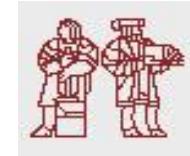




The MIT Information Quality Industry Symposium, 2007



Applying search technology to data matching, consolidation, and cleansing

Jeff Fried

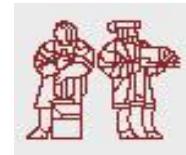
VP Advanced Solutions

Jeff.Fried@fastsearch.com





The MIT Information Quality Industry Symposium, 2007



All Data is Dirty

Forrt Lauderdale

Fort ALuderdale

Fort Laderdale

Fort Laudedale

Fort Lauderdale

Fort Launderdal

Fort Lauderdale

FORT LAUDERDALE FLA.

Fort LAuderdale,

FtLauderdale

Fort Lauderdaleq

Fort Lauderrdale

Fort Laudewrdale

FORT LUADERDALE

Fort. Lauderdale

Fort.Lauderdale

Ft Lauderdale

Ft. Lauderdale

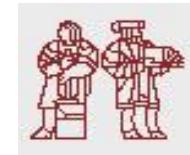
FT. LAUDERDALE

FT.LAUDERDALE

The logo features three orange dots above the word 'fast' in a bold, italicized, orange sans-serif font.



The MIT Information Quality Industry Symposium, 2007



Search technology emphasizes connecting to the User

FAST Connects Users To Information, Products and Communities

USERS

FrontOffice: Search IS the Portal

Search Connects: Service Delivery Framework

BackOffice: Adaptive Information Warehouse

Information

Products

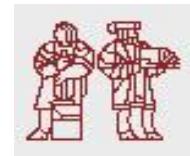
Communities



XML



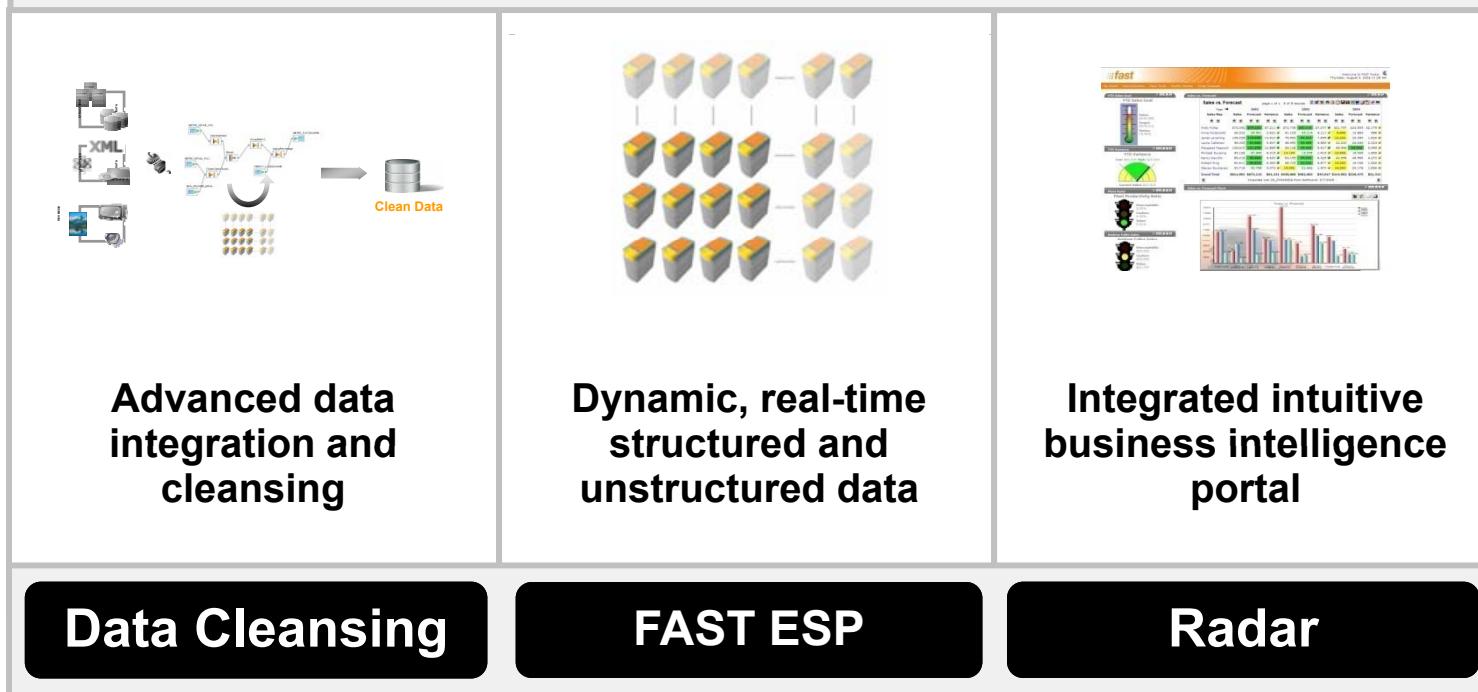
fast



The Adaptive Information Warehouse

**The first business intelligence solution built on search.
Redefining how organizations access, interact with and exploit
information.**

