Linguistic Data Cleansing Case Studies

Damien Islam-Frénoy
Director, Strategic Market Development
Damien.Fenoy@fastsearch.com

FAST Co-Innovates with Customers - Using Search on Structured Data

…and many more (est. 65-75% of customer implementations involve structured data)

2

•2

Case Studies – using linguistic data matching, consolidation, and cleansing

- White pages cleansing and consolidation
- Name & Address consolidation, cleansing, and house-holding
- Fraud detection
- Product catalog cleansing and rollup
- Financial data matching and exploration

fast

Standard Phone No. Search is Severely Limited

U.S. - anywho.com

FIND A BUSINESS OR PERSON BY PHONE NUMBER

Area Code *Required* 781

Telephone Number *Required* 3042400

SEARCH

TIP: Cell phone numbers are not available

You searched for: 781 3042400

Results 1 - 1 of 1 PREVIOUS | NEXT

Reverse Telephone Listings

Fast, Search & Transfer
93 Worcester St
Wellesley, MA 02481

781-304-2400

Wrong address listed!

Maps & Directions | Did you go to school with Search & Transfer Fast?
Public Records Results for Search & Transfer Fast. Preview Results.
Instant Background Check Available - $49.95!
Find a Nearby Business

Norway – gulesider.no

Bjørn Boberg
siv.økonom

Jomfrubråtv. 29 C, 1179 Oslo ... 22 67 94 55
- Sommerbolig
Solvarp, 1794 Sponvika 69 19 45 17
- Mobiltelefon 928 97 133

1 Map | Send as SMS | Save | Call

Send flowers

fast

Contrast to Results with Sesam.no: Creating Cleansed Business Information from Blended Content

Find the business who called our Recruitment Officer

Sesam creates composite objects by joining and cleansing data

Three Years of Financial Data showing growth and profits

Blended Content: A blog on the founder of Boyden International

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

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

Sesam.no: Winner of 2006 Nordic Data Warehouse Award

FAST Data Cleansing
Telstra

Leading telecom and information services company, competing in all telecommunications markets throughout Australia, providing more than 9.94 million Australian fixed line and more than 8.5 million mobile services.

Challenge:

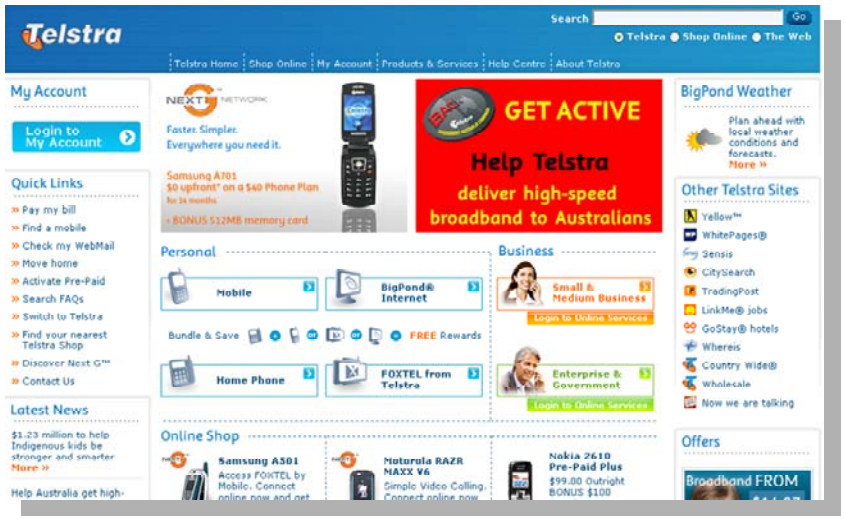
- Over 240 Telstra systems use address data, therefore quality of this data is critical
- Poor data quality in processes such as activation, billing, customer contact, marketing.
- Provide correct legal, mailing and service addresses used for emergency services

Solution:

- Developed a search service that incorporates a configurable query transformation pipeline
- Automatically "corrects" user input when queries do not match any results and quickly expand or modify the search to find a result set.
- Provides an Address Search web service for the enterprise
- Live synchronization against an address database
- External system interactions continue with the database (National Postal System etc)

