



The MIT Information Quality Industry Symposium, 2007



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)





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



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

Wrong address listed!

781-304-2400

[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](#)

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](#)

Norway – [gulesider.no](#)





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

Boyden

Bloggsøk Beta: 1 blogg ga treff på boyden

Skeptics alert: Astrology and Leadership

Personsøk:

Kontaktinfo **Roller**

Jan Gunnar Schumann

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

Boyden As

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

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 dato: 11.01.2005

Kontaktinfo **Regnskapstall**

Jan Gunnar Schumann

Adresse Kollenveien 14 B, 0170 Oslo

Telefon 64 94 53 7

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

Bloggsøk Beta: 1 blogg ga treff på boyden

Skeptics alert: Astrology and Leadership

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

Personsøk:

Kontaktinfo **Roller**

Jan Gunnar Schumann

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

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 dato: 11.01.2005

Kontaktinfo **Regnskapstall**

Jan Gunnar Schumann

Adresse Kollenveien 14 B, 0170 Oslo

Telefon 64 94 53 7

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

Bloggsøk Beta: 1 blogg ga treff på boyden

Skeptics alert: Astrology and Leadership

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

Personsøk:

Kontaktinfo **Roller**

Jan Gunnar Schumann

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

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 dato: 11.01.2005

Kontaktinfo **Regnskapstall**

Jan Gunnar Schumann

Adresse Kollenveien 14 B, 0170 Oslo

Telefon 64 94 53 7

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

Bloggsøk Beta: 1 blogg ga treff på boyden

Skeptics alert: Astrology and Leadership

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

Personsøk:

Kontaktinfo **Roller**

Jan Gunnar Schumann

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

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 dato: 11.01.2005

Kontaktinfo **Regnskapstall**

Jan Gunnar Schumann

Adresse Kollenveien 14 B, 0170 Oslo

Telefon 64 94 53 7

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

Bloggsøk Beta: 1 blogg ga treff på boyden

Skeptics alert: Astrology and Leadership

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

Personsøk:

Kontaktinfo **Roller**

Jan Gunnar Schumann

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

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 dato: 11.01.2005

Kontaktinfo **Regnskapstall**

Jan Gunnar Schumann

Adresse Kollenveien 14 B, 0170 Oslo

Telefon 64 94 53 7

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			

Bloggsøk Beta: 1 blogg ga treff på boyden

Skeptics alert: Astrology and Leadership

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

Personsøk:

Kontaktinfo **Roller**

Jan Gunnar Schumann

Bedriftssøk: 1 treff på jan gunnar schumann

Jan Gunnar Schumann

Styrets leder [Jan Gunnar Schumann \(f1942\)](#)

Styremedlem Kaare Bringa (f 1949)

Varamedlem Erle Li Dragland (f 1957)

Varamedlem

Regnskapstall

	2005	2004	2003
Resultatregnskap			
Sum driftsinntekter	3.249	3.492	2.895
Driftsresultat	2.446	1.431	-47
Resultat før skatt	2.446	1.456	44
Resultat			
Resultat før skatt			
Resultat			



**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

The screenshot shows the Telstra website homepage. At the top, there's a navigation bar with the Telstra logo and links for Home, Shop Online, My Account, Products & Services, Help Centre, and About Telstra. Below this, there are several promotional banners and service tiles. A prominent red banner says 'GET ACTIVE Help Telstra deliver high-speed broadband to Australians'. Other tiles include 'My Account' with a 'Login to My Account' button, 'Quick Links' with various service options, 'Personal' services like Mobile, Home Phone, and BigPond Internet, 'Business' services like Small & Medium Business and Enterprise & Government, and 'Online Shop' featuring Samsung A501, Motorola RAZR MAXX V6, and Nokia 2610. There's also a 'BigPond Weather' section and 'Other Telstra Sites' like Yellow Pages, Sensis, and CitySearch.

