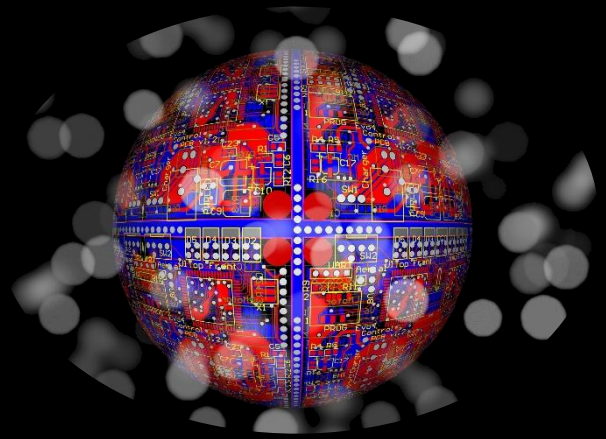
ROBERT GORDON UNIVERSITY ABERDEEN

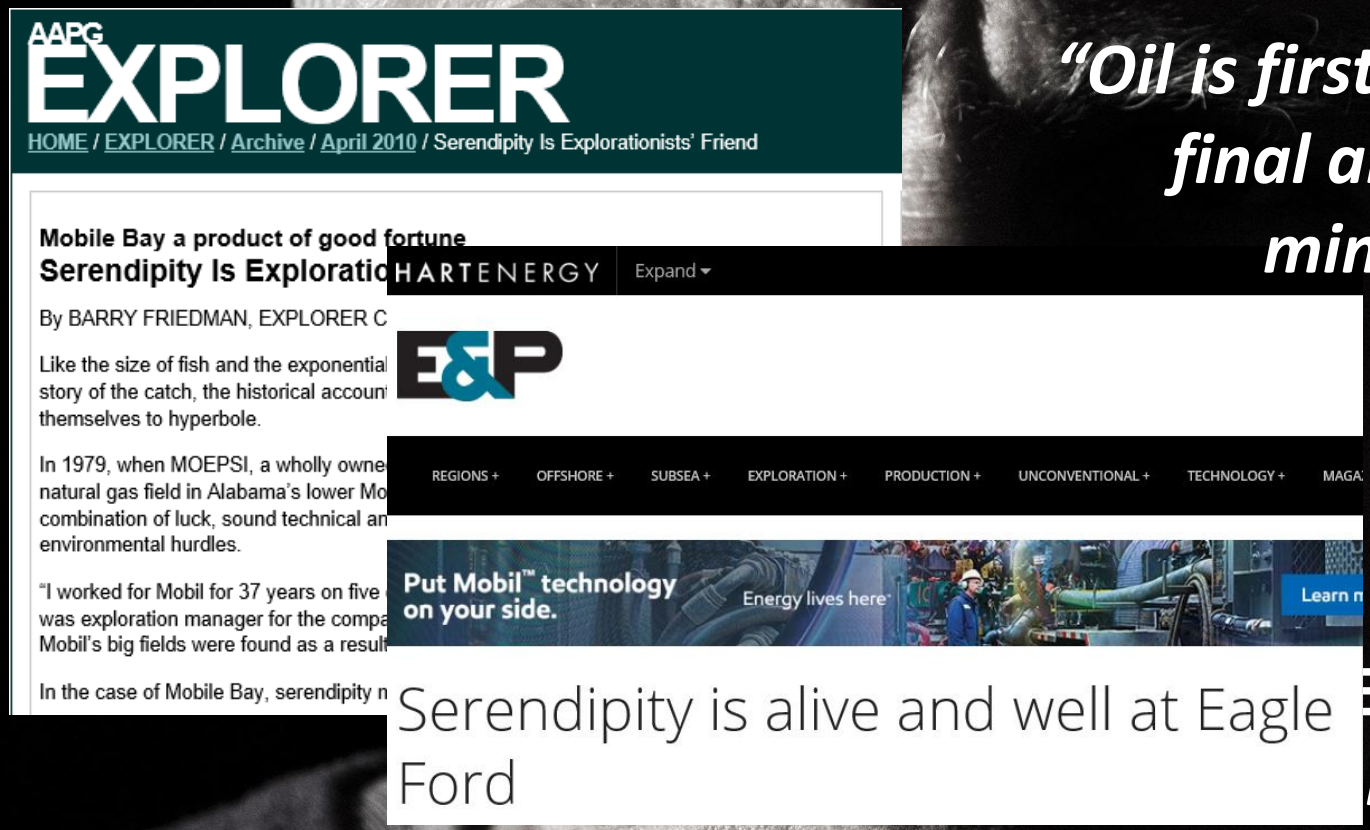
Using machine learning to stimulate ideas for geoscientists

Some ideas for using text analytics in enterprise search to surface the unexpected



Paul H. Cleverley

Robert Gordon University, UK



"Oil is first found, in the final analysis, in the minds of people"

