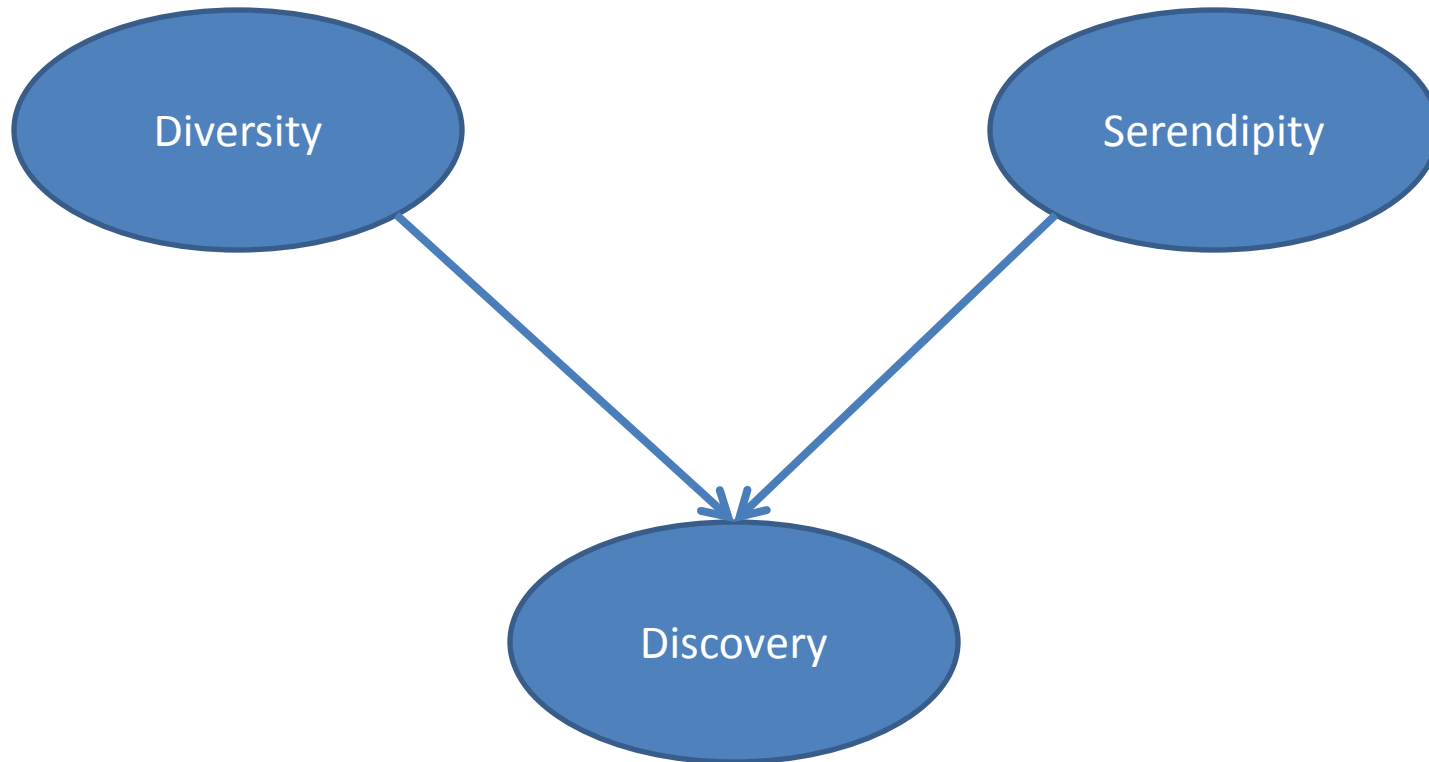


New evaluation measures


Nisheeth

Newer dimensions of evaluation

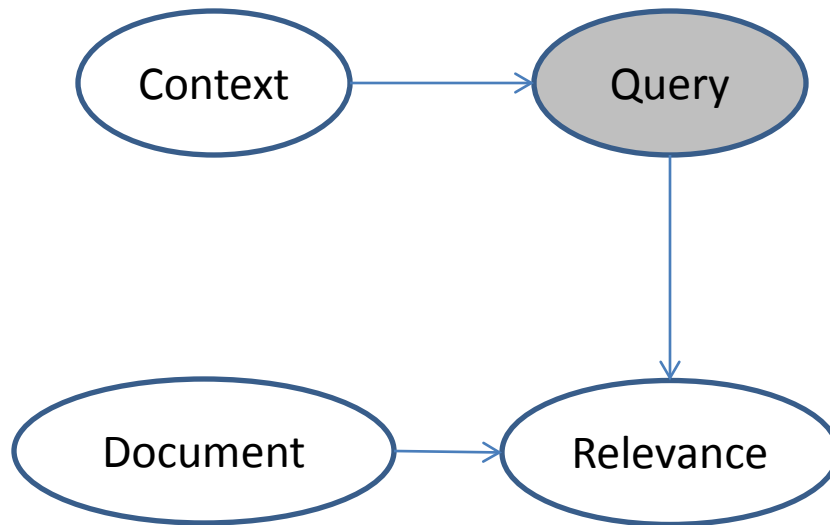


Suggested reading: Santos, R. L., Macdonald, C., & Ounis, I. (2015). Search result diversification. *Foundations and Trends in Information Retrieval*, 9(1), 1-90.

Search result diversification

Google	Bing	Yahoo
<p>Microsoft Windows windows.microsoft.com/ ▼ The official website for the Microsoft Windows operating system. Get the latest information, get downloads, and find support. Support - Free downloads - Windows 7</p> <p>Window - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Window ▼ A window is a transparent or translucent passage of light and, if not closed or covered, allows air to pass. Microsoft Windows - Casement window</p> <p>Window (computing) - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Window_(computing) ▼ In computing, a window is a visual area on a computer screen that displays a graphical user interface. In signal processing, a window function (or tapering function) is a mathematical function that is zero outside a certain range.</p> <p>Sign In home.live.com/ ▼ Outlook.com is free modern email service that lets you easily connect with Office Web Apps.</p> <p>Images for window - Report image</p>	<p>Microsoft Windows windows.microsoft.com ▼ The official website for the Microsoft Windows operating system. Get the latest information, get downloads, and find support. Windows 7 Hotmail Windows XP Meet Support Micro Internet Explorer Prod</p> <p>Microsoft Home Page Devices www.microsoft.com ▼ Microsoft Home. Products, Downloads, Office, Windows Phone, Xbox, and more.