

# Facilitating serendipity in the enterprise search and digital library user interface.

Paul H. Cleverley and Simon Burnett

Department of Information Management Aberdeen Business School Robert Gordon University (RGU), United Kingdom

### Overall PhD Research Topic: What are the causal factors for user satisfaction and suboptimal search task performance in the Enterprise?

#### Include common 'event' themes like:

- Poor search results
- Poor information quality / completeness
- Slow, poor user interface



### Overall PhD Research Topic: What are the causal factors for user satisfaction and suboptimal search task performance in the Enterprise?

#### Include common 'event' themes like:

- Poor search results
- Poor information quality / completeness
- Slow, poor user interface

### And deeper 'systems thinking' themes like:

- Expectations driven by the Internet
- Search expertise, Lack of feedback loops
- Poor Information Management governance
- Attitudes of senior management



Overall PhD Research Topic: What are the causal factors for user satisfaction and suboptimal search task performance in the Enterprise?

Sub-optimal performance could also include the <u>extent to which</u> serendipitous opportunities are stimulated in the search user interface?



# Serendipity the "Happy Accident"

- Fortuitous information encountering (not random)
- Serendipity Space subjective.

|            | Slightly | Moderately | Very |
|------------|----------|------------|------|
| Unexpected |          |            |      |
| Insightful |          |            |      |
| Valuable   |          |            |      |

(Makri and Blandford 2012)







# Background

- Enterprise search & digital library interfaces focus on precision.
- Many researchers feel search interfaces may be constraining the chances of insightful and serendipitous discovery.
- Faceted search increasingly popular as a browsing technique.



## **Motivation for research**

Knowledge organization (ontologies, taxonomies, thesauri) of value to improve precision, recall aid integration and serendipity.

#### BUT

Categories displayed in faceted search that are representative of an information item, rarely convey unexpected or non-obvious contextually associated concepts.



Office of Chief Knowledge Officer (CKO) Johnson Space Centre, NASA Houston



Johnson Space Center

### Word co-occurrence as faceted search filters

Using co-occurring words (to search terms) in the body text of documents may yield more intriguing options.

Word co-occurrence has been used extensively in tag clouds and as filters

#### BUT

Typically only the most popular/statistically common terms are used.

| Audio BBC                     | Case Stu                                       | dies Copyright   | Creative C              | ommons Critical       |
|-------------------------------|--|--|-------------------------|-----------------------|
| Thinking                      | Digital Cam                                    | eras Digital   | Exclusion               | Digital Footprint     |
| Digital I                     | dentity E                                      | Digital Litera   | CY Digital 1            | Media Digital Natives |
| Digital                       | Skills D                                       | igital Tools   | E-books                 | Seolocation Google+   |
| Higher                        | Educa  | tion I-pa  | d INFOG                 | RAPHIC Images         |
| Informat                      | tion Litera                                    | CY Internet 8  | skills JISC             | Learner Checklist     |
| Learner V                     | oice Mob                                       | ile Learning   | News OU                 | PADDLE Reports        |
| affair war                    | Your re  | sults are grouped by fou   | ind editions. Show      | all results velopment |
| nite storey<br>nature tale    | 1. Love t<br>Cowburn                           | oy John Cowburn. (electr<br>, John.                              | onic resource] /        | foursquare            |
| death<br>art                  | Bectronic Vear:<br>Resource Series:<br>Subject | 2003.<br>Marquette studies in ph                                 | losophy;                | cial networking       |
| book love woman               | Table of                                       | books.<br>Contents:  | aspects crinistianity , | CIECTONIC             |
| way letter                    |  | Pt. 1 Self-love and Lov<br>2 Self-love<br>Pt. 2 Solidarity-Love  | ve in General           |                       |
| faith village                 | Found  | Love (23) loves (3)<br>Available Online<br>OSU-Tulsa, Click HERE | to connect to this elec | stronic bool          |
| essay make                    | OSU-Tuk  | a BD436 C598 2003e   | b Check for Availab     | olity                 |
| glove liking                  | 2. Love :<br>Dilman, It                        | its forms, dimensions, ai<br>iam.                                | nd paradoxes / liham l  | Déman.                |
| Association Spelling variatio | Books Year<br>Subject                          | 1998.<br>Love.   |                         |                       |
| Discovery trail               | 1 anis 01 (                                    | 2 Love and Hate: Are   | They Opposites?         |                       |



## Introduction - search term word co-occurrence

• E.g. search term 'metadata' window = sentence.

