



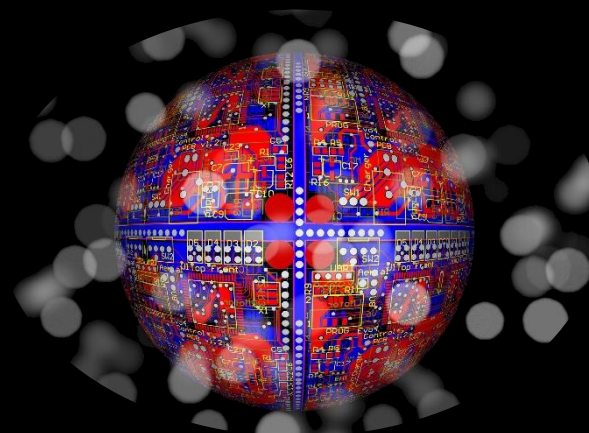
**ROBERT GORDON  
UNIVERSITY ABERDEEN**

## **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**



# AAPG EXPLORER

[HOME](#) / [EXPLORER](#) / [Archive](#) / [April 2010](#) / Serendipity Is Explorationists' Friend

Mobile Bay a product of good fortune

Serendipity Is Explorationists' Friend

By BARRY FRIEDMAN, EXPLORER C

Like the size of fish and the exponential story of the catch, the historical account of themselves to hyperbole.

In 1979, when MOEPSI, a wholly owned natural gas field in Alabama's lower Mobile Bay, was discovered, it was a combination of luck, sound technical analysis and environmental hurdles.

"I worked for Mobil for 37 years on five continents. I was exploration manager for the company. Mobil's big fields were found as a result of serendipity."

In the case of Mobile Bay, serendipity n

HARTENERGY

Expand ▾



REGIONS +

OFFSHORE +

SUBSEA +

EXPLORATION +

PRODUCTION +

UNCONVENTIONAL +

TECHNOLOGY +

MAGAZINE

Put Mobil™ technology on your side.

Energy lives here™

Learn more

Serendipity is alive and well at Eagle Ford

*"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**

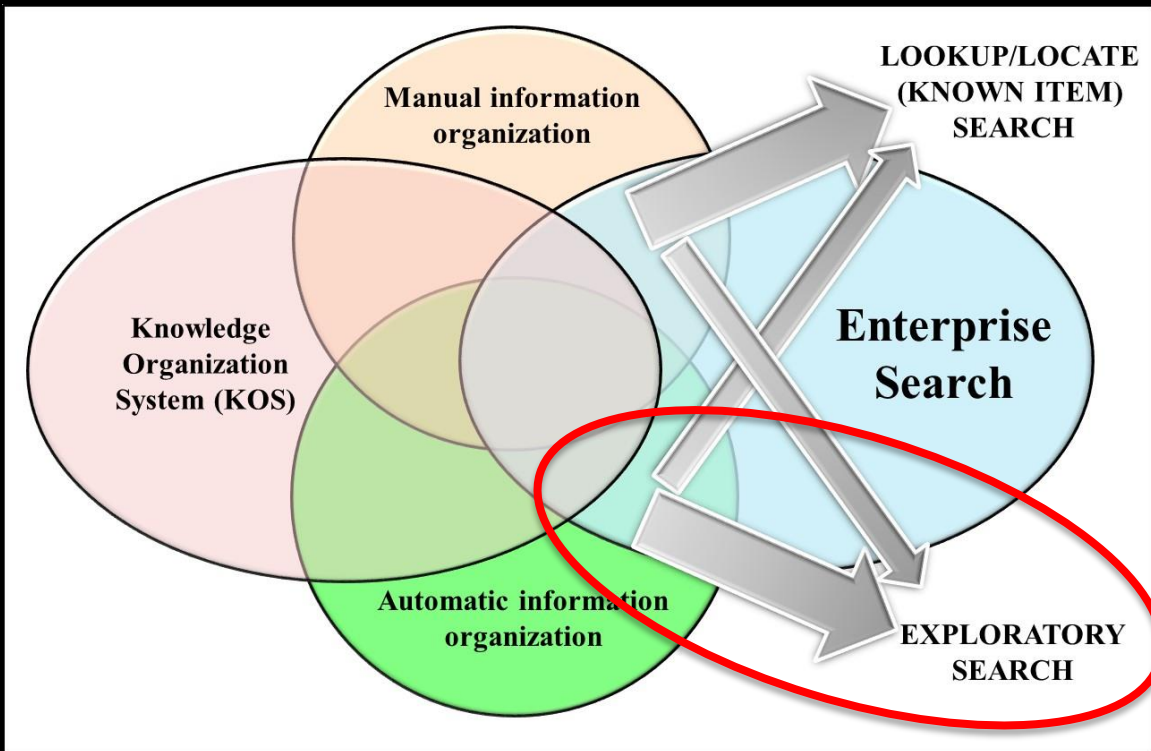
**Immersion in Information-rich  
spaces**