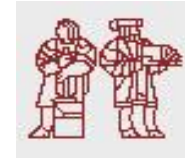
Results with FAST:

- ✓ 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.

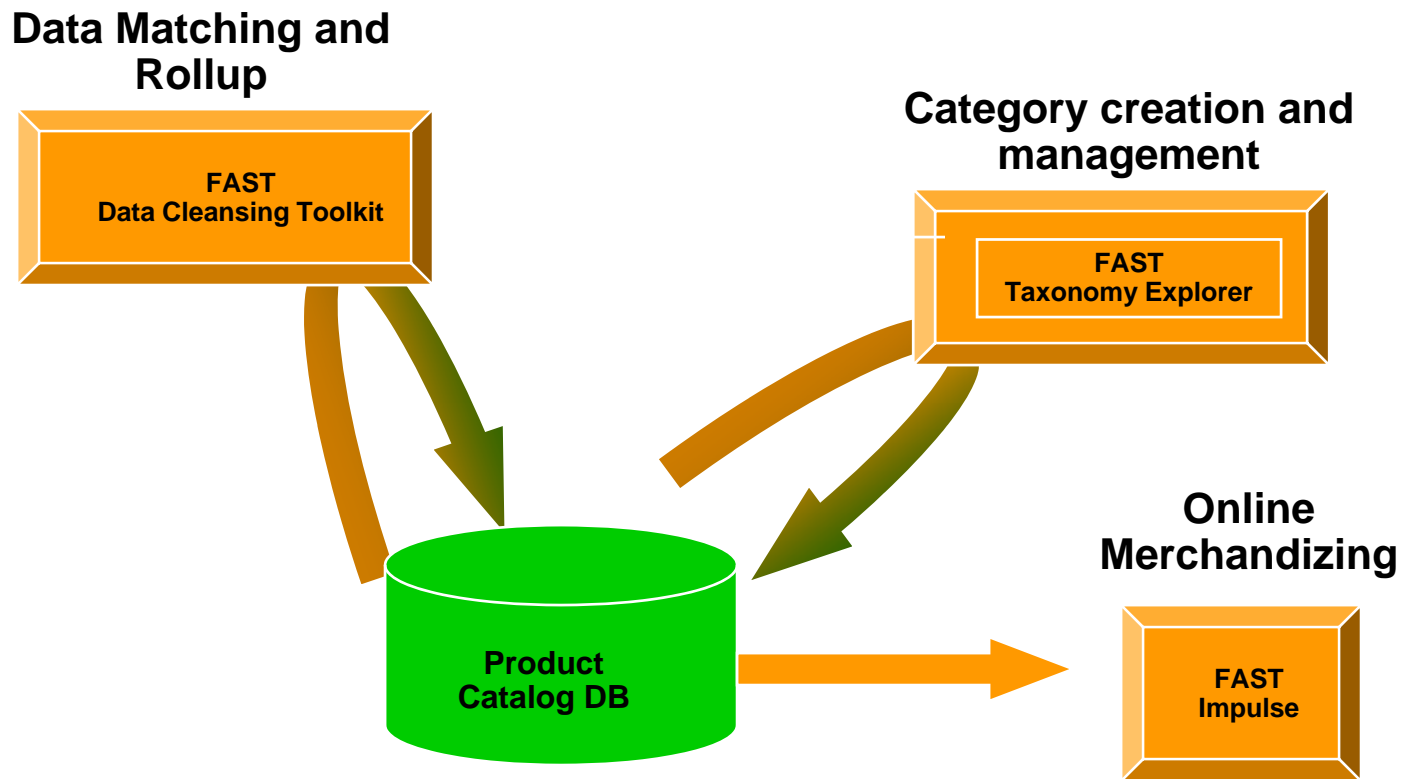


Fraud Detection example - insurance

- 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

The image displays two overlapping screenshots of the Reuters Search application. The left screenshot shows search results for 'IBM', listing various entities like 'International Business Machines Corp' and 'IBM Credit Corp'. The right screenshot shows search results for 'Government & Corporate Bonds', displaying a table of bond details.

