



reifier.

Real Time Fuzzy Matching With Spark and
ElasticSearch

About Us

The only way to do great work is to love what you do.

- Steve Jobs

The problem - lake or swamp?



Duplicates



Challenges

- Quadratic problem
- No standard notion of similarity
- Omissions, typos and other issues
- Different languages

Use Case - Customer Record Dedup



Use Case - Customer Record Dedup




Use Case - Shopping Site Comparison



Use Case - Shopping Site Comparison

Apple iPhone 6 With FaceTime - 64GB, 4G LTE, Space Gray
by [Apple, Mobile Phones](#)
★★★★★ (271 reviews)



Apple iPhone 6 64GB LTE Smartphone Space Gray
★★★★★ 4.8 (1228 reviews)
In stock - Dispatched within 2 hours. ?

AED2,779.00 Availability: In stock
* Required Fields

Bundle

- ☐ PanzerGlass Tempered Glass Screen Protector + AED129.00
- ☐ Stylus Pen+Selfie Pod+Car Charger+ Case+Screen Guard + AED49.00

Other Use Cases

- Cross selling
- Financial Credit Ratings
- Fraud Analytics
- Catalog and inventory management
- Household and individual level analytics.

Lets start wishing...

- Data variety
- Scalable
- No manual configuration of rules or algorithms
- Multi language
- Real time

Reifier - learn

Jobs Support Logout

Do the following pairs match ?

Match 0/0

Non Match 0/0

Unsure 0/0

brooklyn	blake	63	o'connell street		eden hills	810	vic	19101201	2258884
brooklyn	jehne	2	dumaresq street	rowethorpe	lamington	6061	wa	19420723	2935590
<div>YesNoCan't Say</div>									
caitlin	noack	28	hamilton row	sec 375	manly vale	7011	vic		8328282
caitlin	westbrook	631	casson street	blacksdane private sanctuary	gidgegannup	3114	nsw	19360908	9439473
<div>YesNoCan't Say</div>									

Reifier - learn

Jobs Support Logout

Do the following pairs match ?

Match 0/0

Non Match 0/0

Unsure 0/0

<u>brooklyn</u>	blake	63	<u>o'connell street</u>		eden hills	810	vic	19101201	2258884
-----------------	-------	----	-------------------------	--	------------	-----	-----	----------	---------

<u>brooklyn</u>	jehne	2	<u>dumaresq street</u>	rowethorpe	lamington	6061	wa	19420723	2935590
-----------------	-------	---	------------------------	------------	-----------	------	----	----------	---------

Yes

No

Can't Say

<u>caitlin</u>	noack	28	hamilton row	sec 375	manly vale	7011	vic		8328282
----------------	-------	----	--------------	---------	------------	------	-----	--	---------

<u>caitlin</u>	westbrook	631	casson street	blacksdane private sanctuary	gidgegannup	3114	nsw	19360908	9439473
----------------	-----------	-----	---------------	------------------------------	-------------	------	-----	----------	---------

Yes

No

Can't Say

Reifier - learn

Jobs Support Logout

Do the following pairs match ?

Match 10/35

Non Match 24/35

Unsure 1/35

isabellaryankeira streetpyndari

isabellahingston57loureiro street

Yes

No

Can't Say

nicholaswilkey12caladenia streetrivonia

nicholaswickel12jardine streetwuikikakiwi

Yes

No

Can't Say


toowoomba3500vic194004056806487

thornlie3500nsw192901254688367

kingston2480wa1399646

highton3085wa195706202126121

© Nube Technologies

 reifier.

Reifier - learn

Jobs Support Logout

Do the following pairs match ?

Match 24/65

Non Match 38/65

Unsure 3/65

<u>jayden</u>	peckett		<u>burkitt street</u>	bexley		<u>bunnaloo</u>	3169	<u>vic</u>	19700626	4467399
<u>jayden</u>	royle	29	<u>broadsmith street</u>	b		<u>pinnaroo</u>	2036	<u>vic</u>	19121110	3178830
<div>Yes No Can't Say</div>										
<u>nicholas</u>	<u>howie</u>	42	govett aplace	<u>cheshite</u>		west perth	<u>6105</u>	<u>qld</u>	<u>19634028</u>	<u>1019666</u>
<u>nicholas</u>	<u>howie</u>	25	antares crescent	<u>cheshi re</u>		flagstaff hill	<u>6104</u>	<u>qld</u>	<u>19630428</u>	<u>1019666</u>
<div>Yes No Can't Say</div>										

Real Time

Spark + ElasticSearch



Spark Benefits

- Distributed
- Scalable
- Fast
- Machine Learning
- Sampling
- No need to orchestrate multiple jobs



Thank You!

www.nubetech.co

sonal@nubetech.co