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



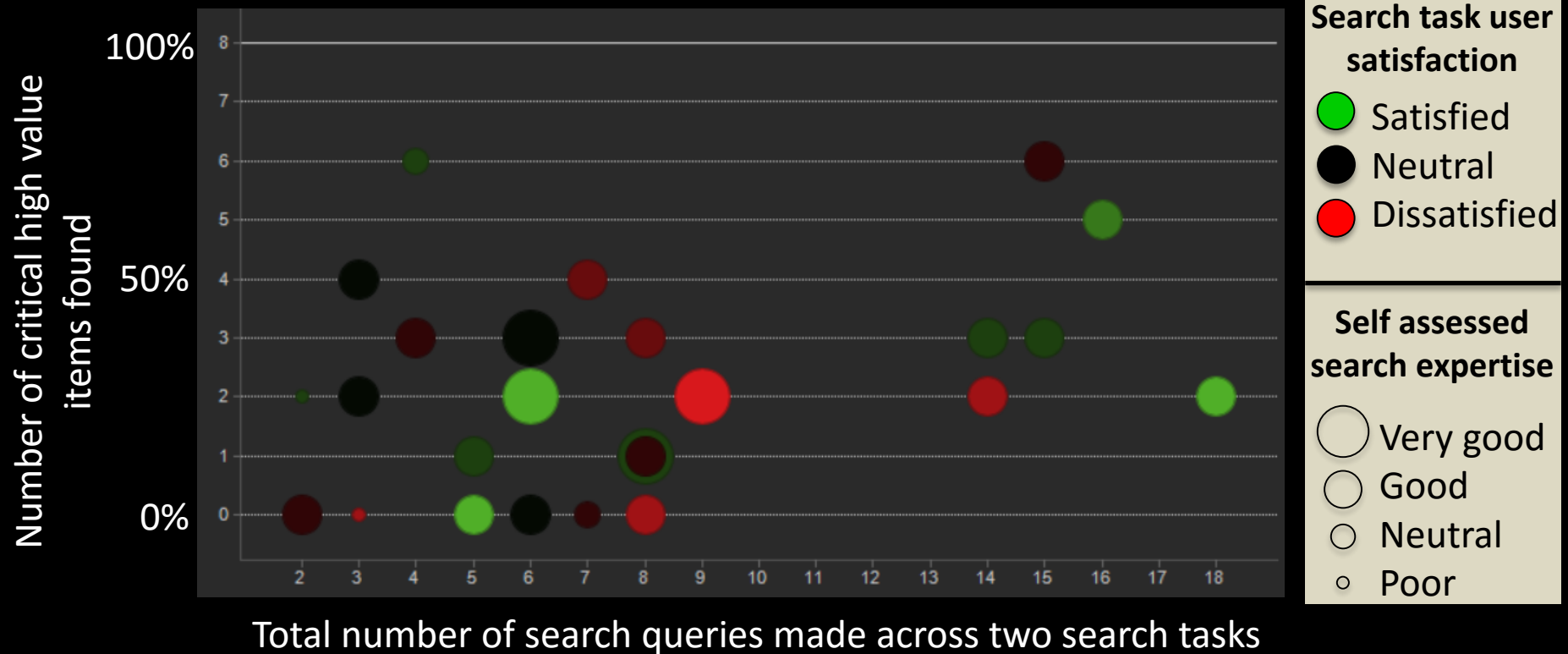
**ROBERT GORDON  
UNIVERSITY ABERDEEN**

# 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)*

# 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



**ROBERT GORDON  
UNIVERSITY ABERDEEN**