AAPG

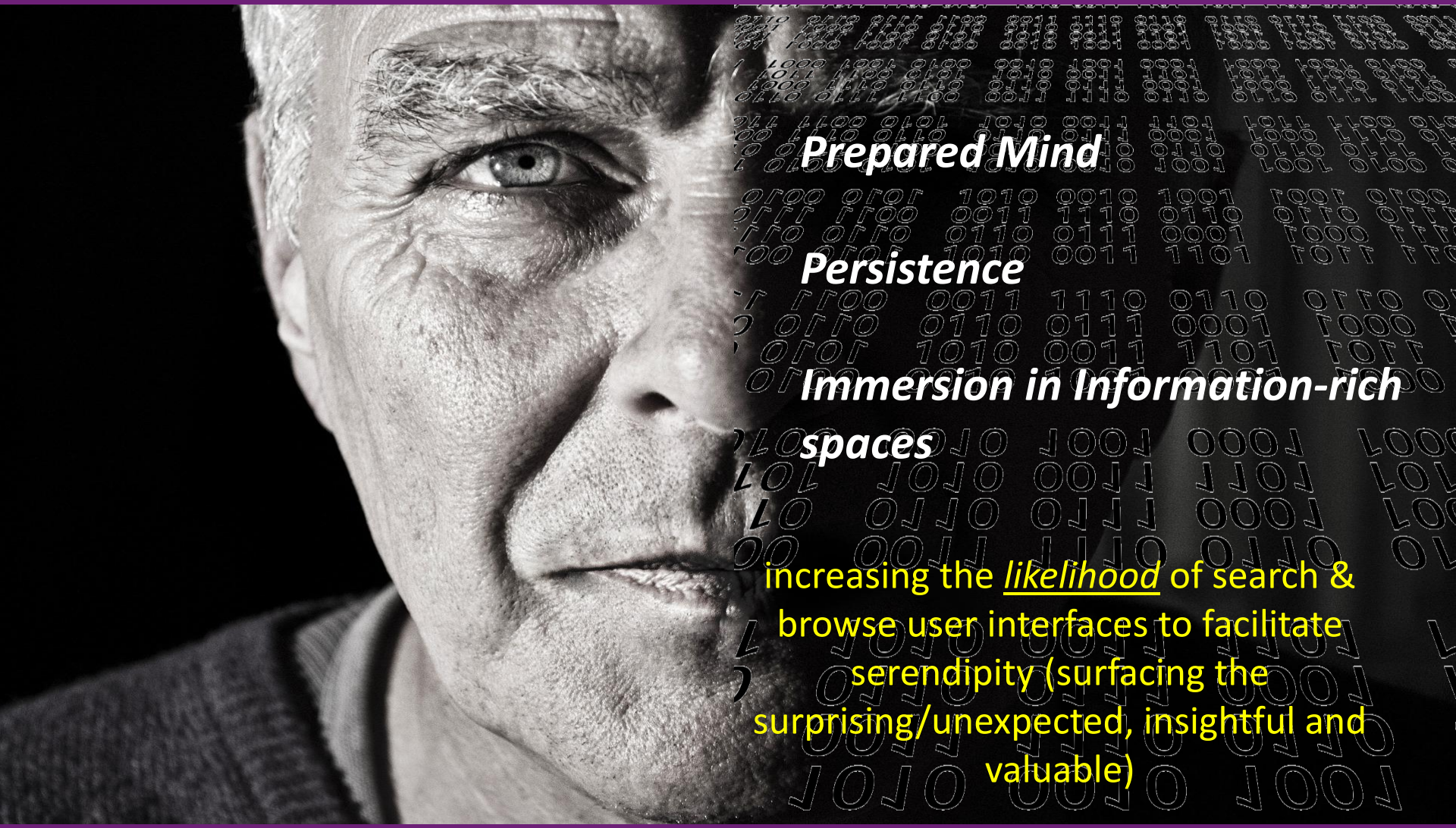
Pratt 1952

"Exploration is in ideas business"

