G	0.945	1
Sony 2GB Memory Stick PRO Duo Card - MSMT2G	0.945	1
Sony 4GB Memory Stick PRO Duo Mark 2 Media Card - MSMT4G	0.945	1
Sony 4GB Memory Stick PRO Duo Card - MSMT4G	0.945	1
Sony 8GB Memory Stick PRO Duo Mark 2 Media Card - MSMT8G	0.945	1
Sony 8GB Memory Stick PRO Duo Card - MSMT8G	0.945	1
Sony Digital SLR Camera With Lens Kit - DSLRA200W	0.945	1
Sony alpha DSLR-A200 Digital SLR Camera with Dual Lens Kit -	0.945	1
Sony Xplod 10-Disc Add-On CDMP3 Changer - CDX565MRF	0.945	1
Sony CDX-565MRF 10-Disc CDMP3 Changer	0.945	1
Linksys EtherFast 4124 24-Port Ethernet Switch - EF4124	0.945	1
Linksys EtherFast EF4124 Ethernet Switch	0.945	1
Sony Learning Remote Control - Silver Finish - RMLV600	0.945	1
Sony Learning Remote Control - RMLV600	0.945	1
Bose SL2 Wireless Black Surround Link - SL2WIRELESS	0.945	1
Bose SL2 Wireless Surround Link	0.945	1
Kenwood KCA-iP200 Black iPod Direct Cable - KCAIP200	0.945	1
Kenwood iPod Direct Cable - kcap200	0.945	1
Sony HD Radio Add On Tuner - XT100HD	0.945	1
Sony XT-100HD HD Radio Tuner	0.945	1
Panasonic PT-AK200U White Home Theater Projector - PTAK200	0.945	1
Panasonic PT-AK200U Home Theater Projector	0.945	1
Panasonic Black 8.5" Portable DVD Player - DVDLS86	0.945	1
Panasonic DVD-LS86 Portable DVD Player	0.945	1
Panasonic Black 8.5" Portable DVD Player - DVDLS83	0.945	1
Panasonic DVD-LS83 Portable DVD Player	0.945	1

Informationen zur Technologie

Thor, A.; Kirsten, T.; Rahm, E.: Instance-based matching of hierarchical ontologies. Proc. of 12. GIFachtagung für Datenbanksysteme in Business, Technologie und Web, 2007.

Köpcke, H.; Rahm, E.: Training Selection for Tuning Entity Matching. 6th International Workshop on Quality in Databases and Management of Uncertain Data (QDB/MUD), 2008.

Thor, A.; Rahm, E.: MOMA - A Mapping-based Object Matching System. Proc. Conf. Innovative Data Research (CIDR), Asilomar, 2007.

Weitere Informationen unter

<http://wdilab.uni-leipzig.de>