fast

Australia's leading telecom provider



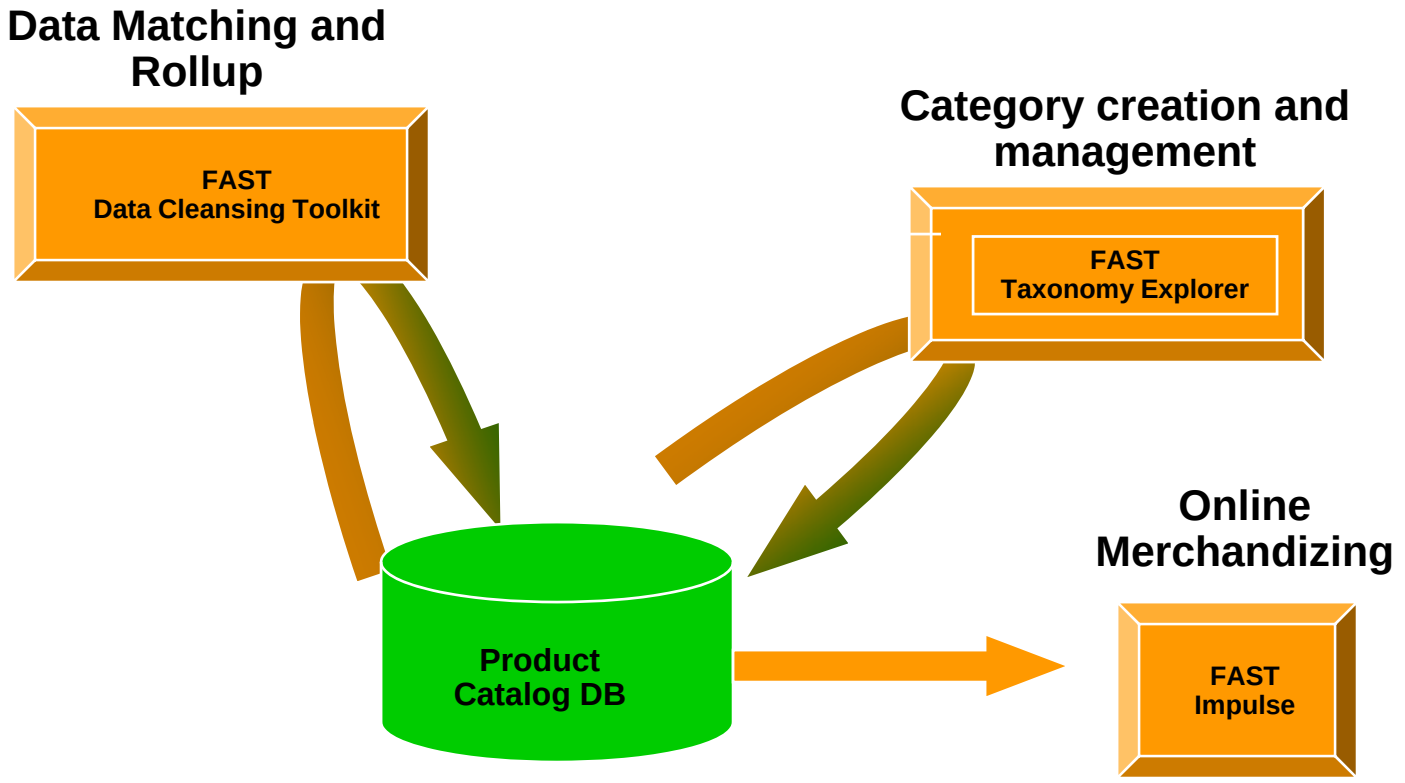
Results with FAST:

- ✓ Reduced Costs: Reduced ineffective truck rolls due to incorrect service address
- ✓ Increased Productivity: Reduced time in contact centers validating addresses
- ✓ Increase Customer Satisfaction: Decrease the incidence of unsuccessful self-service due to poor address search, resulting in a call to a live agent.