</p> <p>Free downloads - Microsoft Windows windows.microsoft.com/en-US/windows Get free downloads for Windows 8, Windows 7, and Windows Vista.</p> <p>Microsoft Windows Update windowsupdate.microsoft.com Latest bug fixes for Microsoft Windows.</p> <p>News about Windows bing.com/news Microsoft to unveil latest Windows 8 San Jose Mercury News - 1 hour ago SAN FRANCISCO—Microsoft is giving away a free Windows 8 upgrade that promises to address some of the</p>	<p>Window - Image Results</p>  <p>More Window images</p> <p>Door and Window Replacement Windows www.doorandwindow.com Cached Browse information about replacement windows and how to find a door or window contractor.</p> <p>Window - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Window Cached Etymology History Types Terminology A window is a transparent or translucent passage of light and, if not closed or covered, allows air to pass.</p> <p>Andersen Windows & Doors www.andersenwindows.com Cached</p>

Intent-based diversification



To diversify, maximize

$$P(D|q) = \sum_c p(c|q) \left(1 - \prod_{d \in D} (1 - p(r, d|q, c))\right)$$

Geographical diversification



Diversification = chunking up area into contexts and proceeding as above

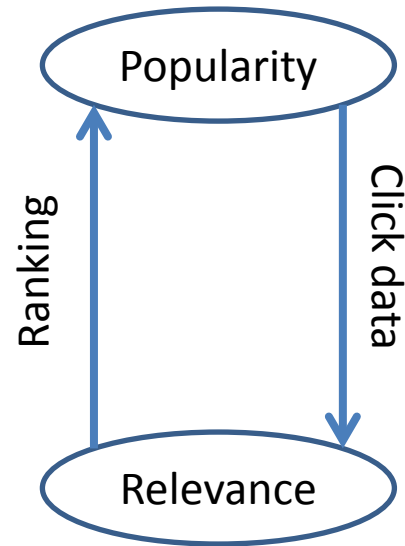
Multi-domain diversification

- Search for upcoming event in specific city in particular month
- Diversification = increasing number {event, city, month} tuples in top k results