We, and others we spoke with, have had mixed results from using technologies such as document management systems for more than point solutions to particular problems. Users resisted moving files to these systems for three primary reasons. First, these systems typically required that users enter metadata information about the file such as a description of it, the author's name, and some keywords for categorization. Any capture of metadata that added more than about ten seconds to saving a file was considered problematic. Second, these systems typically do not work with files that are linked such as many spreadsheets, presentations and documents. Third, many technical applications required that their files be on a LAN or local storage device and do not work with most document management technologies.

Improvements in the management of structured and unstructured data (Garbarini et al. 2008)



## Introduction - search term word co-occurrence

• E.g. search term 'metadata' window = sentence.

We, and others we spoke with, have had mixed results from using technologies such as document management systems for more than point solutions to particular problems. Users resisted moving files to these systems for three primary reasons. First, these systems typically required that users enter metadata information about the file such as a description of it, the author's name, and some keywords for categorization. Any capture of metadata that added more than about ten seconds to saving a file was considered problematic. Second, these systems typically do not work with files that are linked such as many spreadsheets, presentations and documents. Third, many technical applications required that their files be on a LAN or local storage device and do not work with most document management technologies.

Improvements in the management of structured and unstructured data (Garbarini et al. 2008)



### Introduction - search term word co-occurrence

- E.g. search term 'metadata' window = sentence.
- Generates vast amounts of data. How do we select what to display?

We, and others we spoke with, have had mixed results from using technologies such as document management systems for more than point solutions to particular problems. Users resisted moving files to these systems for three primary reasons. First, these systems typically required that users enter metadata information about the file such as a description of it, the author's name, and some keywords for categorization. Any capture of metadata that added more than about ten seconds to saving a file was considered problematic. Second, these systems typically do not work with files that are linked such as many spreadsheets, presentations and documents. Third, many technical applications required that their files be on a LAN or local storage device and do not work with most document management technologies.

Improvements in the management of structured and unstructured data (Garbarini et al. 2008)



#### For the query: Permeability

| SPE (By fi | requency)  | c | GSL (By fr | equency)  | SPE | (by MIM)         | GSL ( | By MIM)        |
|------------|------------|---|------------|-----------|-----|------------------|-------|----------------|
| OC         | c1         | 0 | C          | c1        | OC  | c2 (PMI)         | 0C    | c2 (PMI)       |
| S          | relative   |   | Pr         | porosity  | 5   | disproportionate | RI    | lens           |
| RI         | reservoir  |   | RI         | reservoir | Rep | Millidarcy       | Rep   | functions      |
| Pr         | porosity   |   | S          | low       | Rep | tensors          | В     | reach          |
| 5          | low        |   | Ph         | fault     | Rep | darcies          | RI    | matrix         |
| M          | well       |   | S          | high      | 5   | pseudorelative   | Rep   | md             |
| Ph         | fracture   |   | S          | vertical  | M   | permeameter      | В     | contrasts      |
| RI         | formation  |   | M          | sandstone | M   | liquor           | S     | transient      |
| Pr         | pressure   |   | Ri         | matrix    | M   | modifiers        | RI    | host           |
| S          | high       |   | Rep        | model     | Rep | klinkenberg      | Proc  | entry          |
| M          | oil        |   | Pr         | pressure  | 5   | relative         | S     | pseudorelative |
| M          | gas        |   | M          | rock      | 5   | absolute         | Rep   | ranges         |
| M          | water      |   | S          | effective | S   | uncored          | Rep   | cutoff         |
| H          | measured   |   | Ph         | flow      | RI  | streaks          | Rep   | coefficient    |
| Rep        | data       |   | Ph         | capillary | Pr  | porosity         | Proc  | trapping       |
| Ph         | flow       |   | M          | gas       | в   | contrasts        | M     | Core           |
| M          | sand       |   | S          | relative  | Pr  | anisotropy       | Ph    | capillary      |
| S          | effective  |   | H          | methods   | Rep | kr               | B     | confirm        |
| M          | core       |   | M          | well      | Ph  | hysteresis       | H     | follow         |
| Pr         | saturation |   | Rep        | md        | Rep | stoneley         | Pr    | macroporosity  |

### ROBERT GORDON UNIVERSITY ABERDEEN





#### Discriminatory word co-occurrence: Primary and secondary queries





#### Discriminatory word co-occurrence: Primary and secondary queries





## Research aims

- To what extent can comparison of search results using discriminatory search term word co-occurrence facilitate serendipity in the enterprise?
- Why does this happen?
- When would these techniques be of most value to an organization?