Filter	Count	Results
Coupon		
Less than 3	46737	
Between 3 & 4	34394	
4 & 6	63186	
Greater than or equal to 6	57266	
Maturity		
Before 30/06/2007	50589	
Between 30/06/2007 & 24/04/2009	50536	
25/04/2009 & 24/02/2013	50490	
After 24/02/2013	50549	
Moody's Rating		
NR	24359	
Aaa	33016	
Aa1	3266	
Aa2	4016	
Aa3	8058	
A1	4817	
A2	5389	
A3	3563	
Baa1	2195	
Baa2	2138	
Baa3	1521	
Ba1	2135	

Name [Issuer]	Ticker	Coupon	Maturity	RIC
Toba Pulp Lestari Tbk PT	INRAY	9.13	15/10/2000	69364LAB4=RRPS
Reliance Group Holdings Inc	RELHQ	9.00	15/11/2000	759464AG5=RRPS
Armstrong World Industries Inc		9.00	06/03/2001	04248HAN6=RRPS
Armstrong World Industries Inc		9.00	16/04/2001	04248HAS5=RRPS
Armstrong World Industries Inc		9.00	17/04/2001	04248HAU0=RRPS
Polysindo Eka Perkasa Tbk PT	POLY	13.00	15/06/2001	69364RAB1=RRPS
USG Corp	USG	9.25	15/09/2001	903293AN8=RRPS
Federal-Mogul Corp	FDMLQ	8.37	15/11/2001	313906AF6=RRPS
Pacific Gas and Electric Co	PACGA	8.20	15/11/2001	69430TAY7=RRPS
Federal-Mogul Corp	FDMLQ	8.37	15/11/2001	313906AE9=RRPS
Federal-Mogul Corp	FDMLQ	8.37	15/11/2001	313906AG4=RRPS
Federal-Mogul Corp	FDMLQ	8.33	15/11/2001	313906AA7=RRPS
Federal-Mogul Corp	FDMLQ	8.37	15/11/2001	313906AD1=RRPS
Pacific Gas and Electric Co	PACGA	7.97	24/12/2001	69430TCG4=RRPS
Pacific Gas and Electric Co	PACGA	8.12	04/12/2001	69430TBNO=RRPS
Pacific Gas and Electric Co	PACGA	7.96	19/12/2001	69430TCB5=RRPS
Pacific Gas and Electric Co	PACGA	8.00	20/12/2001	69430TCD1=RRPS
Pacific Gas and Electric Co	PACGA	7.95	12/12/2001	69430TBT7=RRPS
Pacific Gas and Electric Co	PACGA	7.96	24/12/2001	69430TCF6=RRPS
Pacific Gas and Electric Co	PACGA	7.93	13/12/2001	69430TBX8=RRPS
Pacific Gas and Electric Co	PACGA	8.12	05/12/2001	69430TBR1=RRPS
Pacific Gas and Electric Co	PACGA	7.96	24/12/2001	69430TCK5=RRPS
Pacific Gas and Electric Co	PACGA	8.13	04/12/2001	69430TBL4=RRPS
Kaiser Aluminum & Chemical Corp	KSACP	9.88	15/02/2002	483008AE8=RRPS
Multicanal SA	GRPCLM	9.25	01/02/2002	ARMUL1D3=ME
MacSaver Financial Services Inc		7.40	15/02/2002	556109AB2=RRPS
Altos Hornos de Mexico SA de CV		11.38	30/04/2002	022069AF5=RRPS

11238 GBP BRAZ BCO BRASIL LON 1.7351/56 * GBP 1.7399 1.7338 * CHF 1.3173 1.3122 * AUD 0.7415 0.7345 * CAD 1.1591 1.1532 * DKK 6.2934 6.2659 * NOK 6.7990 6.7710 * FED * WGVS 30Y 4.43 - 4.50 * 99.04/00 Stock Chart 22.





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