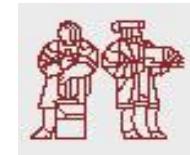
The Adaptive Information Warehouse (AIW)



fast

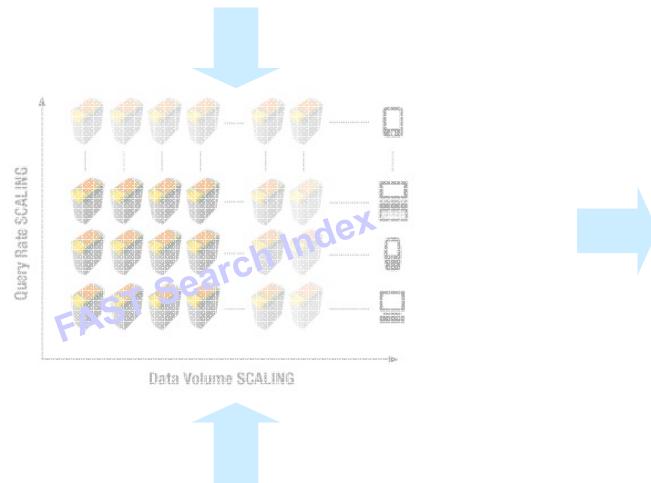


The MIT Information Quality Industry Symposium, 2007



| Customer ID | Last Name | First Name | Address | City | State | Zip |
|-------------|-----------|------------|--------------------------|----------------|-------|-------|
| 12093-28937 | Williams | Joshua | 41 Smiley Ave | Haverhill | MA | 01830 |
| 23801-12938 | Simpson | Mark | 194 No 8 Rd | New Wilmington | PA | 16142 |
| 28934-83112 | Karel | Adrienne | 2435 Park Ave. Southeast | Milpitas | CA | 95035 |
| 38472-20937 | Meyers | John | 10932 Wellington St. | Albuquerque | NM | 87106 |
| 42348-00198 | Welch | Jack | 1338 Ruswood Dr | Abilene | TX | 79601 |
| 72831-21230 | Thatcher | Margaret | 9331 E Cactus Ln S | Sun Lakes | AZ | 85248 |
| 92837-43224 | Johnson | Alice | 872 N. Campbell Ave. | Springfield | MO | 65802 |

Customer Records from Corporate CRM Database



Match,
merge,
cleanse.

| Last Name | First Name | Address | City | State | Zip | Phone No. |
|-----------|------------|-----------------------|-------------|-------|-------|--------------|
| Carroll | Adrian | 2435 Park Ave. S East | Milpitas | CA | 95035 | 408-130-2897 |
| Davidson | Henry | 2997 Shore Dr | Merrick | NY | 11566 | 212-287-8821 |
| Johnson | Alice | 872 N. Campbell Ave. | Springfield | MO | 65802 | 417-862-3465 |
| Meiers | John | 10932 Willinton | Albuquerque | NM | | 505-792-0323 |
| Phillips | Natalie | 4916 N Hoyne Ave | Chicago | IL | 60625 | 773-315-2872 |

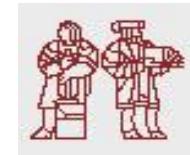
Regional Sales Call Center /

Customer Service Database

fast



The MIT Information Quality Industry Symposium, 2007



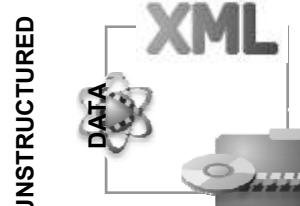
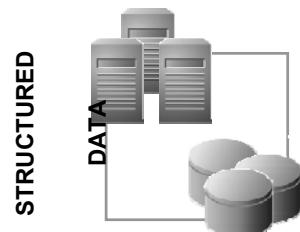
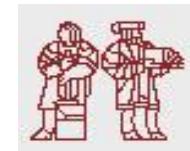
Data matching, cleansing, and consolidation

- Biggest problem for all data warehouses
 - Merge of multiple sources of data
 - No common key
- FAST solution: unique combination of traditional DWH logic (ETL) and fuzzy matching
 - Full ETL tool for the traditional ETL logic
 - Use search engine core for fuzzy match
 - Apply linguistic techniques to data quality issues





The MIT Information Quality Industry Symposium, 2007



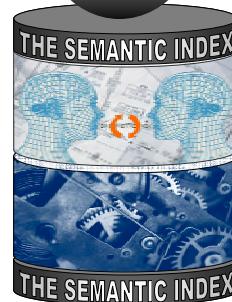
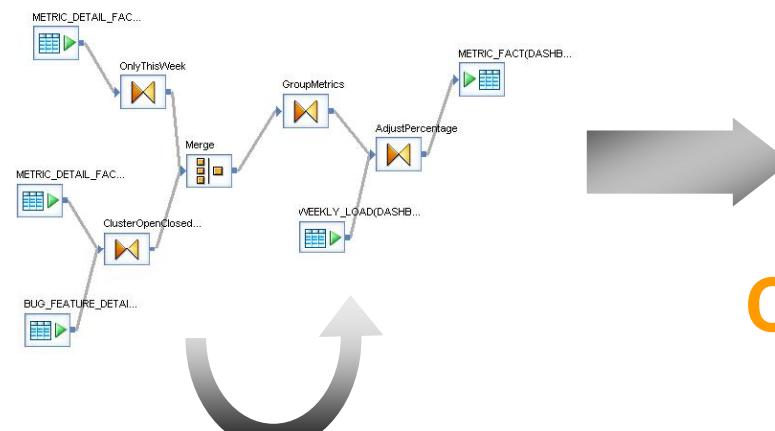
► Content Fusion

fast

Cleansing Example:

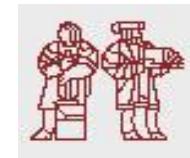
Schibsted built Internet Yellow Pages and White Pages from scratch in 8 weeks outperforming existing providers on data quality