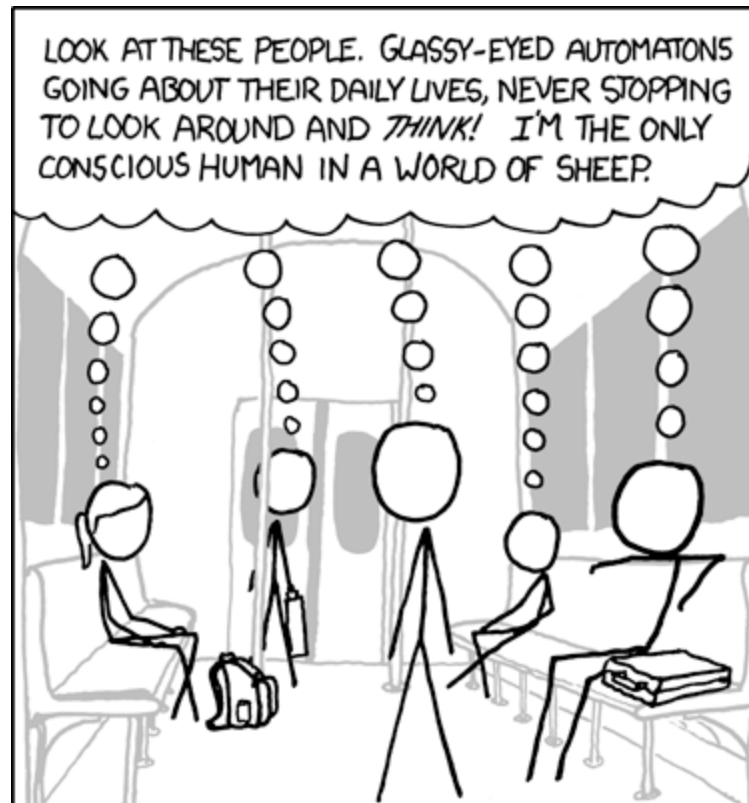
Generalized evaluation criteria

- $NDCG - IA(D) = \sum_c p(c|q)NDCG(D|c)$
- $MRR - IA(D) = \sum_c p(c|q)MRR(D|c)$
- $MAP - IA(D) = \sum_c p(c|q)MAP(D|c)$
- Context-aware judgment of classical measurements simply assumes all documents from the wrong cluster are irrelevant
 - All probabilities upper bounded considerably below 1