Exploration Team Lead

January 2016

ROBERT GORDON UNIVERSITY ABERDEEN



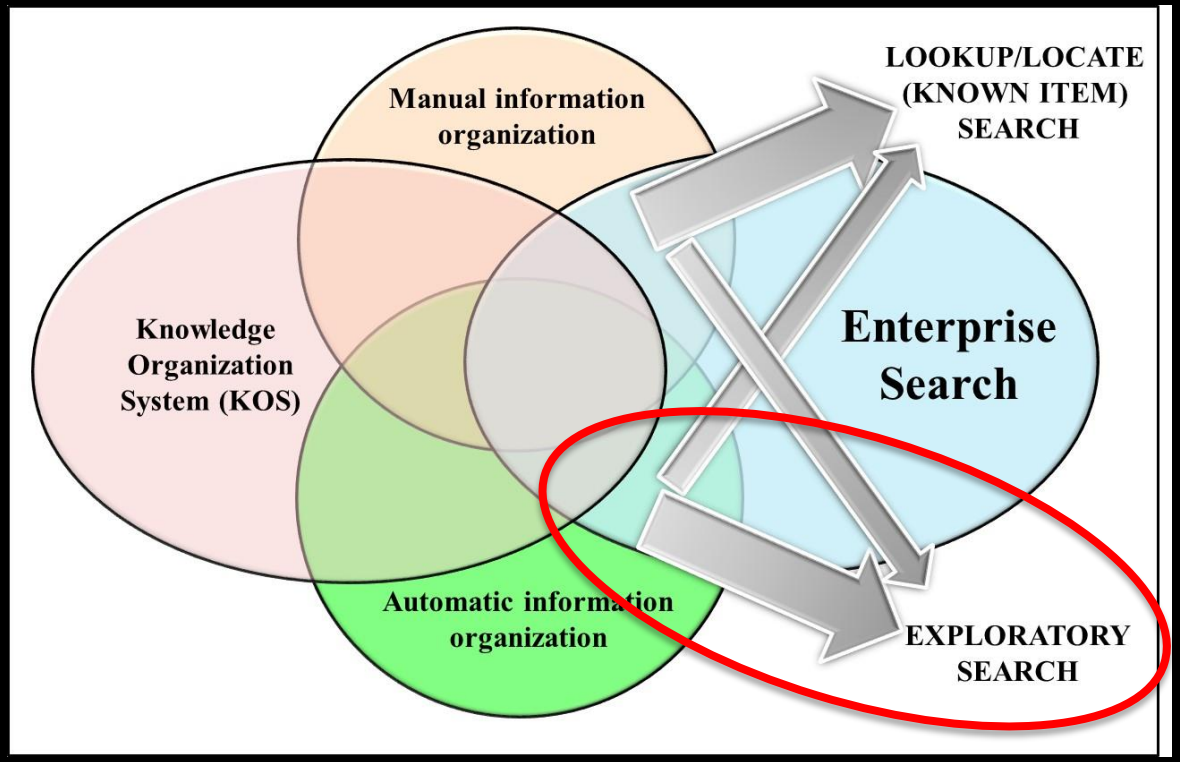
Prepared Mind

Persistence

Immersion in Information-rich spaces

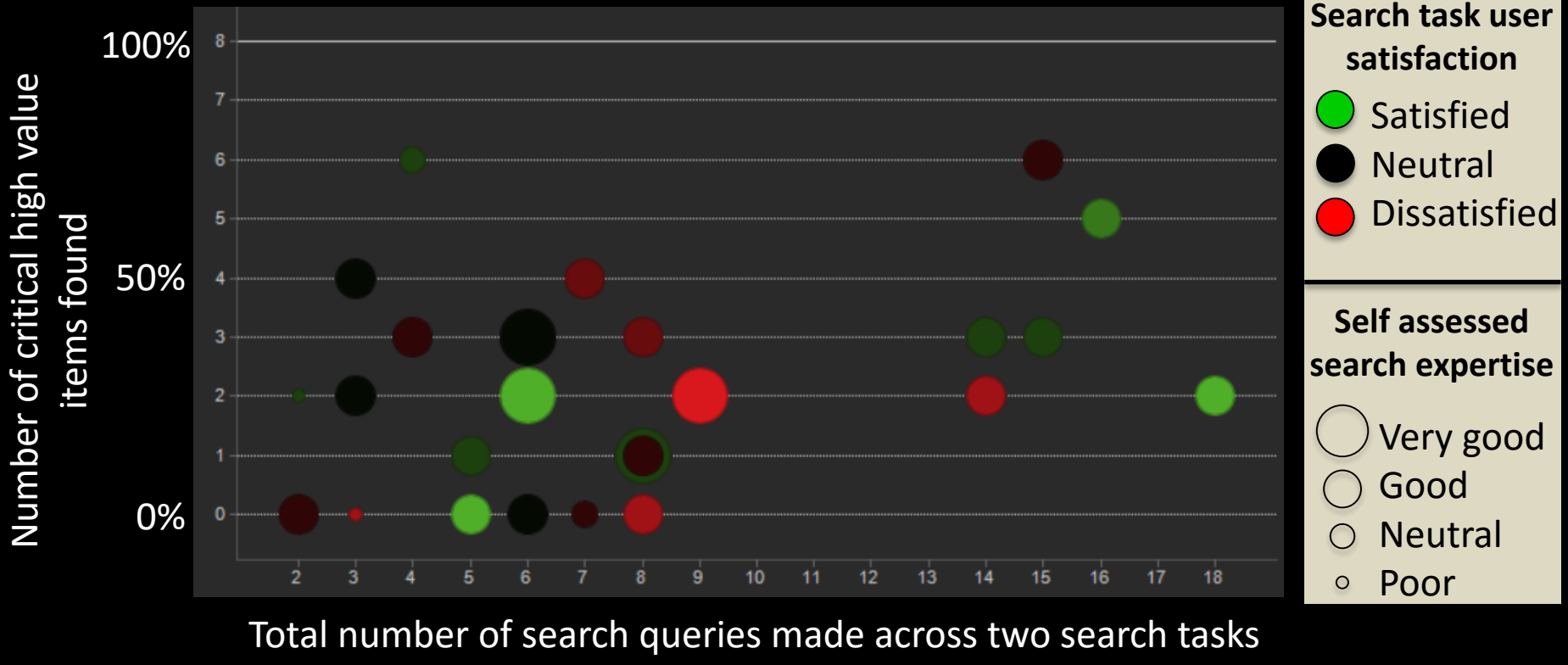
increasing the *likelihood* of search & browse user interfaces to facilitate serendipity (surfacing the surprising/unexpected, insightful and valuable)

Information Searching: Two main goals (Marchionini 2006)



Journal of Knowledge Organization (Cleverley and Burnett 2015)

Enterprise search experiment - 26 Professionals in a case study O&G company
Found only 27% of high value items on average but 60% satisfied.



Journal of the Association for Information Science and Technology (Cleverley and Burnett 2015)

ROBERT GORDON UNIVERSITY ABERDEEN

The Five Dimensions of Serendipity in Digital Environments (McCay-Peet & Toms 2011)

Introducing variety

Introducing the unexpected

Enabling connections between topics

Triggering divergence

Inducing curiosity

Issues and challenges for search user interfaces

Social search is useful but ill suited some tasks. With collaborative filtering (social voting), always a risk of just '*discovery through the rear view mirror*'