98 % Cleansing Precision

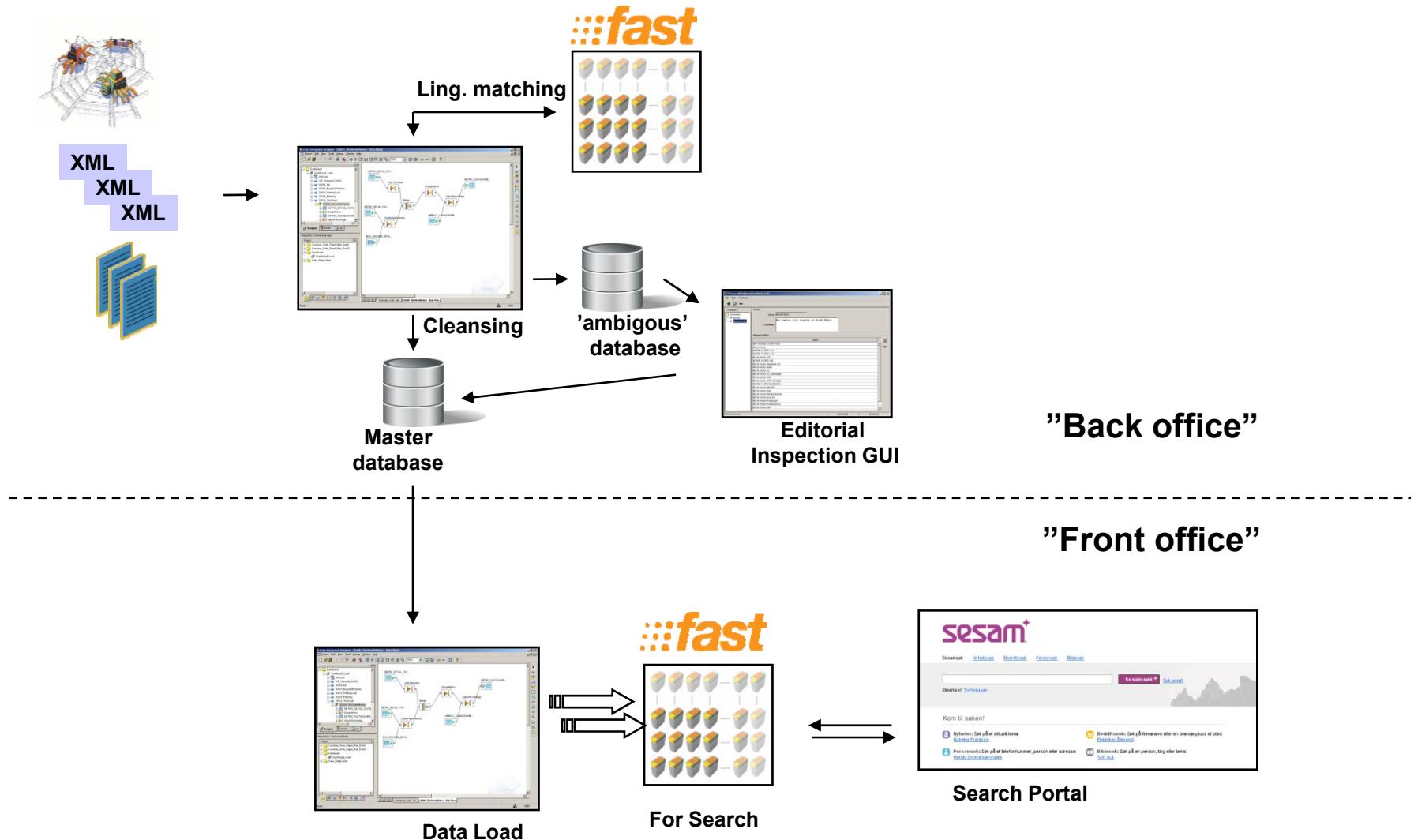


Clean Data

► **Semantic and Phonetic Cleansing Analysis**



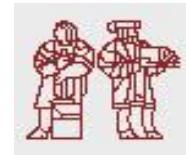
Example system configuration



fast



Data Warehouse of the Year Award 2006



sesam*

Find the business who called our Recruitment Officer!

Sesam creates composite objects by joining and cleansing data

Bedriftssøk: 1 treff på 92897133

Boyden As

Resultatregnskap

| | 2005 | 2004 | 2003 |
|---------------------|-------|-------|-------|
| Sum driftsinntekter | 3.249 | 3.492 | 2.895 |
| Driftsresultat | 2.446 | 1.431 | -47 |
| Arenasresultat | 2.446 | 1.456 | 44 |

Three Years of Financial Data showing growth and profits

Blended Content: A blog on the founder of Boyden International

B Bloggsøk Beta: 1 blogg ga treff på boyden

Tips oss om bloggen din: Legg til blogg

Skeptics alert: Astrology and Leadership

...A friend of my father's, Sidney Boyden, told me he was born in 1942 and I proved it statistically. Sidney Boyden was founder of Boyden & Associates, a global executive search firm.

The Chairman is born in 1942, lives in Vinterbro and runs a competing firm as well ☺

Firma: Boyden global executive search

Kontaktperson: Bjørn Boberg

E-post: bjorn.boberg@boyden.no

Hjemmeside: www.boyden.com

Kontaktinformasjon: bjorn.boberg@boyden.no / tlf. 92897133

Registrert: 1.01.2005

The CEO, Bjørn Boberg called us. The cellphone number is registered to him.

Personssøk:

Kontaktinfo

Navn: Jan Gunnar Schumann

Adresse: Kollenveien 14, 0171 Oslo

Telefon: 64 94 53 71

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

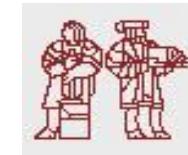
Stensberggata 17 B, 0170 Oslo

Sesam Bedriftssøk er beriket av innhold fra internettkatalogen.no - en ren bransjekatalog. [Les mer på internettkatalogen.no](#)