The filter bubble

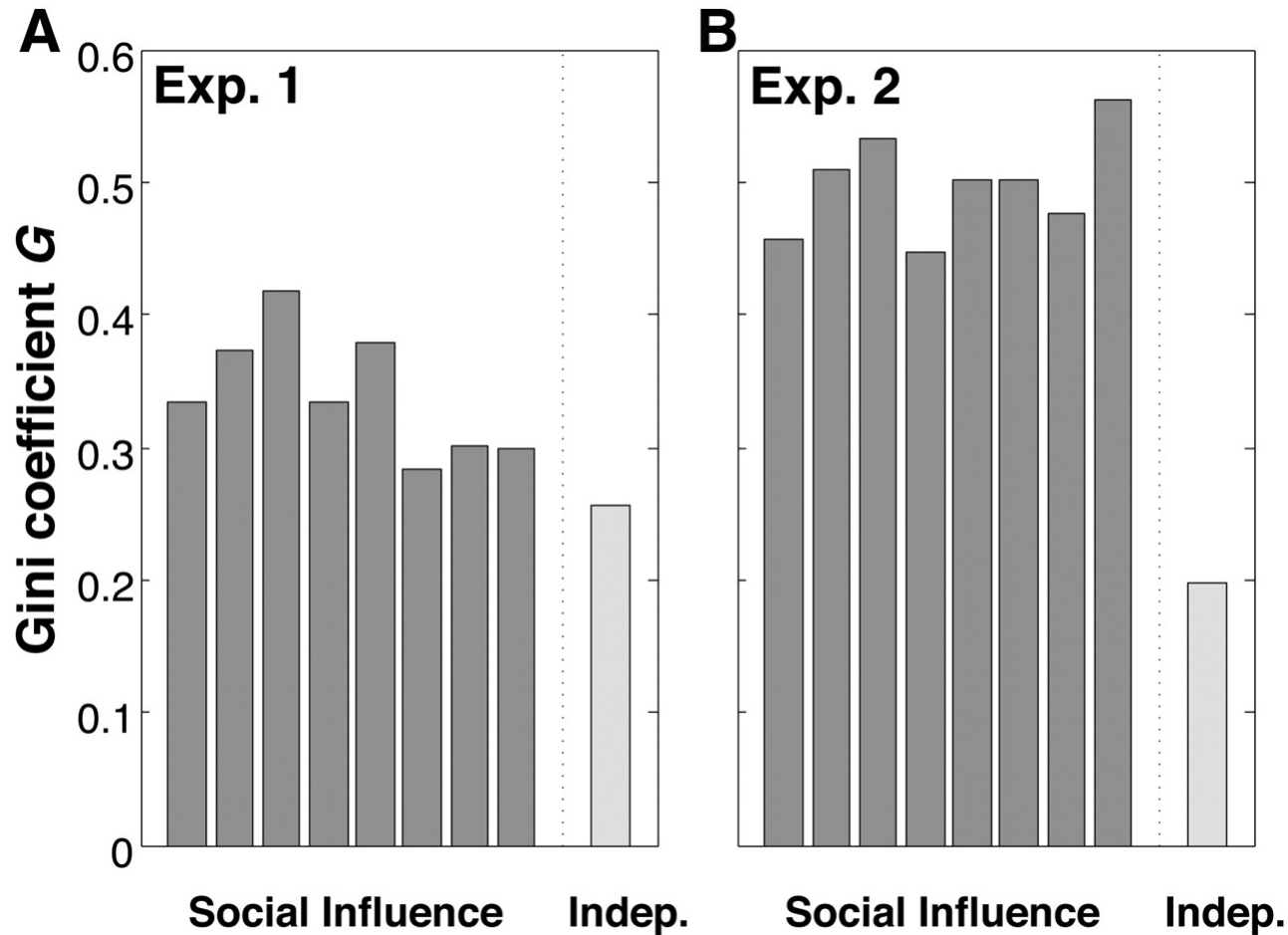


Micro-bubble



Personalized search reinforces pre-existing preferences, suppressing exogenous influences

Macro-bubble



(Salganik, Dodds & Watts, 2006)

Global query log utilization promotes preferential attachment to a small set of relevant items