## Methodology – Creation of the stimuli

1. 100,000+ Scientific Literature Articles **Society of Petroleum Engineers Literature Geological Society of London Literature** 5. Produce Semi-Interactive Stimuli **American Geological Institute Literature** colour coded displayed inspired by DNA visualizations F-1 MALE 2. Sample Initial **Queries and** 4. Colour coding to a high Secondary comparison 3. Word co-occurrence Level taxonomy queries **Algorithms (Python)** 



#### Discriminatory word co-occurrence: Primary and secondary queries







Your search for seismic AND "gulf of mexico" AND magnetotelluric has returned 32 results in 0.103 seconds

| seismic   |  | Filter your results by Publication date: All dates Journal or conference: All publications | Sort t | Items per page:       10         Mark all       Add marked to cart         Marine Magnetotelluric (MMT) Data Interpretation in the Gulf of Mexico for Subsalt Imaging   |  |
|---|--|--|--------|---|--|
| Algorithm A<br>data seis<br>3D 3D:<br>reservoir tim<br>well seis<br>time-lapse seis<br>amplitude seis<br>interpretation 4D:<br>velocity seis  | Gulf of N<br>Algorithm B<br>smic data<br>seismic<br>te-lapse seismic<br>smic amplitude<br>smic surveys<br>smic reflection<br>seismic<br>smic response  | Publisher: All publishers Type: All types  |        | Sandberg, Stewart, WesternGeco         Roper, Tom, WesternGeco Electromagnetics         Campbell, Tracy, WesternGeco Electromagnetics         19659-MS OTC Conference Paper - 2008         View rights & permissions         ☆ ☆ ☆ ☆ ☆         Add to cart ♥ Export citation )         Quick Abstract   |  |
| acquisition         mag           production         surf           reflection         20.0           exploration         seif           acquired         higt           conventional         seif           response         seif           httpd         seif           db         seif           velocities         seif           velocities         seif           processing         seif           imaging         seif           sufface         seif           information         seif           attributes         pro | rine seismic<br>face seismic<br>seismic<br>smic seismic<br>smic seismi<br>h-resolution seismi<br>smic source<br>smic inversion<br>h resolution<br>smic acquisition<br>smic acquisition<br>smic velocity<br>smic imaging<br>rehole seismic<br>smic volume<br>ean bottom<br>smic amplitudes<br>smic differences<br>aduction data |  |        | Marine Magnetotellurics For Base Salt Mapping: Gulf<br>of Mexico Field-Test At the Gemini Structure       Image: Constable of the second structure         Hoversten, G. Michael, University of California at Berkeley<br>Constable, Steve, Scripps Institution of Oceanography<br>Morrison, H. Frank, University of California at Berkeley and Lawrence<br>Berkeley National Laboratory         1998-0425 SEG Conference Paper - 1998            ★★★★★<br>Add to cart<br>Export citation         ) |  |









Your search for seismic AND "gulf of mexico" AND magnetotelluric has returned 32 results in 0.103 seconds

| seismic  | Filter your results by Publication date: All dates Journal or conference: All publications Publisher: All publishers   | Sort by:       Relevance       Items per page:       10       Image: Sort by:         Mark all       Add marked to cart       Export marked citations       Image: Sort by:       Ima  |
|--|--|--|
| Gulf of           Algorithm A         Algorithm B           data         seismic data           3D         3D seismic           reservoir         time           well         seis           amplitude         seis           sequencial         factor           acquisition         4D seis           sequencial         seis           sequencial         seis           sequencial         seis           sequencial         seis           sequencial         seis           sequencial         sedimenta           production         seris           exploration         seis           acquired         hiti | Type:<br>All types<br>magnetotelluric (MMT) survey was<br>system in the Gulf of Mexico. For<br>base of salt, when compared with w<br>se of salt was determined, but add<br>iry host provides indication of lithol<br>vity zones and low velocity was also<br>effects of the method, was also inv | 19659-MS OTC Conference Paper - 2008<br>View rights & permissions<br>Action Action Conference Paper - 2008<br>View rights & permissions<br>Action Conference Paper - 2008<br>Action C |
| conventional         seismic source           response         seismic inversion           marine         high resolution           Introduction         seismic acquisition           4D         seismic inversion           resolution         bersimic imaging           processing         seismic velocities           technology         seismic velocities           surface         seismic amplitudes           information         seismic differences           attributes         production data  |  | Hoversten, G. Michael, University of California at Berkeley<br>Constable, Steve, Scripps Institution of Oceanography<br>Morrison, H. Frank, University of California at Berkeley and Lawrence<br>Berkeley National Laboratory<br>1998-0425 SEG Conference Paper - 1998<br>☆☆☆☆☆☆<br>Add to cart 🛒 Export citation  |



### 53 Practicing geoscientists in two organizations





## Methodology – Focus Groups (Morgan 1997)





## **Results – Qualitative Themes**

Game playing 'spot the difference'

"It is like open up the box for me and I'll pick what does not fit with my brain, like one of those games".