fast

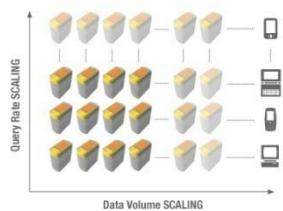


Data Cleansing Scenario One: "Master" DB exists



One 'Master' exists, other datasources are cleansed towards it

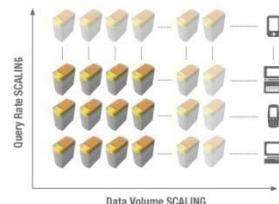
| | id | lastname | firstname | address |
|----|-------------|--------------|-----------|----------------------|
| 1 | 172-32-1176 | White | Johnson | 10932 Bigge Rd. |
| 2 | 213-46-8915 | Green | Marjorie | 309 43rd St. #4 |
| 3 | 238-95-7766 | Carson | Cheryl | 589 Darwin Ln. |
| 4 | 267-41-2394 | O'Leary | Michael | 22 Cleveland Av. #14 |
| 5 | 274-80-9391 | Straight | Dean | 5420 College Av. |
| 6 | 341-22-1782 | Smith | Meander | 10 Mississippi Dr. |
| 7 | 409-56-7008 | Bennet | Abraham | 6223 Bateman St. |
| 8 | 427-17-2319 | Dull | Ann | 3410 Blonde St. |
| 9 | 472-27-2349 | Grovesby | Burt | PO Box 782 |
| 10 | 486-29-1786 | Lockley | Charlene | 18 Brasilia J. |
| 11 | 527-72-3088 | Greene | Muninger | 22 Grateful H. |
| 12 | 535-57-5118 | Holloway | Helen | 10 Mississippi Dr. |
| 13 | 672-71-3249 | Yokonami | Atsuko | 3 Silver Ct. |
| 14 | 712-45-1987 | del Castillo | Innes | 2286 Cray Pl. |
| 15 | 722-51-5454 | DeFrance | Michel | 3 Baling Pl. |
| 16 | 731-63-1234 | Ward | John | 44 Upstart Rd. |
| 17 | 749-80-9391 | MacFester | Eugene | 4709 Meridian |
| 18 | 764-95-7991 | Ward | John | 4709 Meridian |



Step one:

- The Customer Master is indexed by FAST ESP

| | id | lastname | firstname | address | city | state | zip |
|---|-------------|----------|-----------|----------------------|------------|-------|-------|
| 1 | 172-32-1176 | White | Johnson | 10932 Bigge Rd. | Merlo Park | CA | 94025 |
| 2 | 213-46-8915 | Green | Marjorie | 309 43rd St. #411 | Oakland | CA | 94618 |
| 3 | 238-95-7766 | Carson | Cheryl | 589 Darwin Ln. | Berkley | CA | 94705 |
| 4 | 267-41-2394 | O'Leary | Michael | 22 Cleveland Av. #14 | San Jose | CA | 95128 |
| 5 | 274-80-9391 | Straight | Dean | 5420 College Av. | Oakland | CA | 94609 |
| 6 | 341-22-1782 | Smith | Meander | 10 Mississippi Dr. | Lawrence | KS | 66044 |
| 7 | 409-56-7008 | Bennet | Abraham | 6223 Bateman St. | Berkeley | CA | 94705 |
| 8 | 427-17-2319 | Dull | Ann | 3410 Blonde St. | Palo Alto | CA | 94301 |
| 9 | 472-27-2349 | Grovesby | Burt | PO Box 782 | Palo Alto | CA | 94301 |



Step two:

A datasource is cleansed according to the Master

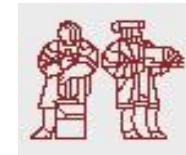
| | id | lastname | firstname | address | city | state | zip |
|---|-------------|----------|-----------|----------------------|------------|-------|-------|
| 1 | 172-32-1176 | White | Johnson | 10932 Bigge Rd. | Merlo Park | CA | 94025 |
| 2 | 213-46-8915 | Green | Marjorie | 309 43rd St. #411 | Oakland | CA | 94618 |
| 3 | 238-95-7766 | Carson | Cheryl | 589 Darwin Ln. | Berkley | CA | 94705 |
| 4 | 267-41-2394 | O'Leary | Michael | 22 Cleveland Av. #14 | San Jose | CA | 95128 |
| 5 | 274-80-9391 | Straight | Dean | 5420 College Av. | Oakland | CA | 94609 |
| 6 | 341-22-1782 | Smith | Meander | 10 Mississippi Dr. | Lawrence | KS | 66044 |
| 7 | 409-56-7008 | Bennet | Abraham | 6223 Bateman St. | Berkeley | CA | 94705 |
| 8 | 427-17-2319 | Dull | Agn | 3410 Blonde St. | Palo Alto | CA | 94301 |
| 9 | 472-27-2349 | Grovesby | Burt | PO Box 782 | Palo Alto | CA | 94301 |

Result: a corrected version of that datasource

fast



Data Cleansing Scenario Two: No "Master" DB



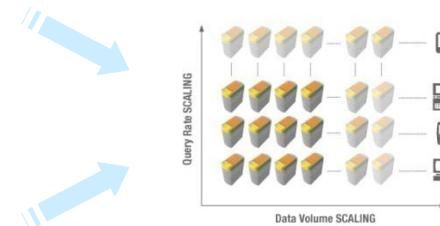
Two or more databases exist, and a merge is needed

Incident records from Central Region

| | id | lastname | firstname | address | city | state | zip |
|---|-------------|-----------------|------------------|----------------------|-------------|--------------|------------|
| 1 | 172-32-1176 | White | Johnson | 10932 Bigge Rd. | Menlo Park | CA | 94025 |
| 2 | 213-46-8915 | Green | Marjorie | 309 63rd St. #411 | Oakland | CA | 94618 |
| 3 | 238-95-7766 | Carson | Cheryl | 589 Darwin Ln. | Berkeley | CA | 94705 |
| 4 | 267-41-2394 | O'Leary | Michael | 22 Cleveland Av. #14 | San Jose | CA | 95128 |
| 5 | 274-80-9391 | Straight | Dean | 5420 College Av. | Oakland | CA | 94609 |
| 6 | 341-22-1782 | Smith | Meander | 10 Mississippi Dr. | Lawrence | KS | 66044 |
| 7 | 409-56-7008 | Bennet | Abraham | 6223 Bateman St. | Berkeley | CA | 94705 |
| 8 | 427-17-2319 | Dull | Ann | 3410 Blonde St. | Palo Alto | CA | 94301 |
| 9 | 472-27-2249 | Gringlesby | Burt | PO Box 792 | Coweln | CA | 95428 |