Massive economic implications



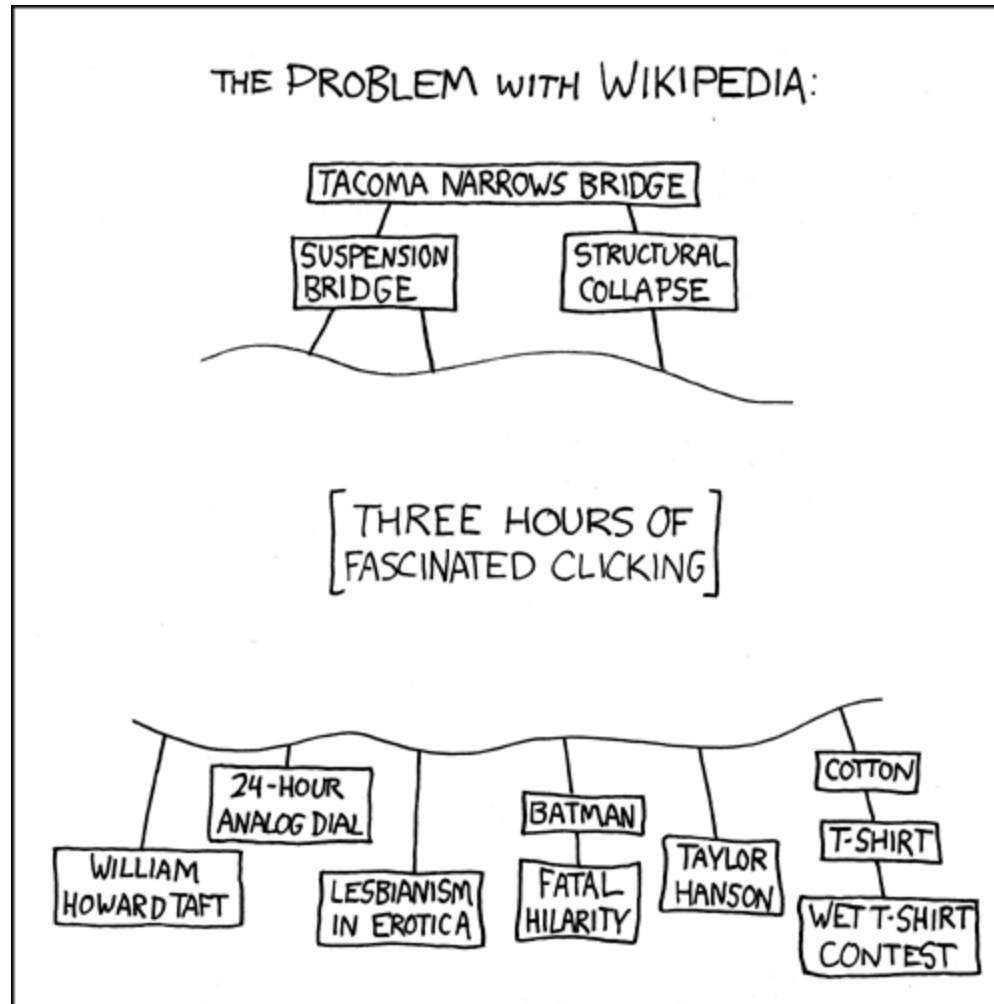
Defining and measuring serendipity

- Definition: unexpectedly encountering meaningful information, not necessarily directly relevant to query
- Usually measured against a baseline
 - $SRDP(SE) = \frac{\sum_{i \in UNEXP} rel_i}{|UNEXP|}$
- But can explicitly probe also (Rahman & Wilson, 2015)
 - Ask people why they clicked on a link using checkboxes at random intervals
 - Measure frequency with which they mention clicking links not directly relevant to query

Promoting serendipity

- Unexpectedness
 - Add noise to pre-existing ranking results (Thom-Santelli, 2007)
- Meaningful
 - Add salient perceptual cues that stand out in UI display (Burkell et al, 2009)
 - Push notifications (Campos & de Figuerido, 2001)
- Tangential relevance
 - Open problem

Shining example



Cross-links boost serendipity

How to introduce cross-links usefully in search UI?

Glowing counter-example

How Does Facebook Choose What To Show In News Feed?

$$\text{News Feed Visibility} = * \text{I} \times \text{P} \times \text{C} \times \text{T} \times \text{R}$$

Interest Post Creator Type Recency

Interest

Interest of the user
in the creator

Post

This post's
performance
amongst
other users

Creator

Performance of past
posts by the content
creator amongst
other users

Type

Type of post
(status, photo,
link) user prefers

Recency

How new is the post

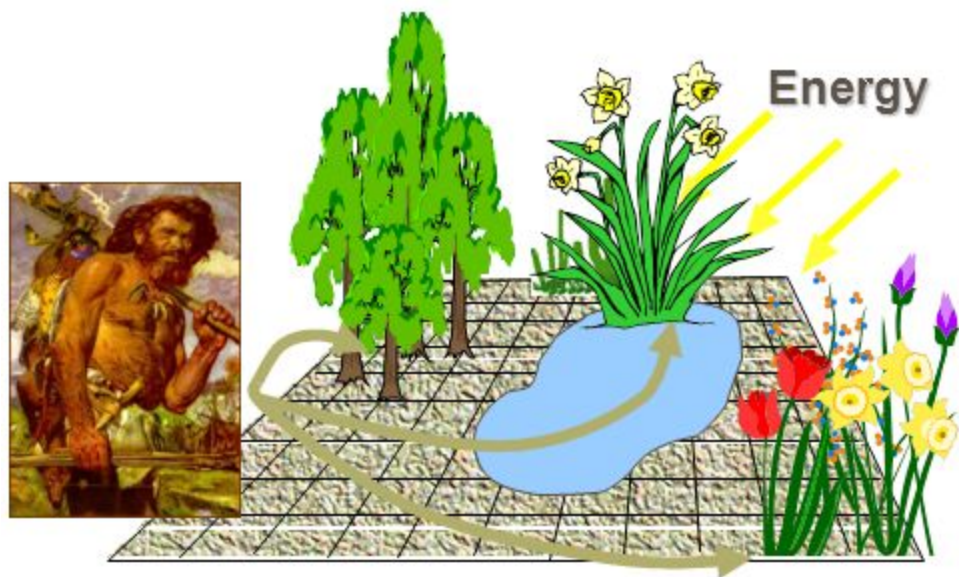
* This is a simplified equation. Facebook also looks at roughly 100,000 other high-personalized factors when determining what's shown.

Design principles for promoting serendipity

- Conceptual connections, not statistical ones
- Flat UI display
- Non-commercial motivation
 - Crowd-sourced content
 - Avoid lawyers
- How to do this for dynamic content?
 - Develop a model of what the person might find interesting

