



VALIDATION ANALYTICS – DATA YOU NEED AT THE SPEED YOU NEED IT

Cynthia Cook

Software Architect | Intel Corporation

9/21/2017

What Problem Are We Trying to Solve

- Validation is facing a data problem at exponential scale
- Data needs at collection time are different from data needs at the point of validation and are difficult to predict
- Critical pieces of information are often siloed or not accessible
- Late decisions and bad decisions cause rework that compounds the problem
- Algorithm development becoming more costly

entire

Accuracy and Trust

- One piece of inaccurate data affects all systems downstream
- Mistrust remediation typical involves more validation on dependent systems (more data)

Mistrust is a significant source of waste in an increasingly complex system



Traditional Attempts



- Faster hardware / software
- Data warehousing & software convergence
- Human expertise
- Standardization & simplification
- UX

Intelligent Analytics

Intel is currently engaged with an external partnership on a cognitive computing intelligent analytics validation effort

Algorithm-less noise reduction

- Identify duplicate errors to reduce waste

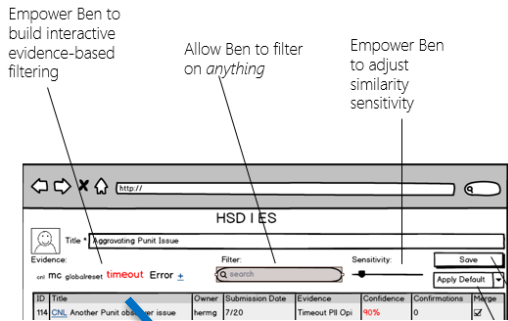
Prioritize debug research data points

- Suggest anomalous and important data first and then peel the onion

Learn naturally from human expertise

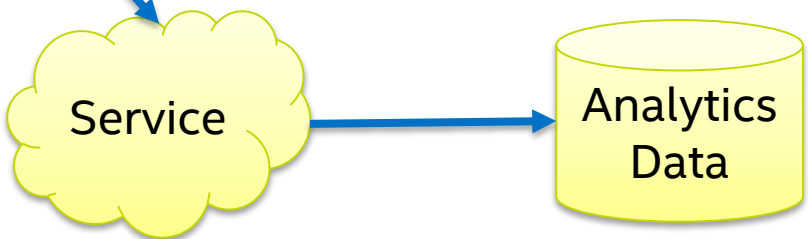
- Store human discovery as naturally occurring data that helps the system improve itself

Intelligent Analytics

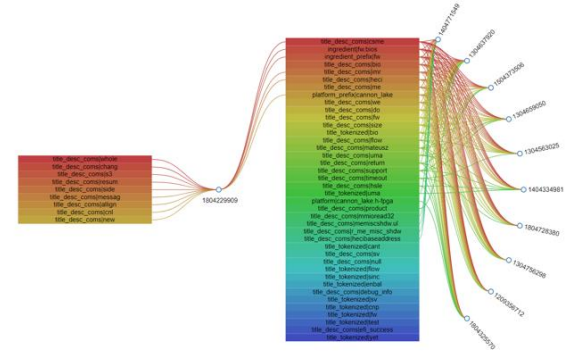


Model 1
current best

Human
current best



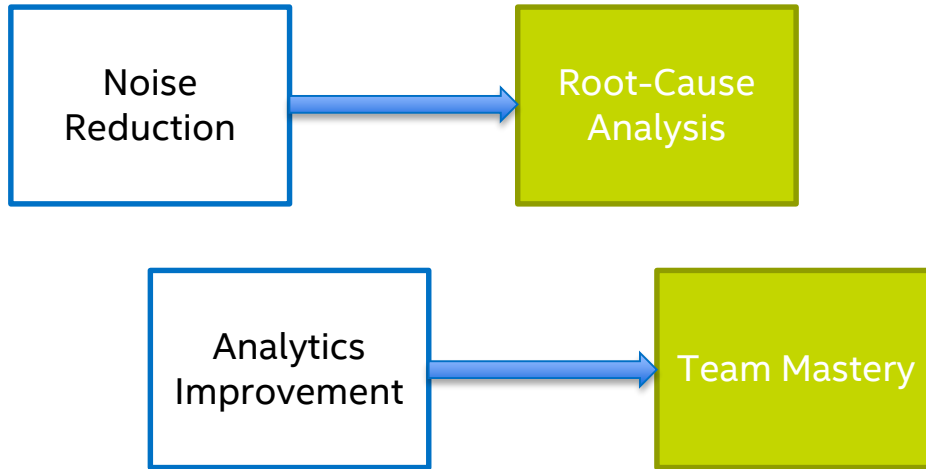
Evidence Visualizations



Same Position Different Position Not Found

paper ID	source ID	position found	result 1	result 2	result 3	result 4	result 5	result 6	result 7	result 8	result 9	result 10
10431471	10431400	1	10431400	10431401	10431402	10431403	10431404	10431405	10431406	10431407	10431408	10431409
10432478	12972152	0	10432478	10432479	10432480	10432481	10432482	10432483	10432484	10432485	10432486	10432487
10432847	10432871	2	10432872	10432873	10432874	10432875	10432876	10432877	10432878	10432879	10432880	10432881
10433219	10433068	1	10433069	10433070	10433071	10433072	10433073	10433074	10433075	10433076	10433077	10433078
10432738	10431878	0	10431879	10431880	10431881	10431882	10431883	10431884	10431885	10431886	10431887	10431888
10432210	14907752	1	14907753	14907754	14907755	14907756	14907757	14907758	14907759	14907760	14907761	14907762
10427953	10440911	0	10440912	10440913	10440914	10440915	10440916	10440917	10440918	10440919	10440920	10440921
10429676	15481700	0	15481701	15481702	15481703	15481704	15481705	15481706	15481707	15481708	15481709	15481710
10421443	12770928	0	12770929	12770930	12770931	12770932	12770933	12770934	12770935	12770936	12770937	12770938
10424464	10443471	0	10443472	10443473	10443474	10443475	10443476	10443477	10443478	10443479	10443480	10443481

Where Do We Go From Here?



Ethics of Data Presentation