Incidents from East Region

| | id | lastname | firstname | address | city | state |
|----|-------------|-----------------|------------------|--------------------|----------------|--------------|
| 1 | 807-91-6654 | Pantley | Sylvia | 1956 Arlington Pl. | Rockville | MD |
| 2 | 527-72-2246 | Greene | Morningstar | 22 Graybar Hou... | Nashville | TN |
| 3 | 722-51-5454 | DeFrance | Michel | 3 Balding Pl. | Gary | IN |
| 4 | 712-45-1887 | del Castillo | Innes | 2086 Cram Pl. # | Ann Arbor | MI |
| 5 | 341-22-1782 | Smith | Meander | 10 Mississippi Dr. | Lawrence | KS |
| 6 | 859-46-2035 | Ringe | Annie | 67 Seventh Av. | Salt Lake City | UT |
| 7 | 950-72-3567 | Ringe | Albert | 67 Seventh Av. | Salt Lake City | UT |
| 8 | 172-32-1176 | White | Johnson | 10932 Bigge Rd. | Menlo Park | CA |
| 9 | 486-29-1786 | Locksley | Charlene | 18 Broadway Av. | San Francisco | CA |
| 10 | 427-17-2319 | Dull | Ann | 3410 Blonde St. | Palo Alto | CA |
| 11 | 846-50-7186 | Hunter | Sheryl | 3410 Blonde St. | Palo Alto | CA |
| 12 | 672-71-3249 | Yokomoto | Akiko | 3 Silver Ct. | Walnut Creek | CA |
| 13 | 752-20-7291 | Stringer | Birk | 5201 Talmank | Oakland | CA |



Cleansed Unioned Incident list

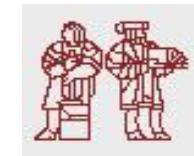
| | id | lastname | firstname | address |
|----|-------------|-----------------|------------------|-----------------|
| 1 | 172-32-1176 | White | Johnson | 10932 Bigge R |
| 2 | 213-46-8915 | Green | Marjorie | 309 63rd St. # |
| 3 | 238-95-7766 | Carson | Cheryl | 589 Darwin Ln |
| 4 | 267-41-2394 | O'Leary | Michael | 22 Cleveland / |
| 5 | 274-80-9391 | Straight | Dean | 5420 College / |
| 6 | 341-22-1782 | Smith | Meander | 10 Mississippi |
| 7 | 409-56-7008 | Bennet | Abraham | 6223 Bateman |
| 8 | 427-17-2319 | Dull | Ann | 3410 Blonde S |
| 9 | 472-27-2249 | Gringlesby | Burt | PO Box 792 |
| 10 | 486-29-1786 | Locksley | Charlene | 18 Broadway / |
| 11 | 527-72-3246 | Greene | Morningstar | 22 Graybar Hc |
| 12 | 648-92-1872 | Blotchet-Halls | Reginald | 55 Hillsdale Bl |
| 13 | 672-71-3249 | Yokomoto | Akiko | 3 Silver Ct. |
| 14 | 712-45-1867 | del Castillo | Innes | 2286 Cram Pl. |
| 15 | 722-51-5454 | DeFrance | Michel | 3 Balding Pl. |
| 16 | 724-08-9931 | Stringer | Dirk | 5420 Telegrap |
| 17 | 724-80-9391 | MacFeather | Stearns | 44 Upland Hts |
| 18 | 752-20-7291 | Korean | Livis | 5720 McAlister |

- Duplication identification and removal
- Approximate matching to help oversee minor errors
- Multi-field compare to ensure secure identification
- Tunable relative importance of each field
- More than 20 million names known by the platform
- Spell checking and error correction included
- Private attributes from either source are merged to a super-set





Search Is a Matching Service



3-Gram

Score

ABUKABLAN, KHALID HUSSEIN OCEAN BLVD 33062 POMPANO BEACH

131 776

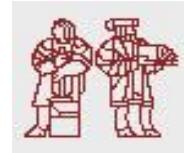
3-Gram

Search improves Data Quality through de-duplication

fast



Linguistic Capabilities Provide High Data Matching Accuracy

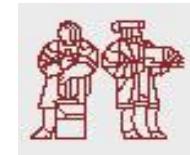


- **Score strings on similarity**
 - **Lai Nancy, 501 N. Clinton St. ≈ Nancy J. Lai, 501 North Clinton**
- **Levenshtein edit distance algorithms**
 - **Calculates the difference on two strings (names)**
 - **John Meyers ≈ John P. Meiers; Britney Spears ≈ Britney Spaers**
- **Phonetic match**
 - **Calculates an integer from the strings, and matches on 'sounds like'**
 - **Christin ≈ Kristin; Kierkegaard ≈ Kierkegård**
- **Scope Fields (new in FAST Instream 4.1 and FAST ESP 5.0)**
 - **Finds relevancy in a scope or a context – for example 'Person' vs. 'Address'.**
- **'Charset' normalization**
 - **For multinational chars and name combinations**
 - **André = Andre, Jürgen = Jurgen etc.**





The MIT Information Quality Industry Symposium, 2007



Linguistic Fundamentals

Lexicon Base

Special terminology lexica

Geographical and people's names

Spellcheck dictionaries

Subject-specific ontologies

Synonymy Dictionaries

Part-of-speech Dictionaries

Inflection Dictionaries

Language-specific Common Words

Basic Linguistic Algorithms

Vectorization

Part-of-speech Tagging

Language normalization

Pattern extraction

Stemming / Lemmatization



Entity Extraction

Categorization

Include all forms (10 languages)

Stop word elimination

Spell checking

Lang ID (77 languages)

Find similar

Suggest Synonyms

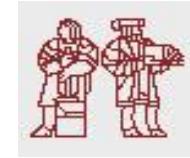
Drill-down & Refinement

Applications

fast

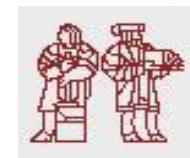


Where do Linguistics and other search technologies apply to data quality?