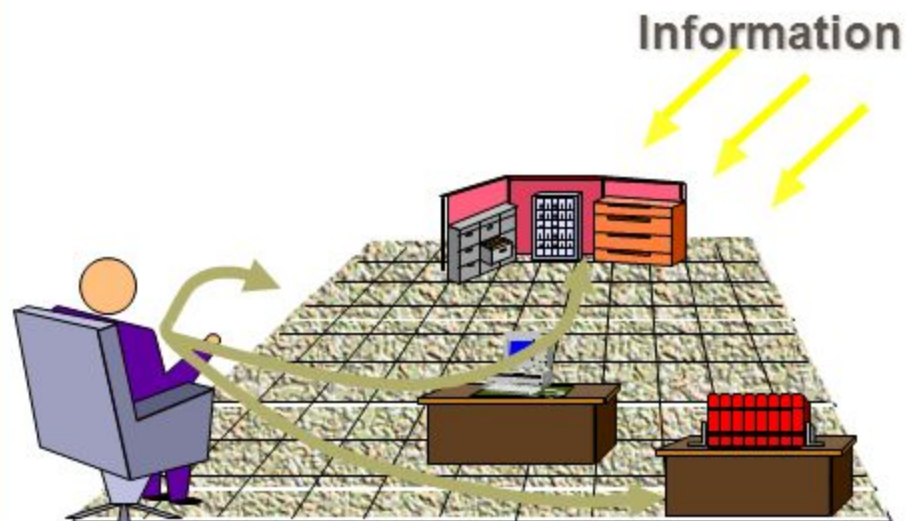
OPTIMALITY THEORY

Optimal Foraging Theory



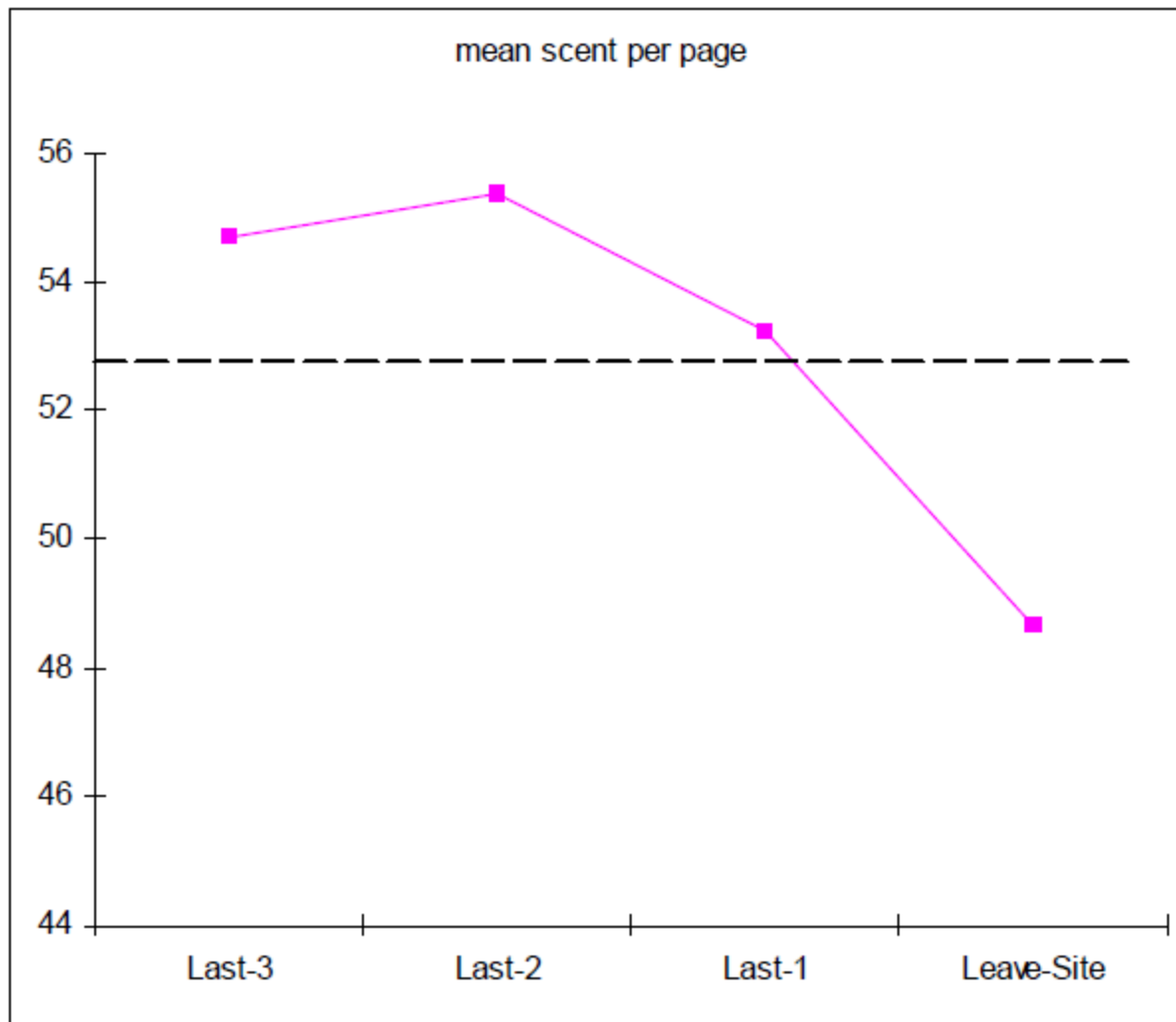
$$\text{Max} \left[\frac{\text{Energy}}{\text{Time}} \right]$$