## 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'*

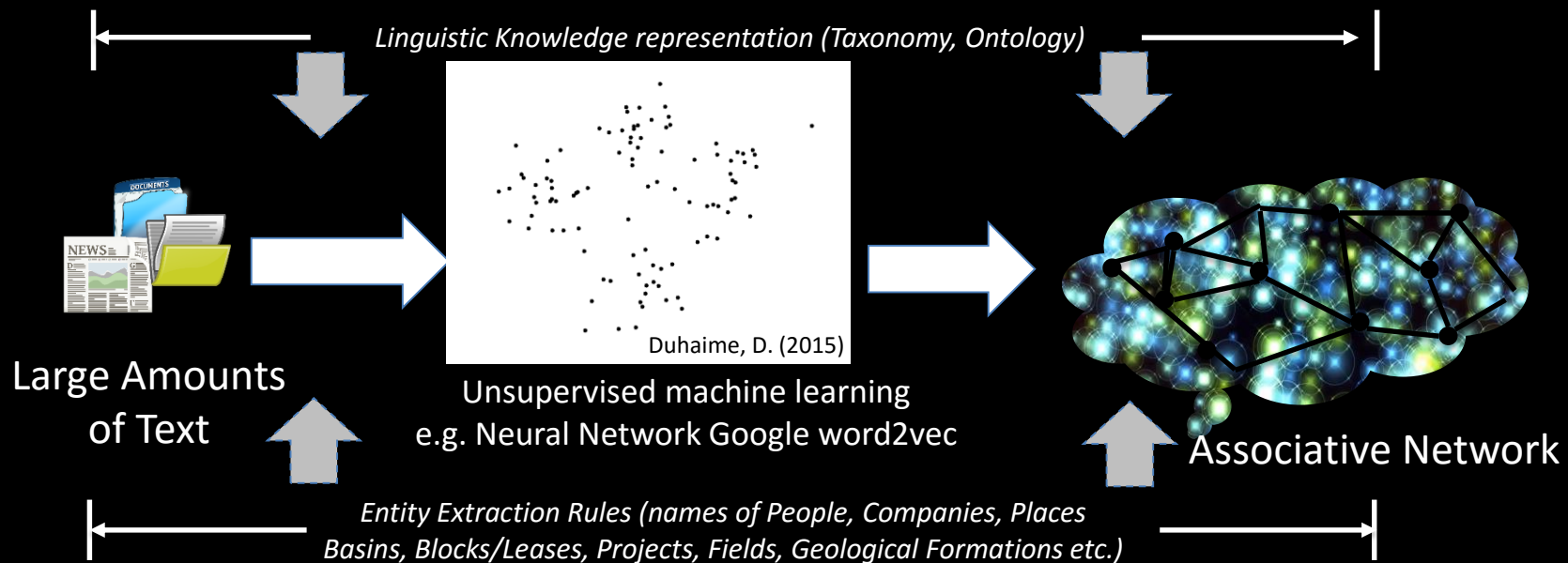
Faceted search refiners (prompts, a form of recommendation) to filter search results tend to use manual tags. Manually added tags are often poorly added to documents (if at all). A few tags cannot represent 'the unusual' or 'contextual' buried in and across reports