- On textual and/or multimedia data
- In combination with other rules-based techniques
- In hard-to-match situations
 - Match and cleanse with significantly fewer rules
 - Match and cleanse with much higher accuracy
- Where structuring “blobs” of text is useful
 - Entity Extraction
 - Relationship Extraction

fast



1 Data Consolidation and ETL Studio

Direct access to RDBMs
for info from some Telco's



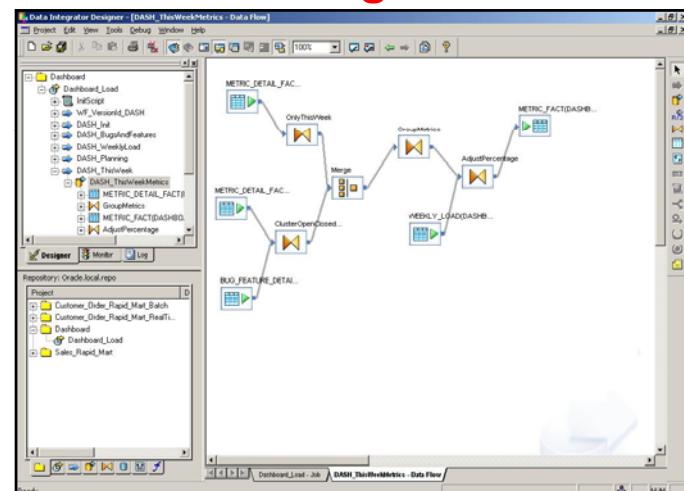
XML feed from
other Telco's



Flat files (CSV or fixed)
from the 'laggards'



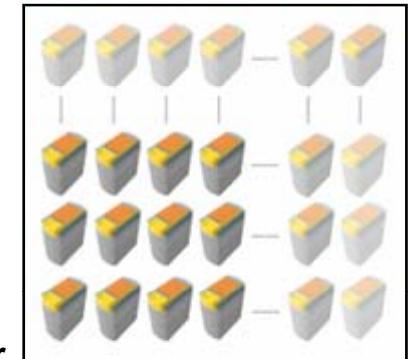
Logic for ETL



Logic for matching
and cleansing

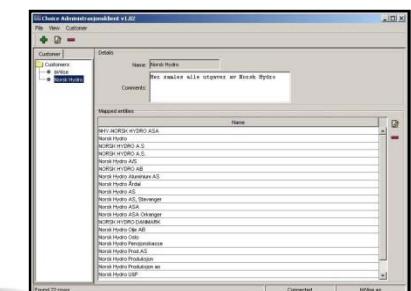
Lookup to ESP for
Matching

Ordered hits (by quality)



Cleansed data
Back to ESP for
next round
Cleansing
or for 'search'

Ambiguous
data
(close hits or
unidentified)



'Error' database for manual inspection,
correction, storage/learning

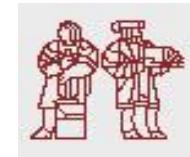


Master database for
persistant storage

fast



The MIT Information Quality Industry Symposium, 2007

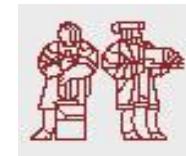


Key Cleansing Benefits

- Increased accuracy in search
- Improved entity recognition and relationship mapping
- Better analytics and intelligence
- Improvement and reuse of data for all applications



Cleansing: Impact on Discovery



mozart Search

Artist (100)
 mozart (2810)
 wolfgang amadeus MOZART (1980)
 Mozart, Wolfgang amadeus (370)
 W. A. MOZART (175)
 Mozart (Complete Works) (152)
 w.a. mozart (139)
 w.a.mozart (47)
 WOLFGANG AMADEUS MOZART (1756-1791) (39)
 MOZART WOLFGANG AMADEUS (37)
 Mozart, Wolfgang Amadeus (1756-1791) (34)
 Mozart, W.A. (28)
 Mozart, W. A. (27)
 London Mozart Players (17)
 wolfgang amadeus mozart (17)
 WA Mozart (16)
 Wolfgang A. Mozart (16)
 Wolfgang Amadeus Mozart (12)
 MOZART FESTIVAL ORCHESTRA (12)
 Wolfgang Mozart (11)
 W A Mozart (10)
80 more...

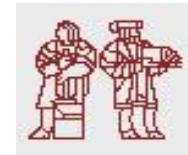
Artist (6)
 Wolfgang Amadeus Mozart (1998)
 Mozart Festival Orchestra (10)
 Leopold Mozart (7)
 Mozart & Haydn (4)
 Mozart - Beethoven (4)
 Beethoven Mozart (3)

fast



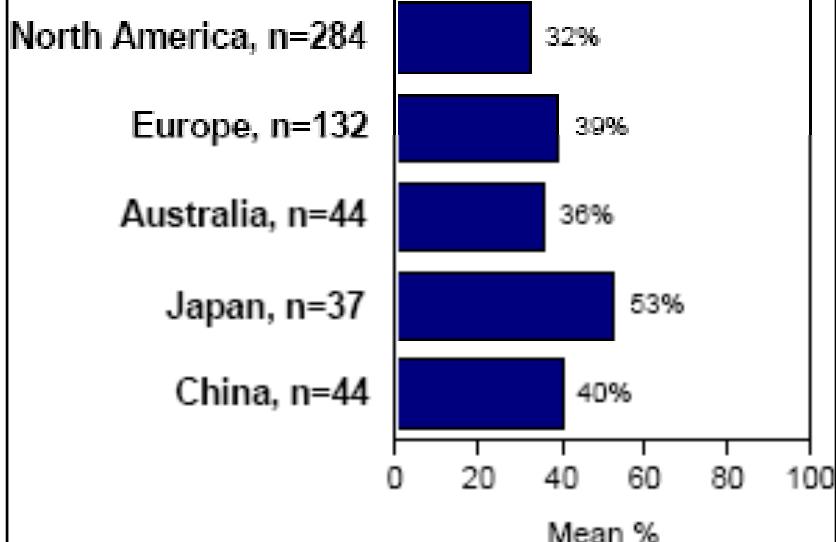
The Largest BI Limitations Today

007



*If We Build It,
<50% of Them Will Come*

“Is everyone using BI who should be using BI?”