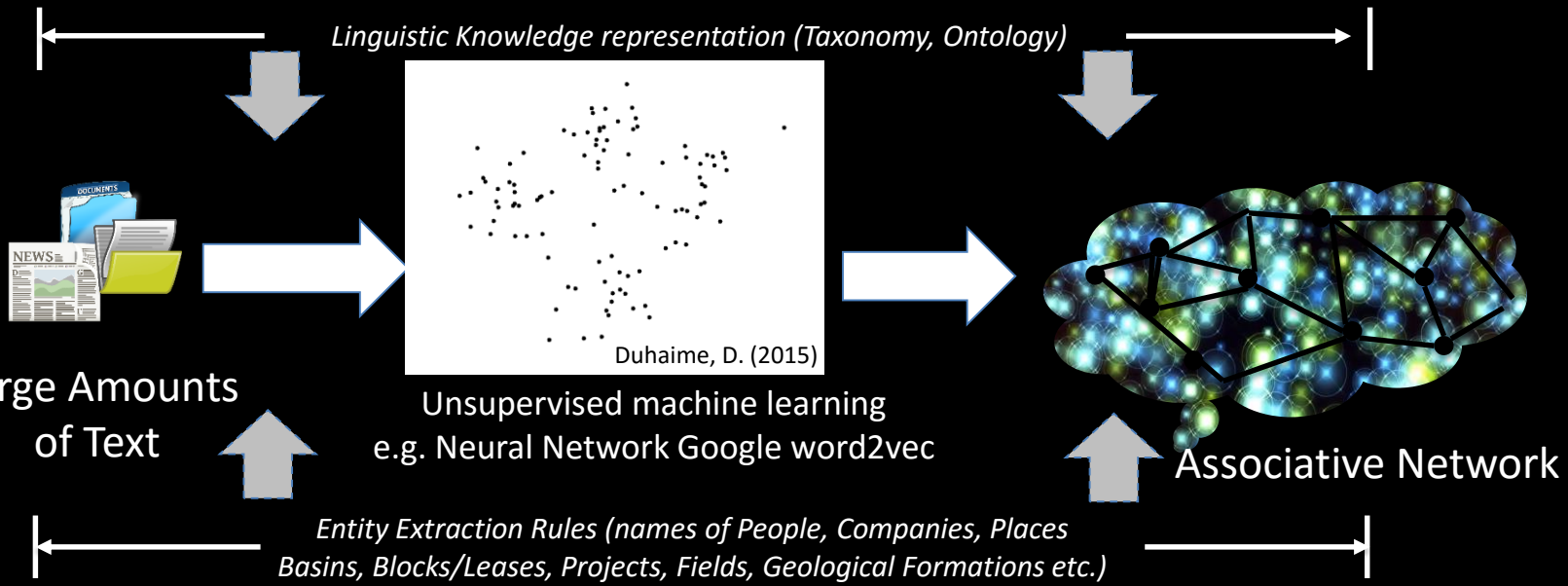
Enterprise search tends to focus on Lookup/Known Item not exploratory search. People don't complain in exploratory search *'we don't know what we don't know'*

Faceted search refiners (prompts, a form of recommendation) to filter search results tend to use manual tags. Manually added tags are often poorly added to documents (if at all). A few tags cannot represent 'the unusual' or 'contextual' buried in and across reports

Suggesting the interesting & surprising but minimize the distracting. How?

Possible solution: Text embedding's/word co-occurrence

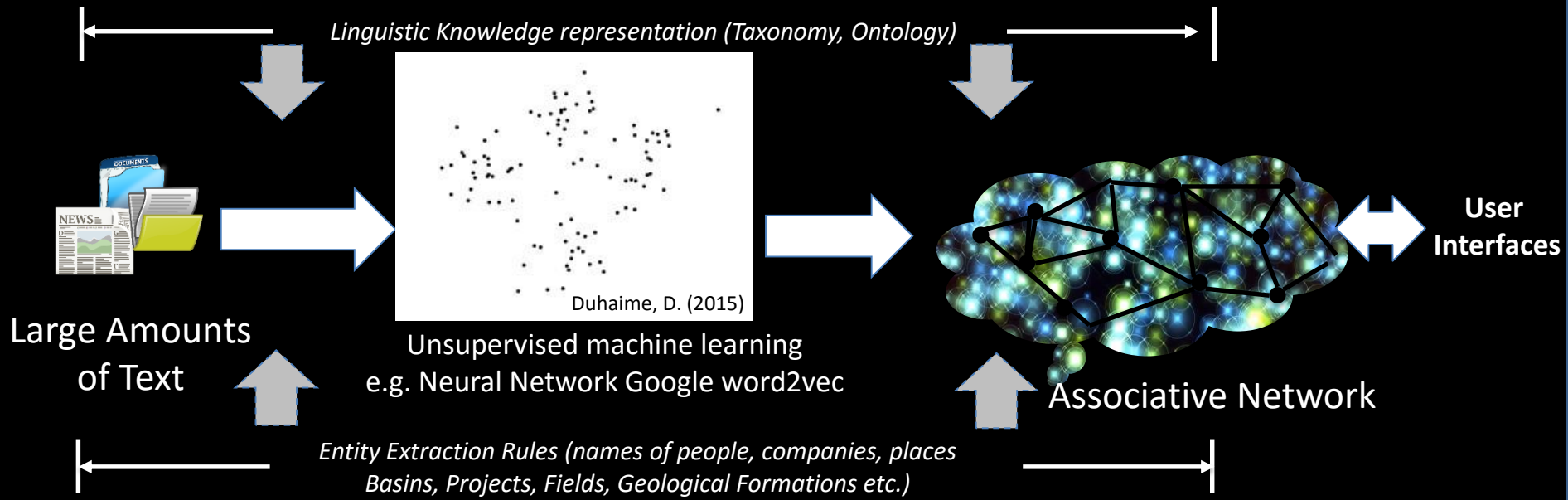
Using word co-occurrence in text, an associative network can be created through fast automated unsupervised machine learning. Where each word can be converted to numbers (vector) creating an n-dimensional vectorspace.