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



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

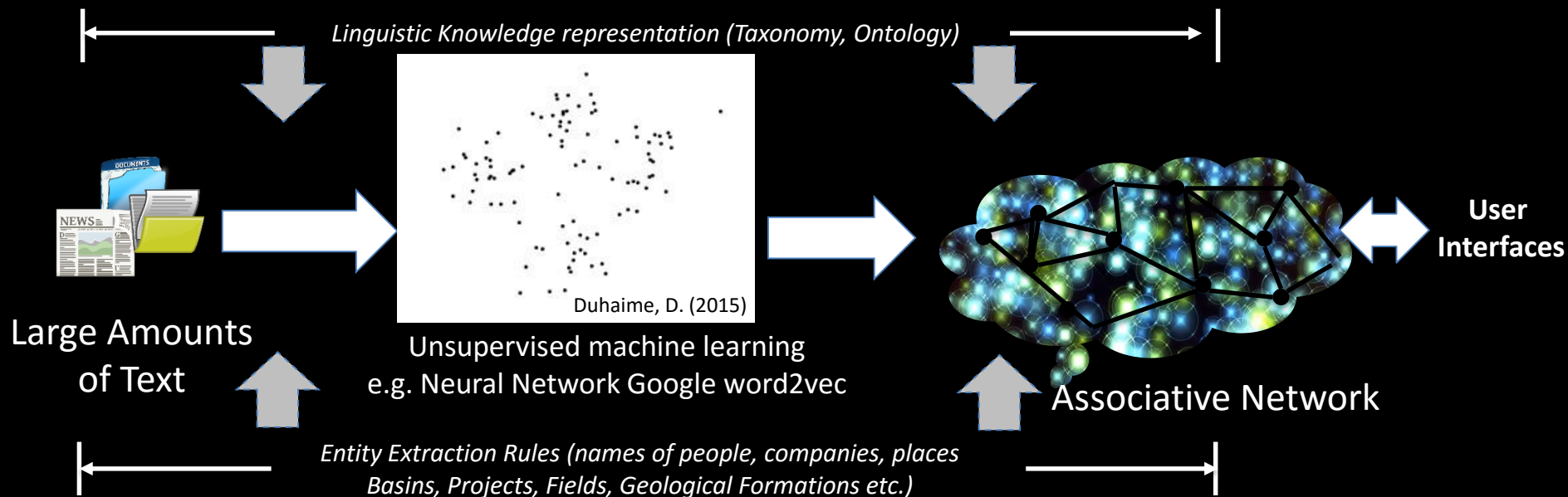
Using word co-occurrence in text, an associative network can be created through fast automated unsupervised machine learning. Where each word can be converted to numbers (vector) creating an n-dimensional vectorspace.





# 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.



# Word association networks can be visualized

The majority of

+ new text

Many tilted fault block **traps** have **eroded** crests, and thereby include a component of stratigraphic entrapment, but most of such **traps** are essentially structural

Visualize

visual summary

polysingularity

original text reading path: trap > eroded > (save)



[How to Use?](#) | [Full Screen](#) | [Delete](#) | [Make Public](#)

[http://texttexture.com/index.php?text\\_id=70285](http://texttexture.com/index.php?text_id=70285)



**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

**Primary Terms**

turbidite, submarine fan

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

miocene

eocone

oligocene

pliocene

Submit



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



# Used results in two example stimulants to date

## Discriminatory

turbidite, submarine fan

| miocene       |                          |                        |         |               |
|---------------|--------------------------|------------------------|---------|---------------|
| Unigram       | Bigram                   | Unique                 | Unigram | Unigram       |
| use           | 29 early miocene         | 13 valley california   | 4       | sandstone     |
| late          | 29 san joaquin           | 10 turonian reservoirs | 4       | late          |
| turbidites    | 27 western rif           | 8 tertiary ozouri      | 4       | sediments     |
| reservoirs    | 25 thrust sheet          | 8 tertiary marine      | 4       | flysch        |
| middle        | 23 late oligocene        | 8 society              | 4       | fan           |
| fan           | 22 turbidites deposited  | 7 senonian tertiary    | 4       | age           |
| eocene        | 2 stratigraphic pinchout | 4 sandstones cherts    | 4       | sandstones    |
| stratigraphic | 2 sandstone units        | 7 pre salt             | 4       | reservoirs    |
| campos        | 23 salt deformation      | 7 ozouri formation     | 4       | fig           |
| deep          | 19 oligocene early       | 7 ozouri               | 4       | coarse        |
| section       | 18 middle eocene         | 7 oligocene lowest     | 4       | formation     |
| slope         | 17 fan deposits          | 7 joaquin valley       | 4       | deep          |
| reservoir     | 17 et al                 | 7 joaquin basin        | 4       | stratigraphic |
| figure        | 17 use shell             | 6 higher               | 4       | sands         |
| thrust        | 16 source rocks          | 6 growth faults        | 4       | field         |
| early         | 16 sea level             | 6 formation miocene    | 4       | et            |
| deposits      | 16 potential reservoirs  | 6 fluvial fan          | 4       | deposits      |