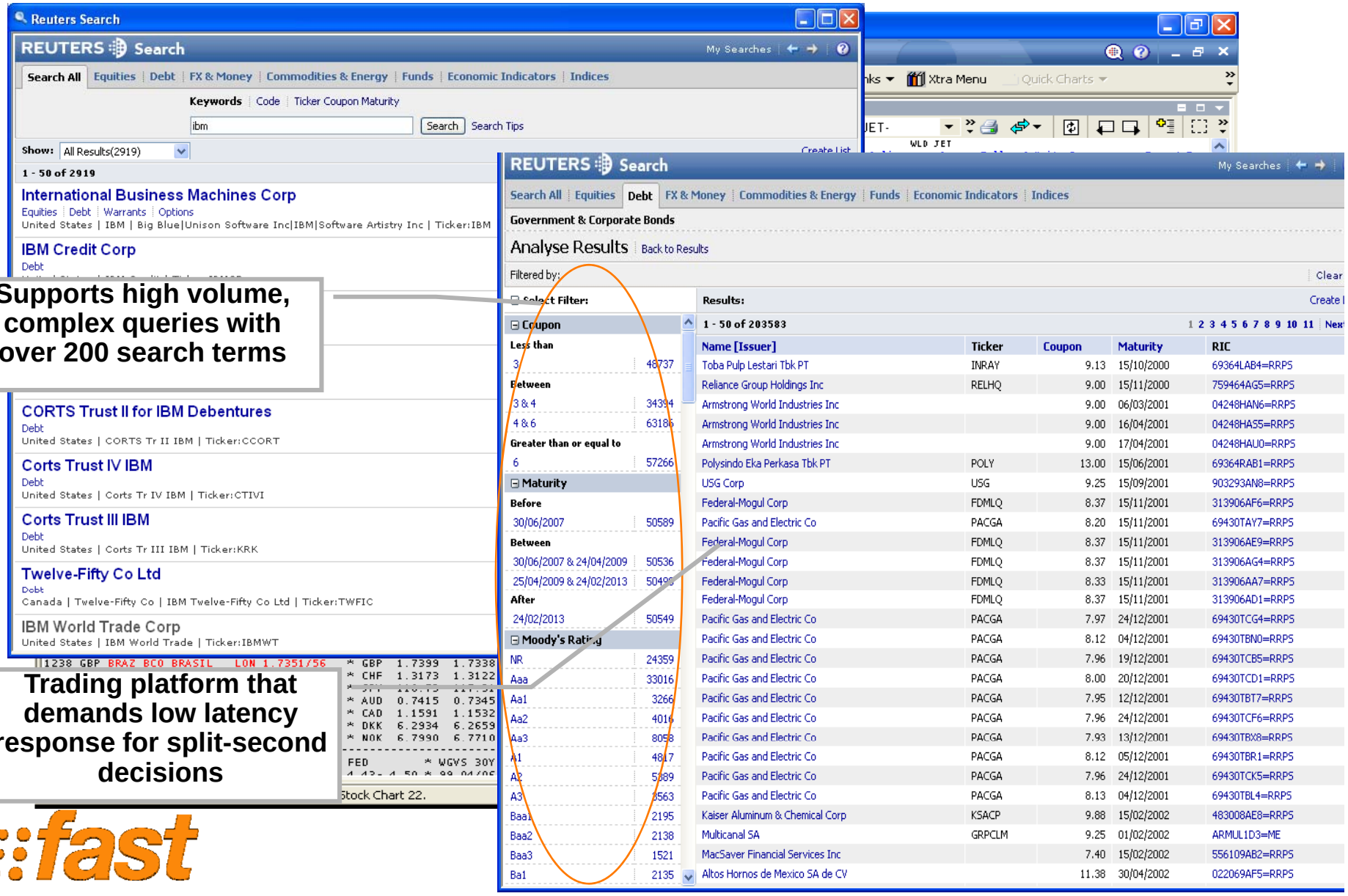
Fraud Detection example - insurance

- Detect “**patient tourism**”: Care Providers cooperating to pull as much money from the insurance company as possible when treating a patient that has been in an accident. After investigation, the first doctor passes the patient to other doctors (typically friends) that also investigate and do some treatment, and then pass on to next "friend". That way they are able to get a higher total amount from the insurance company.
- Detect "**ping-ponging**" of patients: This is a specialized version of Use Case 1, where N=2 (two doctors involved). For example, this could be your primary care physician (PCP) sending you to a specialist, the specialist sending you back to your PCP ("for checkup"), and then your PCP sending you back to the specialist ("for additional treatment touch-up"), and so on.
- Detect "**treatment prolongation**": Find doctors that cause the longest "time-from-the-accident-till-patient-was-cured". The insurance company pays their salary in this period. FAST is creating a system for coordinators to assigning Care Providers when new patients came in. Type in treatment type and get a list of doctors, sorted on which doctor caused shortest treatment time, lowest price and lowest invalid-level.
- Detect "**improper drug prescriptions**": Find doctors who are prescribing unnecessary or improper drug treatments, for example by prescribing patented drugs when a generic is available or by prescribing more drugs than are required for treatment. Then the doctors get kickbacks from the pharmacy.

Product Catalog Cleansing and Rollup



Extreme Performance For Demanding Environments



Supports high volume, complex queries with over 200 search terms

Trading platform that demands low latency response for split-second decisions

fast

The MIT Information Quality Industry Symposium, 2007

Thank you!

fast

find the real value of search

Damien Islam-Frénoy
Director, Strategic Market Development
Damien.Fenoy@fastsearch.com

fast