ROBERT GORDON UNIVERSITY ABERDEEN

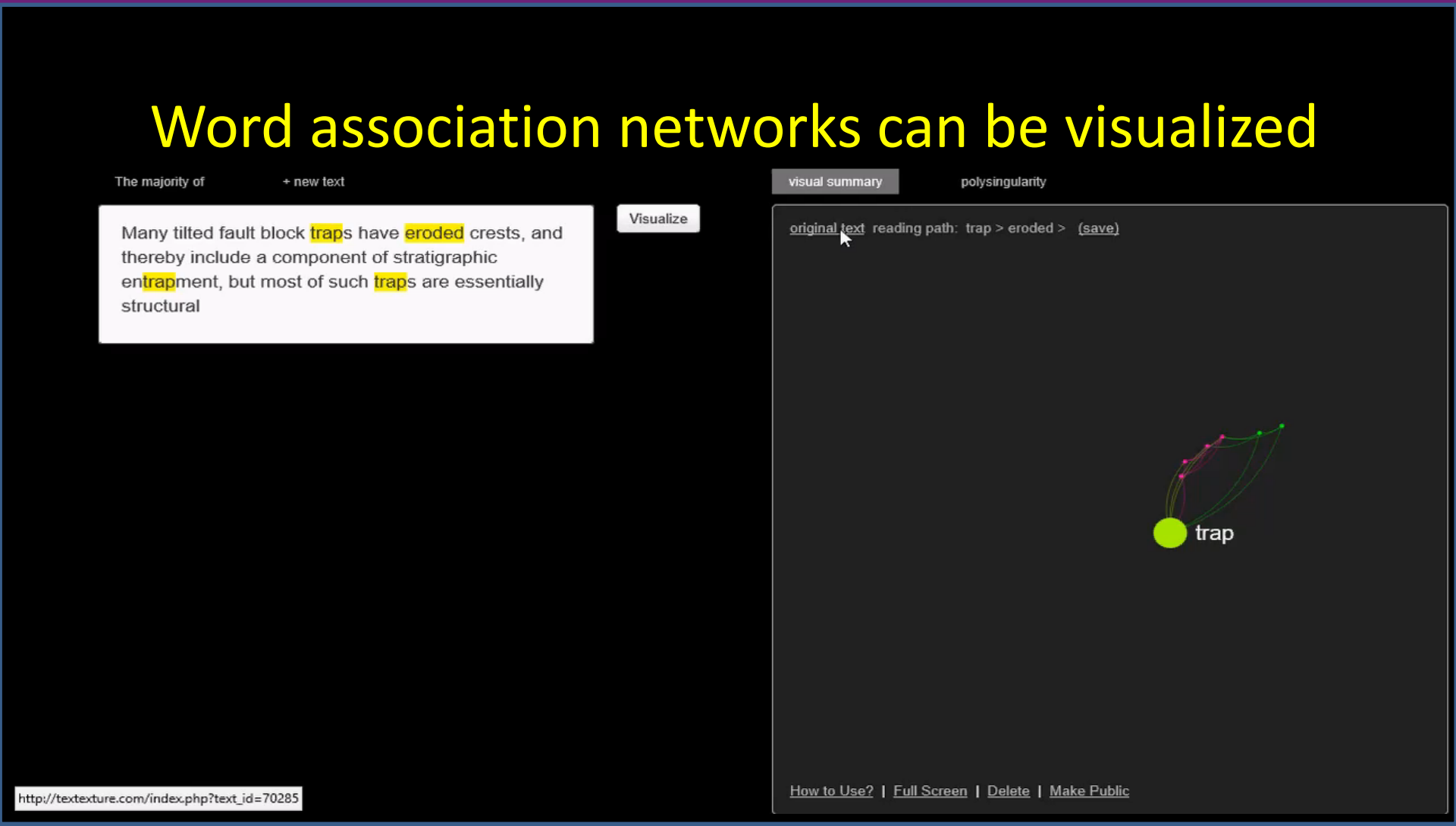
Possible solution: Text embedding's/word co-occurrence

Using word co-occurrence in text, an associative network can be created through fast automated unsupervised machine learning. Where each word can be converted to numbers (vector) creating an n-dimensional vectorspace.



ROBERT GORDON UNIVERSITY ABERDEEN

Word association networks can be visualized



ROBERT GORDON UNIVERSITY ABERDEEN

What are the information need characteristics for word co-occurrence based search filters? (54 geoscientist & engineers)

List A	List B	List C
Drilling	lost circulation	differentially
Problems	problems such	freeing
Hole	well control	spotting
Lost	poor hole	incidents
Incidents	hole instability	sticking
Well	hole cleaning	risked
Risk	drilling operations	troubles
Cost	freeing differentially	jarring
Loss	while drilling	caving
Circulation	tight hole	sloughing
…	…	…
Reduced	open hole	costly

Journal of Information Science (Cleverley and Burnett 2014)

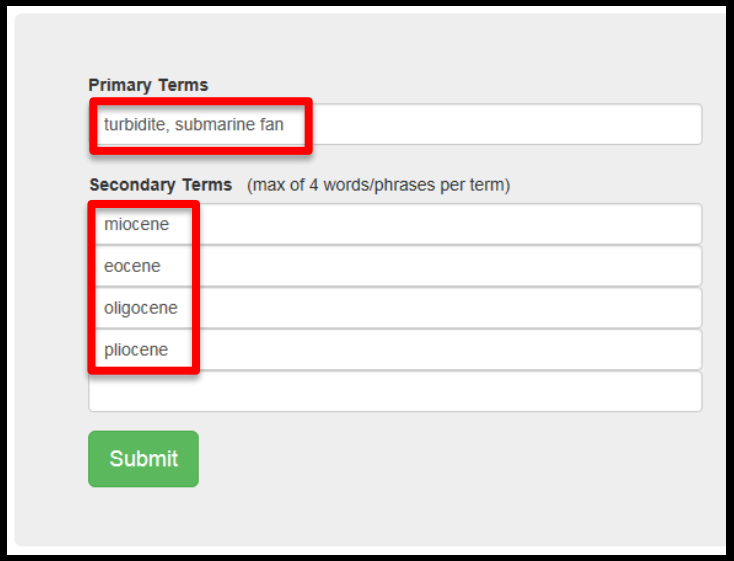
A = Unigram (Ranked by frequency of occurrence)
B = Bigram (Ranked by frequency of occurrence)
C = Unigram (Ranked by P. Mutual Information Measure)

D = Two Term Clusters (Topic Modelling)
E = Three Term Clusters (Topic Modelling)
F = Four Term Clusters (Topic Modelling)