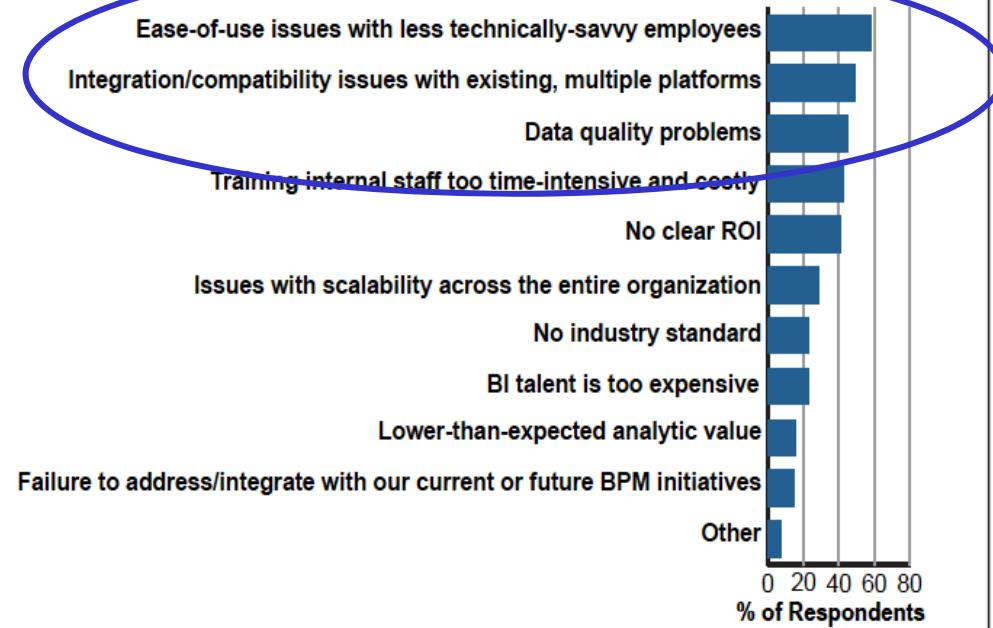


Source: Gartner

What else is wrong with BI today?

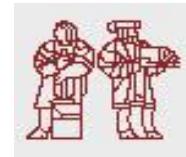
Deployment Challenges

What are the barriers to adopting BI tools enterprise-wide?





BI Built on Search Redefines How Organizations Access, Interact With and Exploit Business Intelligence



Drivers of Pervasive BI:



- Better business decisions with accurate, enriched data
- Increased productivity with ad-hoc, real-time performance
- Increased adoption with easy and intuitive access to intelligence

ease of use

democratize information

perform deeper analytics

currently just a handful of power users

high quality data

real-time information

break down silos

timely cross-enterprise information

integrate structured and unstructured data

speed-up decision making

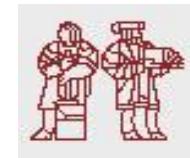
suppress noise in data

fast

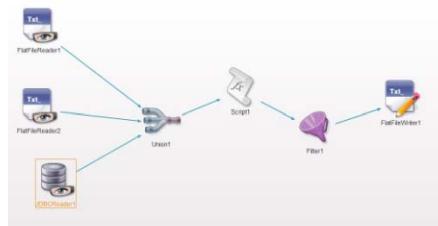




BI Built on Search Redefines How Organizations Access, Interact with and Exploit Information

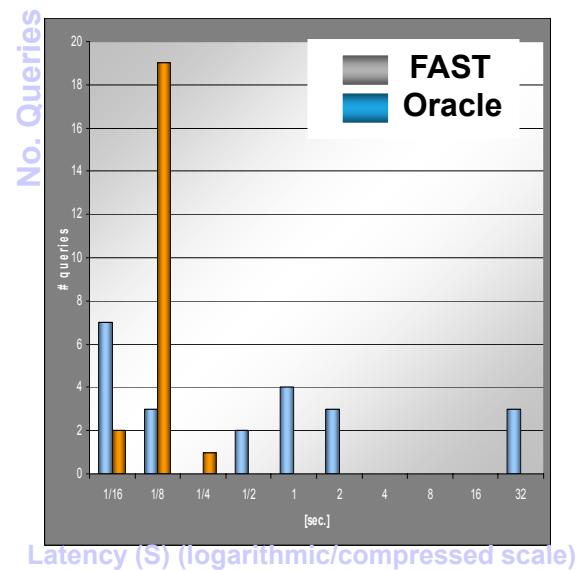


1.



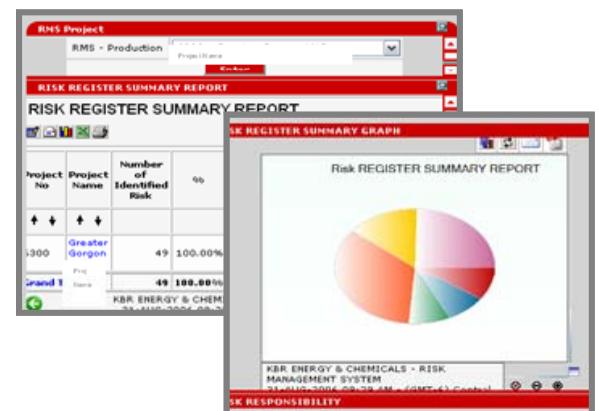
Better Business decisions with accurate data

2.