Information Foraging Theory



$$\text{Max} \left[\frac{\text{Useful info}}{\text{Time}} \right]$$

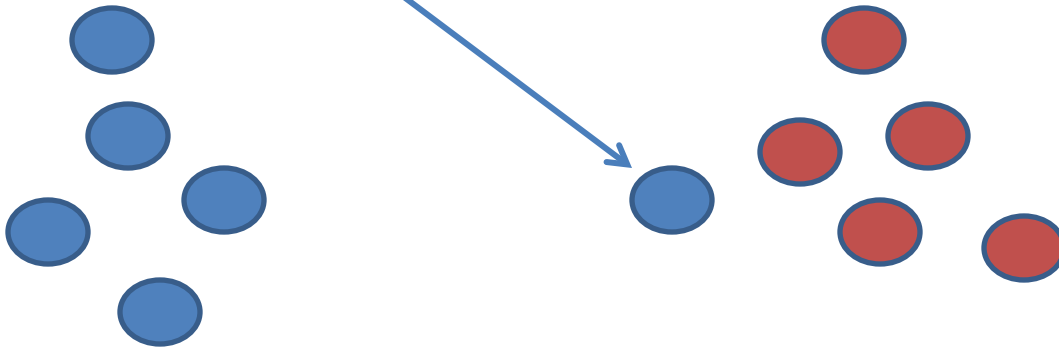
People are sensitive to information scent



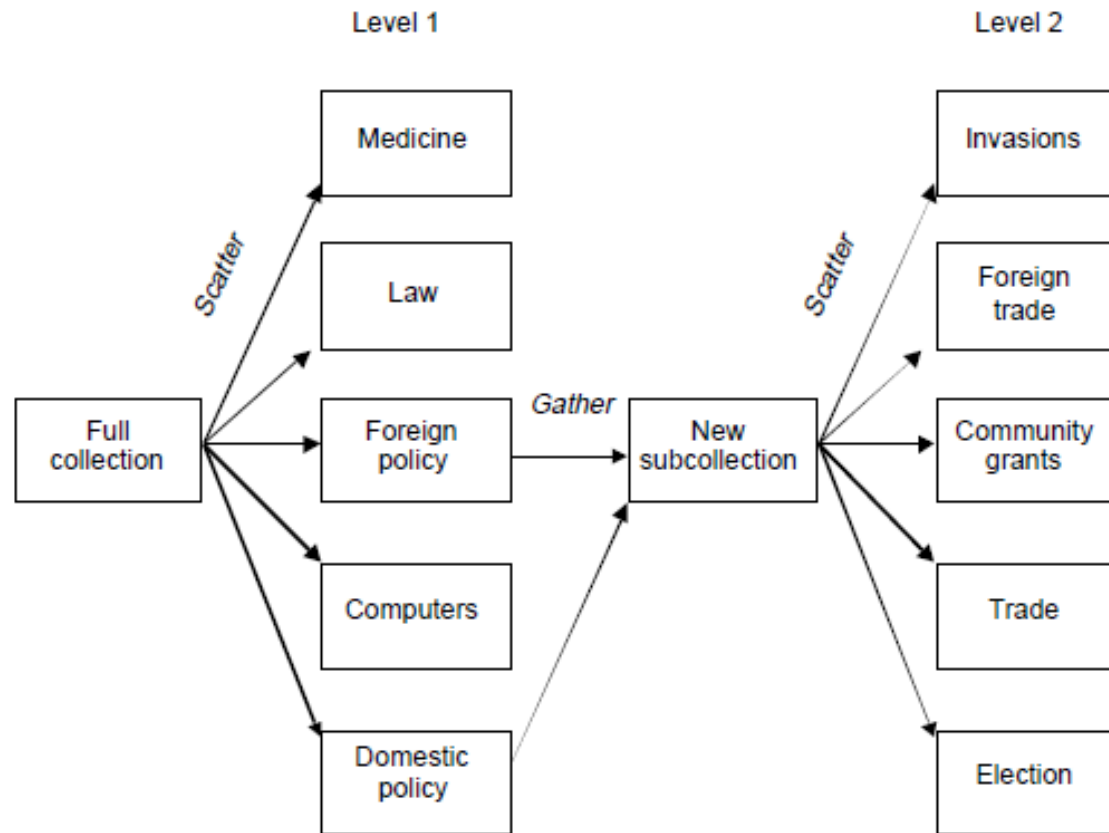
(Pirolli & Fu, 2003)