ROBERT GORDON UNIVERSITY ABERDEEN

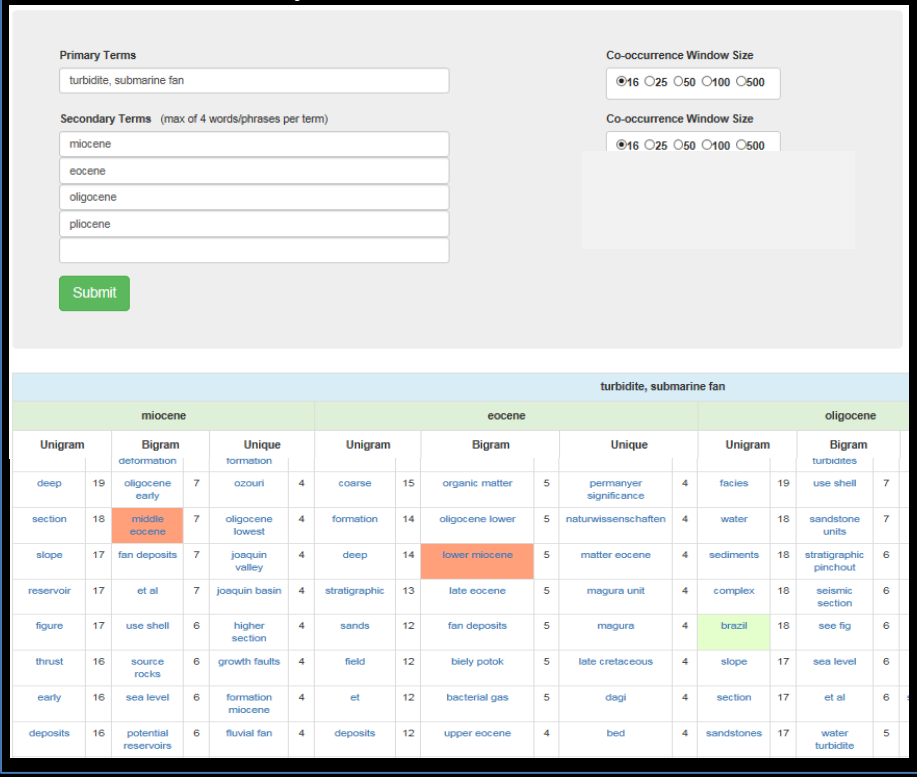
Used results in two example stimulants to date

Discriminatory



Used results in two example stimulants to date

Discriminatory



Primary Terms

turbidite, submarine fan

Co-occurrence Window Size

◉16 ○25 ○50 ○100 ○500

Secondary Terms (max of 4 words/phrases per term)

miocene
eocene
oligocene
pliocene

Co-occurrence Window Size

◉16 ○25 ○50 ○100 ○500

Submit

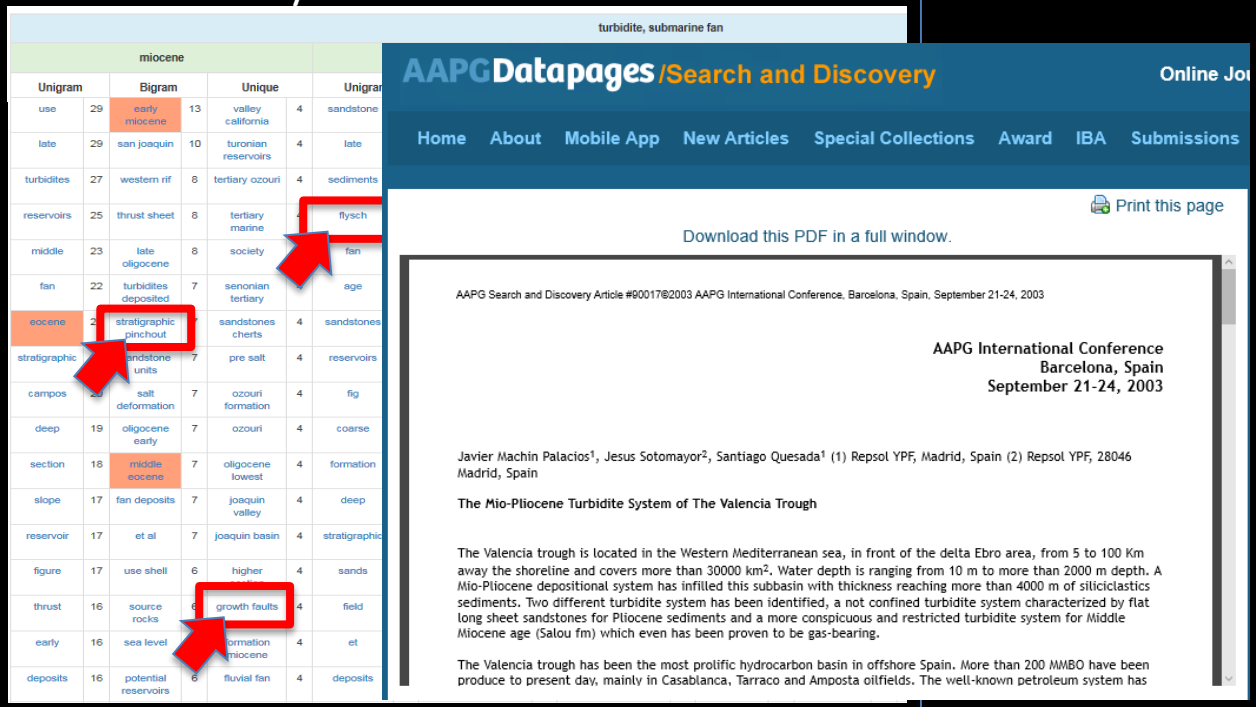
turbidite, submarine fan

miocene						eocene						oligocene			
Unigram		Bigram		Unique		Unigram		Bigram		Unique		Unigram		Bigram	
		deformation		formation										turbidites	
deep	19	oligocene early	7	ozouri	4	coarse	15	organic matter	5	permanyer significance	4	facies	19	use shell	7
section	18	middle eocene	7	oligocene lowest	4	formation	14	oligocene lower	5	naturwissenschaften	4	water	18	sandstone units	7
slope	17	fan deposits	7	joaquin valley	4	deep	14	lower miocene	5	matter eocene	4	sediments	18	stratigraphic pinchout	6
reservoir	17	et al	7	joaquin basin	4	stratigraphic	13	late eocene	5	magura unit	4	complex	18	seismic section	6
figure	17	use shell	6	higher section	4	sands	12	fan deposits	5	magura	4	brazil	18	see fig	6
thrust	16	source rocks	6	growth faults	4	field	12	biely potok	5	late cretaceous	4	slope	17	sea level	6
early	16	sea level	6	formation miocene	4	et	12	bacterial gas	5	dagi	4	section	17	et al	6
deposits	16	potential reservoirs	6	fluvial fan	4	deposits	12	upper eocene	4	bed	4	sandstones	17	water turbidite	5