#### **Surprisingness - Learning**

".. some of them attract my attention because they are very unique... I am looking for unique things that trigger my attention,".

"the algorithm made clear I underestimated the importance of Carbonates in Malaysia, immediately important to the research I am doing now" **Enablers/constraints** 

Belief System / Personality Traits

"This is overwhelming maybe too much information"

"Excitement was the first thought I had.."

Organization maturity for Enterprise Search **Business uses** 

**Competitor Intelligence** 

**Idea generation** 

Help find analogues

**Project framing/research** 

**Taxonomy Development** 



## Results – Quantitative Questionnaire



To what extent do current search interfaces in your organization facilitate serendipitous discovery? **41% - To a moderate/large extent**  To what extent could discriminatory word co-occurrence techniques facilitate serendipitous discovery? 73% - To a moderate/large extent







## Stories from practice

"Word associations highlighted new and unexpected terms such as 'metamorphic sole' associated with the secondary keyword 'platform'. This surprising result led us to consider a new geological element which could impact our opportunity"

> New Venture Geologist, Multinational Oil Company (December2014)



### **Summary**

- Some enterprise search user interfaces may be more likely to facilitate serendipitous encounters than others.
- Limitations to semi-interactive stimulant. Need for more interactive tool to further research phenomena.



#### Citice 365 Preview

| pro        | oject  | Q         |
|------------|--|-----------|
| Ever       | ything People Conversations Videos   | Réports   |
| pe 🐴:      | 2012 Packaging project   | · · · · · |
| Tob<br>Ker | y Nixon Sr. Product Manager, Project Lead Yukari<br>motes Product Manager, Co. Project Lead Chice Proceed    | 200       |
| Factor     | earch Project Manager Did you not like the one I<br>ect scope  |           |
| hub        | elbert1-myuhanepoint.com/ _/Weekly Meeting Polyentat_  |           |
| (E)        | Contoso Electronics Packaging Project  | 20. 20.   |
| 2012       | about the person, not the packaging Emotional<br>metion is keen to prove marketing                           | - 21-     |
| wyie inter | ives brand logality It's time to break new ground with our   |           |
| hub        | elitestLsharepoint.com//2909 Rackaging Review.pptx   |           |
| n          | Contract Standard 2010 Marketica Contractor  | 14        |
| Pro        | ect team Toby Noon Sr. Product Manager. Project Land   | 2540      |
| - 71       | kari Kemmotsu Product Manager, Co-Project Lead Chice<br>cast Research Project Manager Let's out made for the |           |
| City City  | -  |           |
| Fub        | eNestLsharepoint.com//Contoso 2010 Marketing Cong  |           |
|            | Project Falcon Architecture  |           |
| Feb        | ellextLsharepoint.com/Project Falcon Architecture  |           |
|            |  |           |



#### Citice 365 Preview

| 100 B 100         |   |          |
|-------------------|---|----------|
|                   | project.  | Q        |
|                   | Everything People Conversations Videos  | Reports  |
| Result type       | 型 2012 Packaging project  | -        |
| Show all          | Toby Nixon Sr. Product Manager, Project Lead Yukari<br>Kennepitia Product Manager Co-Project Lead Chice Brusser   | 200 C    |
| PowerPoint        | Research Project Manager Did you not like the one I<br>Project scope  |          |
| Author            | hubelterzi-ray-charepoint.com/(Weekly Meeting Preventat   |          |
| Steph             | Contoso Bectionics Padaging Project   | 190. 200 |
| junnill Has       | It's about the person, not the packaging Emotional<br>connection is key to good marketing The right connection  | - 2      |
| David Longneuit   | inspires brand logally It's time to break new ground with our<br>exclusion  |          |
| Jukan Isla        | FubilitiestLohangpoinLoony/2009 Packaging Neview.pptx   |          |
| Gernet Vergen     | La constante de   |          |
| SHOW MORE         | Contoso Biectronics 2010 Marketing Campaign   | 311. E   |
| Intriguing?       | Project team Toby Necon Sr. Product Manager, Project Lead<br>Yukari Sammotsu Product Manager, Co Project Lead Chice<br>Brassard Research Project Manager Let's get ready for the<br>Dev | 1111     |
| Accident          | FubertestLsharepoint.com//Contoso 2010 Marketing Camp   |          |
| Bermuda           | Project Falcon Architecture   |          |
| Inland Revenue    | hubeltestLsharepoint.com//Project Falcon Architecture   |          |
| Meteorite         |   |          |
| Sale<br>SHOW MORE | Project Falson Overview   |          |



Thank you for listening Paul H. Cleverley <u>http://www.rgu.ac.uk/dmstaff/cleverley-paul</u>

p.h.cleverley@rgu.ac.uk