Implications

- People will look at content in patches/clusters
- When the information rate within a cluster drops, they will move to the next
- Reducing transit time between clusters will cause quicker jumps
- Hypothesis: This will be interesting



Presenting search results better



Scatter-gather (Pirulli & Card, 1999)

Scatter-gather UI

Scatter/Gather 0

Scatter/Gather

Text Database: trec1
This Window: 0
Parent Window: -1

Options:

Scatter/Gather
Show Titles
Similarity Search
Boolean Search
Backup
Hide Window

Search Terms:

Make Query Record Set SimSearch Threshold

Goto Window:

0

Cluster 0 (38940)

cell, patient, radiation, dose, beam, disease, treatment, ima
AP: Early Results In Hospital Patient Study Sho (aid, study, percent, health,)
DOE: Doses of secondary radiation appearing as (radiation, dose, exposure, s)
AP: Poll: AIDS Test Confidentiality Opposed in (percent, study, drug, report)

Cluster 5 (122371)

user, software, computer, network, ibm, version, pc, line, d
ZF: Networks for the '90s: getting down to busi (user, software, application,)
ZF: Windows 3.0 wins with the users. (News Anal (application, user, window, s)
ZF: Visual programming: is a picture worth 1,00 (software, program, user, app)

Cluster 1 (23564)

court, judge, law, attorney, appeal, lawyer, trial, justice,
AP880502-0111 (court, state, u.s., federal,)
WSJ: Supreme Court to Review New York Law That (court, law, federal, case, j)
WSJ: Law -- Legal Beat: Court Affirms Verdict i (court, case, state, law, fed)

Cluster 6 (46144)

section, 1988, 1989, public, regulation, office, action, agen
FR: Subpart 1852.1_Instructions for Using Provi (1, section, information, 2,)
FR: Community Development Block Grants (section, 1, state, rule, 2,)
FR: Proprietary Trading Systems (commission, file, 1989, rule)

Cluster 2 (37760)

fire, mile, spokesman, monday, friday, thursday, sunday, wedn
AP: Troops Violated Rules in Village Raid Last (police, kill, report, offici)
AP: Up To 50,000 Reported Killed In Armenian Qu (official, report, government)
AP: Two Palestinians Wounded in Clash With Sold (police, kill, army, attack,)

Cluster 7 (48384)

0, phase, reaction, concentration, 6, cm, investigate, metal,
DOE: The composition of NO/sub x/ generated in (2, 3, 1, rate, reaction, ch,)
DOE: Using a 40 cm bent crystal spectrograph of (2, 1, 3, refs, gas, tab,)
DOE: X-ray method is used to study phase compos (2, 4, 3, 5, solution, study,)

Cluster 3 (106437)

american, bush, country, house, soviet, party, leader, think
AP: U.S. Invasion of Panama Draws Fire From Lat (state, government, u.s., uni)
WSJ: --- President's Power Is Slipping; Soviet (u.s., government, soviet, of)
WSJ: The Iraqi Invasion of Kuwait -- Persian Gu (u.s., soviet, government, of)

Cluster 8 (116210)

coal, fuel, flow, technique, oil, pressure, production, mean
DOE: A specialized software system which allowe (data, time, measurement, mea
DOE: We present results of our theoretical and (refs, tab, model, 1, 4, proc)
DOE: Distributed control systems provide their (process, control, developmen)

Cluster 4 (174353)

share, stock, trade, billion, bank, sale, sell, york, buy, e
WSJ: Reliance Electric Mulls Public Offering; B (million, company, stock, sha)
WSJ: Enterprise: @ Staying Private Helped Rainb (company, million, sale, busi)
AP: WASHINGTON: ... he said. (company, million, share, off)

Cluster 9 (28670)

field, nuclear, plant, waste, reactor, magnetic, plasma, theo
DOE: The effect of the radial electrostatic fie (field, magnetic, plasma, wav)
DOE: The operational principle of nuclear power (plant, nuclear, analysis, de)
DOE: Wilson 'renormalization group' fixed point (field, theory, equation, fun)