ROBERT GORDON UNIVERSITY ABERDEEN

Used results in two example stimulants to date

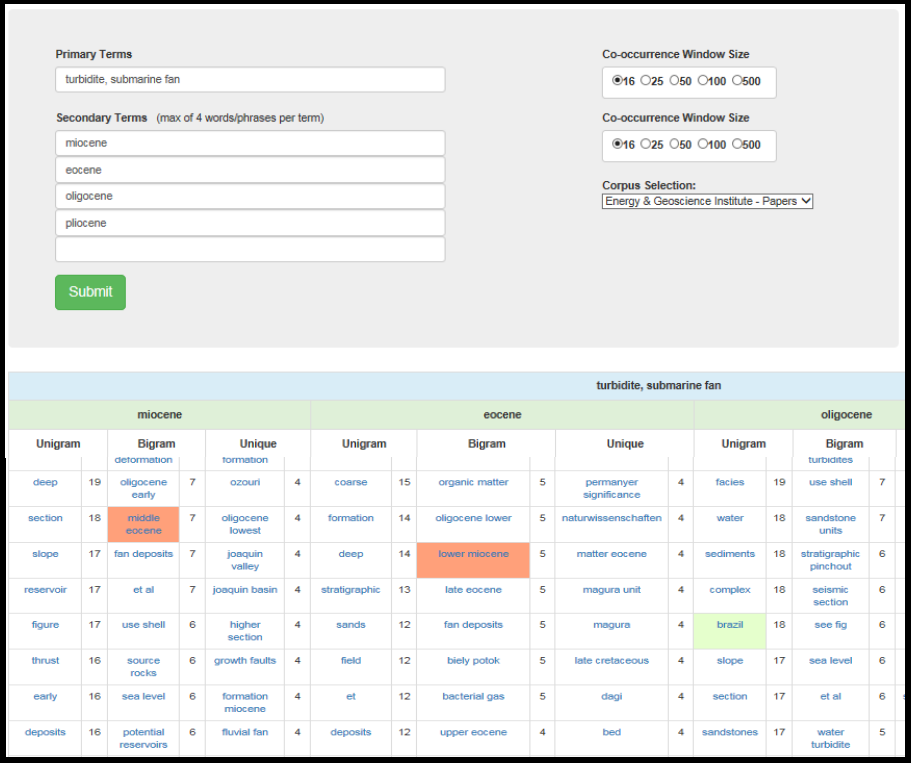
Discriminatory



ROBERT GORDON UNIVERSITY ABERDEEN

Used results in two example stimulants to date

Discriminatory



Primary Terms

turbidite, submarine fan

Secondary Terms (max of 4 words/phrases per term)

miocene

eocene

oligocene

pliocene

Co-occurrence Window Size

◉16 ○25 ○50 ○100 ○500

Co-occurrence Window Size

◉16 ○25 ○50 ○100 ○500

Corpus Selection:

Energy & Geoscience Institute - Papers

Submit

turbidite, submarine fan

miocene						eocene						oligocene					
Unigram		Bigram		Unique		Unigram		Bigram		Unique		Unigram		Bigram			
		deformation		formation										turbidites			
deep	19	oligocene early	7	ozouri	4	coarse	15	organic matter	5	permanyer significance	4	facies	19	use shell	7		
section	18	middle eocene	7	oligocene lowest	4	formation	14	oligocene lower	5	naturwissenschaften	4	water	18	sandstone units	7		
slope	17	fan deposits	7	joaquin valley	4	deep	14	lower miocene	5	matter eocene	4	sediments	18	stratigraphic pinchout	6		
reservoir	17	et al	7	joaquin basin	4	stratigraphic	13	late eocene	5	magura unit	4	complex	18	seismic section	6		
figure	17	use shell	6	higher section	4	sands	12	fan deposits	5	magura	4	brazil	18	see fig	6		
thrust	16	source rocks	6	growth faults	4	field	12	biely potok	5	late cretaceous	4	slope	17	sea level	6		
early	16	sea level	6	formation miocene	4	et	12	bacterial gas	5	dagi	4	section	17	et al	6		
deposits	16	potential reservoirs	6	fluvial fan	4	deposits	12	upper eocene	4	bed	4	sandstones	17	water turbidite	5		