**AAPG Datapages / Search and Discovery**
Online Jo

[Home](#) [About](#) [Mobile App](#) [New Articles](#) [Special Collections](#) [Award](#) [IBA](#) [Submissions](#)

Print this page

[Download this PDF in a full window.](#)

AAPG Search and Discovery Article #90017@2003 AAPG International Conference, Barcelona, Spain, September 21-24, 2003

**AAPG International Conference**  
**Barcelona, Spain**  
**September 21-24, 2003**

Javier Machin Palacios<sup>1</sup>, Jesus Sotomayor<sup>2</sup>, Santiago Quesada<sup>1</sup> (1) Repsol YPF, Madrid, Spain (2) Repsol YPF, 28046 Madrid, Spain

**The Mio-Pliocene Turbidite System of The Valencia Trough**

The Valencia trough is located in the Western Mediterranean sea, in front of the delta Ebro area, from 5 to 100 Km away the shoreline and covers more than 30000 km<sup>2</sup>. Water depth is ranging from 10 m to more than 2000 m depth. A Mio-Pliocene depositional system has infilled this subbasin with thickness reaching more than 4000 m of siliciclastics sediments. Two different turbidite system has been identified, a not confined turbidite system characterized by flat long sheet sandstones for Pliocene sediments and a more conspicuous and restricted turbidite system for Middle Miocene age (Salou fm) which even has been proven to be gas-bearing.

The Valencia trough has been the most prolific hydrocarbon basin in offshore Spain. More than 200 MMBO have been produce to present day, mainly in Casablanca, Tarraco and Amposta oilfields. The well-known petroleum system has



# 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

Submit

### Co-occurrence Window Size

☒16 ☐25 ☐50 ☐100 ☐500

### Co-occurrence Window Size

☒16 ☐25 ☐50 ☐100 ☐500

### Corpus Selection:

Energy & Geoscience Institute - Papers

### turbidite, submarine fan

| miocene   |        |                      |         |                   |        |               |        |                 |         |                        |        |
|-----------|--------|----------------------|---------|-------------------|--------|---------------|--------|-----------------|---------|------------------------|--------|
| Unigram   | Bigram | Unique               | Unigram | Bigram            | Unique | Unigram       | Bigram | Unique          | Unigram | Bigram                 | Unique |
| deep      | 19     | oligocene early      | 7       | ozouri            | 4      | coarse        | 15     | organic matter  | 5       | permaryer significance | 4      |
| section   | 18     | middle eocene        | 7       | oligocene lowest  | 4      | formation     | 14     | oligocene lower | 5       | naturwissenschaften    | 4      |
| slope     | 17     | fan deposits         | 7       | joaquin valley    | 4      | deep          | 14     | lower miocene   | 5       | matter eocene          | 4      |
| reservoir | 17     | et al                | 7       | joaquin basin     | 4      | stratigraphic | 13     | late eocene     | 5       | magura unit            | 4      |
| figure    | 17     | use shell            | 6       | higher section    | 4      | sands         | 12     | fan deposits    | 5       | megura                 | 4      |
| thrust    | 16     | source rocks         | 6       | growth faults     | 4      | field         | 12     | biely potok     | 5       | late cretaceous        | 4      |
| early     | 16     | sea level            | 6       | formation miocene | 4      | et            | 12     | bacterial gas   | 5       | dagl                   | 4      |
| deposits  | 16     | potential reservoirs | 6       | fluvial fan       | 4      | deposits      | 12     | upper eocene    | 4       | bed                    | 4      |