Increased productivity with ad-hoc, real-time performance

3.

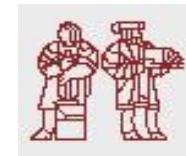


Increased Adoption with easy and intuitive access to intelligence

fast



Analytical Portals built on Search



REUTERS Search

Search All Equities Debt FX & Money Commodities & Energy Funds Economic Indicators Indices M & A

Companies Warrants Options Futures

Analyse Results Back to Results

Filtered by:

Select Filter:

Pretax Profit (USD)

Less than -1522840 8

Between -1522840 & 928711 8

928711 & 8.1075m 9

Greater than or equal to 8.1075m 8

Revenue Growth 5 Years

Less than -4 6

Between -4 & 7 7

7 & 27 6

Greater than or equal to 27 6

Revenue (USD)

Less than 23,3184m 8

Between 23,3184m & 43,9762m 8

43,9762m & 176,453m 9

Greater than or equal to 176,453m 5

Reuters 3000 Xtra Search

Welcome to FAST Radar
Thursday, August 3, 2006 11:35 AM

fast

My Portal Administration Data Tools Dealer Metrics Sales Forecast

YTD Sales Goal

Value: \$438,886 Target: \$578,014 Status: 75.93%

YTD Variance

Low: \$40,000 High: \$20,000 Current Value: \$21,513

Plant Ratio

Unacceptable: 3.00% Caution: 4.00% Value: 5.01%

Andrew Fuller Sales

Unacceptable: \$20,000 Caution: \$25,000 Value: \$21,757

Sales vs. Forecast

Sales vs. Forecast

page 1 of 1 9 of 9 records

| Sales Rep | 2002 | | | 2003 | | | 2004 | | |
|--------------------|------------------|------------------|-----------------|------------------|------------------|-----------------|------------------|------------------|-----------------|
| | Sales | Forecast | Variance | Sales | Forecast | Variance | Sales | Forecast | Variance |
| Andy Fuller | \$72,042 | \$79,253 | \$7,211 | \$72,739 | \$80,016 | \$7,277 | \$21,757 | \$23,935 | \$2,178 |
| Anne Dodsworth | 26,310 | 28,941 | 2,631 | 41,103 | 45,214 | 4,111 | 9,895 | 10,883 | 988 |
| Janet Leverling | 108,026 | 118,836 | 10,810 | 76,563 | 84,222 | 7,659 | 18,224 | 20,050 | 1,826 |
| Laura Callahan | 56,033 | 61,640 | 5,607 | 48,590 | 53,455 | 4,865 | 22,220 | 24,444 | 2,224 |
| Margaret Peacock | 128,810 | 141,699 | 12,889 | 54,136 | 59,553 | 5,417 | 49,946 | 54,945 | 4,999 |
| Michael Suyama | 43,126 | 47,445 | 4,319 | 14,144 | 15,559 | 1,415 | 16,643 | 18,308 | 1,665 |
| Nancy Davolio | 86,216 | 94,846 | 8,630 | 63,195 | 69,520 | 6,325 | 42,696 | 46,966 | 4,270 |
| Robert King | 60,611 | 66,674 | 6,063 | 48,725 | 53,602 | 4,877 | 15,240 | 16,766 | 1,526 |
| Steven Buchanan | 30,716 | 33,789 | 3,073 | 19,692 | 21,662 | 1,970 | 18,342 | 20,178 | 1,836 |
| Grand Total | \$611,892 | \$673,123 | \$61,231 | \$438,886 | \$482,803 | \$43,917 | \$214,962 | \$236,475 | \$21,513 |

Imported into CR_PYRAMIDS from Northwind: 3/7/2005

Sales vs. Forecast Chart

Sales vs. Forecast

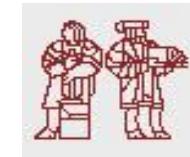
Legend: Sales, Forecast, Sales, Sales

| Sales Rep | Sales 1 | Forecast 1 | Sales 2 | Forecast 2 |
|------------------|---------|------------|---------|------------|
| Andy Fuller | 72,042 | 79,253 | 72,739 | 80,016 |
| Anne Dodsworth | 26,310 | 28,941 | 41,103 | 45,214 |
| Janet Leverling | 108,026 | 118,836 | 76,563 | 84,222 |
| Laura Callahan | 56,033 | 61,640 | 48,590 | 53,455 |
| Margaret Peacock | 128,810 | 141,699 | 54,136 | 59,553 |
| Michael Suyama | 43,126 | 47,445 | 14,144 | 15,559 |
| Nancy Davolio | 86,216 | 94,846 | 63,195 | 69,520 |
| Robert King | 60,611 | 66,674 | 48,725 | 53,602 |
| Steven Buchanan | 30,716 | 33,789 | 19,692 | 21,662 |

fast



FAST Vision For Pervasive Business Intelligence



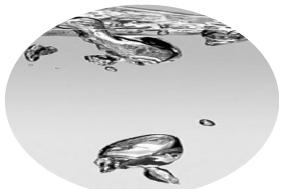
Ad-hoc, FAST access



Exabyte scale information management
Structured/Semi-structured/Unstructured



EASY analytics



Data Improvement
“Content Capture and Refinement”



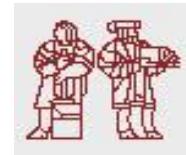
Connections across information
“Semantic Integration”



Real Time information



The MIT Information Quality Industry Symposium, 2007



Thank you!

fast
find the real value of search

Jeff Fried
VP Advanced Solutions
Jeff.Fried@fastsearch.com

fast