Similarity

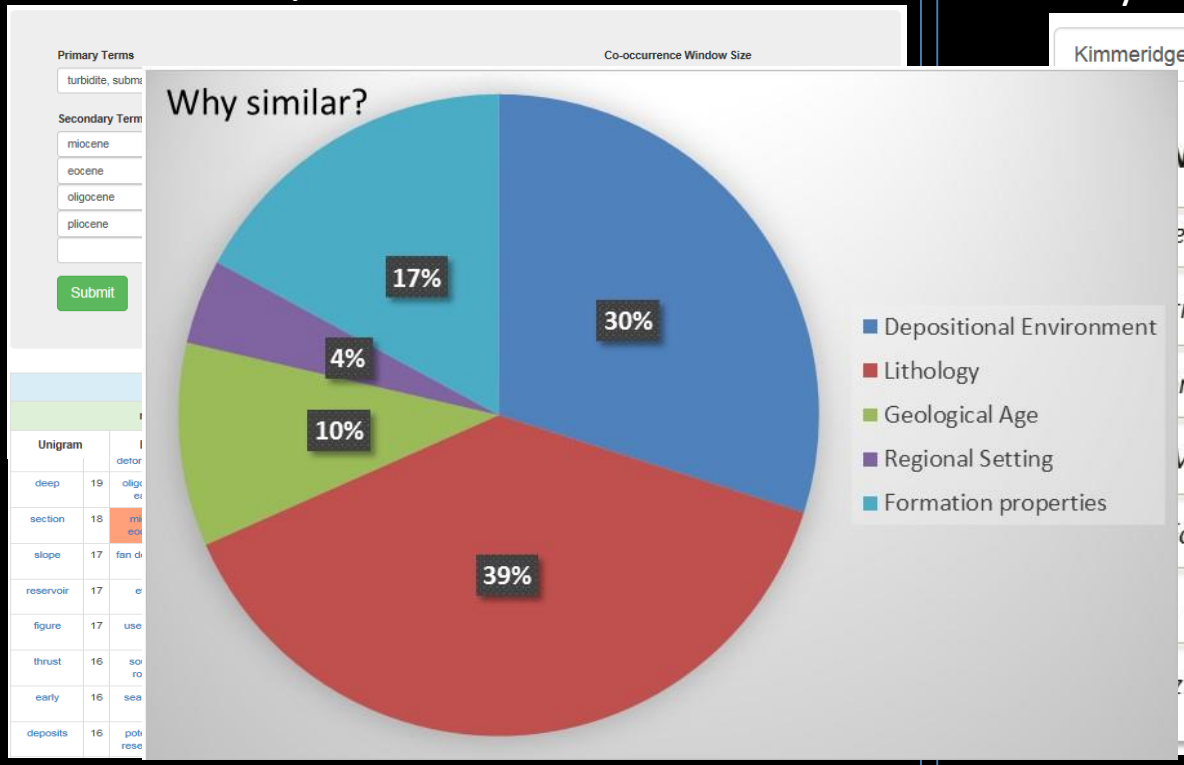
Kimmeridge Clay Formation,

Find Similar

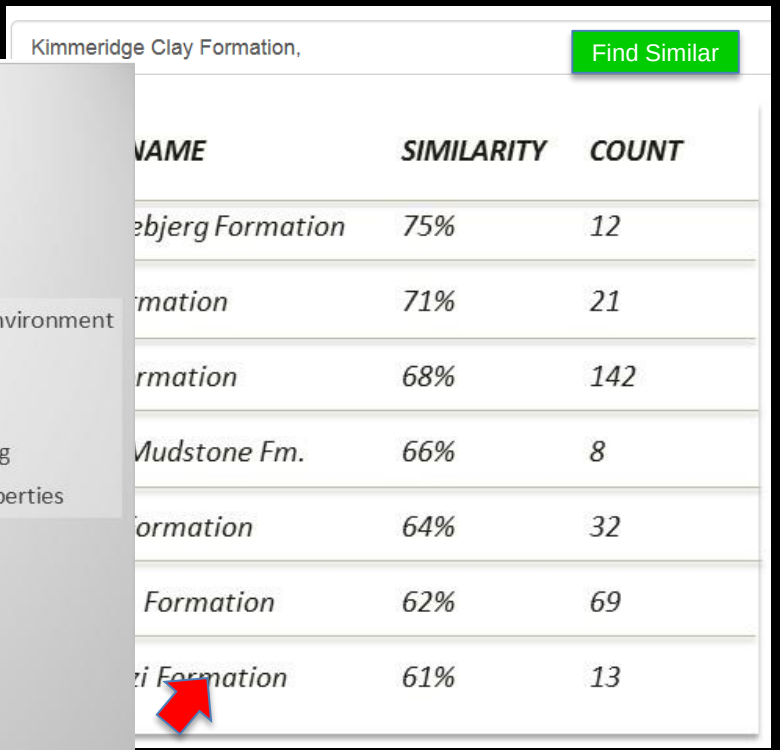
ROBERT GORDON UNIVERSITY ABERDEEN

Used results in two example stimulants to date

Discriminatory



Similarity

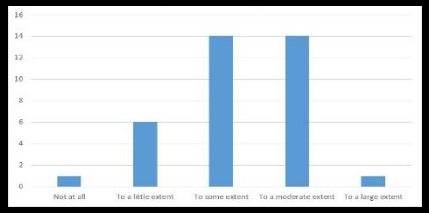


ROBERT GORDON UNIVERSITY ABERDEEN

To what extent can word co-occurrence facilitate serendipity?

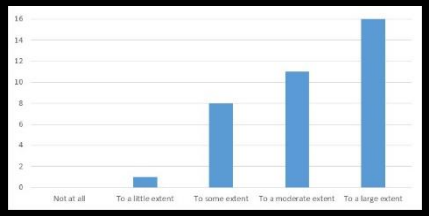
(53 geoscientists in two oil and gas companies)

Journal of Information & Knowledge Management (Cleverley and Burnett 2015)



To what extent do current search interfaces in your organization facilitate serendipitous discovery?

41% - To a moderate/large extent



To what extent could word co-occurrence techniques facilitate serendipitous discovery?

73% - To a moderate/large extent

ROBERT GORDON UNIVERSITY ABERDEEN

Summary: Opportunity to move ‘search’ from *just* a *time-saving* utility to a *value adding* creative assistant

“Using these techniques, word associations highlighted new and unexpected terms …. This surprising result led us to consider a new geological element which could impact our opportunity”

New Venture Geologist
Multinational Oil & Gas Company

“The research has direct implications for many of our own studies at NASA including the efficacy of enterprise search..
The concept of ‘Facilitating Serendipity’ offers an exemplary model .. that can be communicated with groups and individuals at many levels.”

Office of the Chief Knowledge Officer (CKO)
NASA Johnson Space Centre (JSC)

Email: p.h.cleverley@rgu.ac.uk Blog: www.paulhcleverley.com

ROBERT GORDON UNIVERSITY ABERDEEN