## Similarity

Kimmeridge Clay Formation,

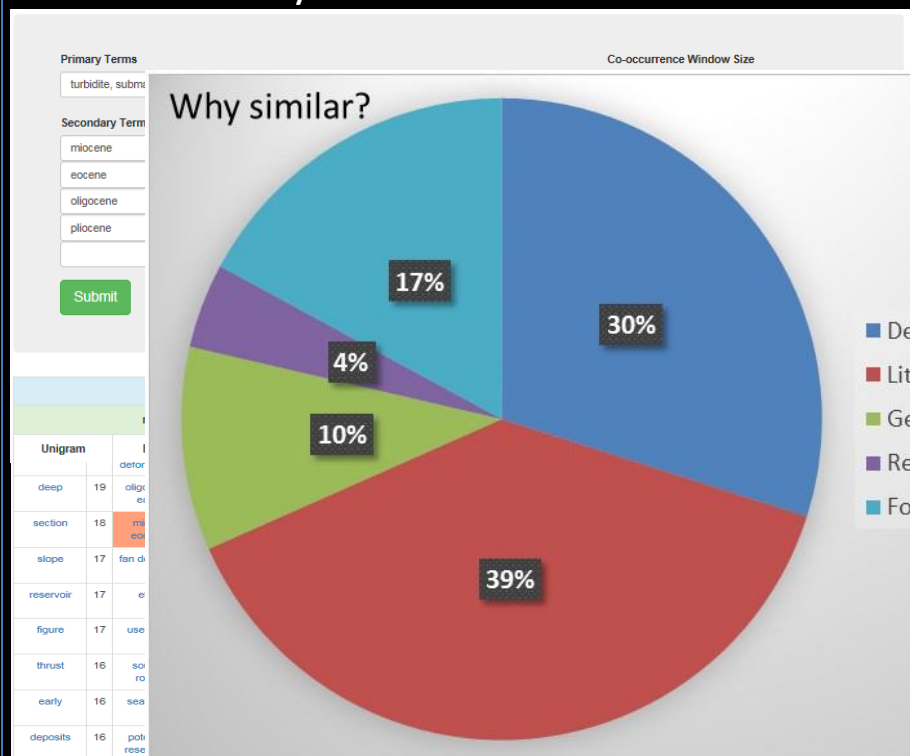
Find Similar



ROBERT GORDON  
UNIVERSITY ABERDEEN

# Used results in two example stimulants to date

## Discriminatory



## Similarity

Kimmeridge Clay Formation,

Find Similar

| NAME             | SIMILARITY | COUNT |
|------------------|------------|-------|
| ebjerg Formation | 75%        | 12    |
| formation        | 71%        | 21    |
| formation        | 68%        | 142   |
| Mudstone Fm.     | 66%        | 8     |
| formation        | 64%        | 32    |
| Formation        | 62%        | 69    |
| Formation        | 61%        | 13    |

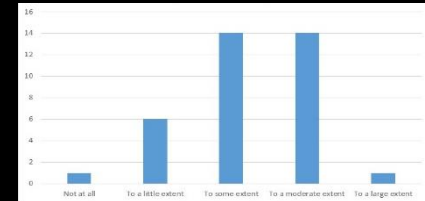




# To what extent can word co-occurrence facilitate serendipity? (53 geoscientists in two oil and gas companies)

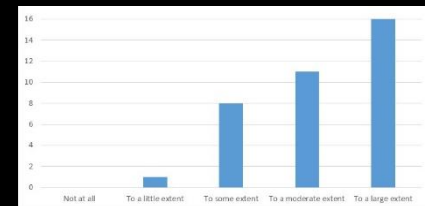


*Journal of Information & Knowledge Management (Cleverley and Burnett 2015)*



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](mailto:p.h.cleverley@rgu.ac.uk) Blog: [www.paulhcleverley.com](http://www.paulhcleverley.com)



**ROBERT GORDON  
UNIVERSITY ABERDEEN**