

Search Engine Strategies  
February 27, 2006, NYC

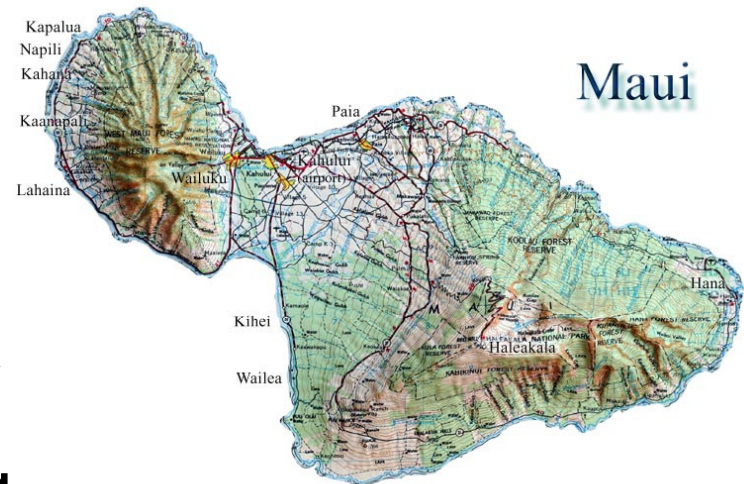
**Click Streams,  
Complexity,  
and Contribution:**  
Modeling Searcher Behavior  
Using Markov Models

**Alan Rimm-Kaufman, PhD**  
Rimm-Kaufman Group LLC  
[www.rimmkaufman.com](http://www.rimmkaufman.com)



# Click streams

Vacation ►  
hawaiian flight ►  
Mauii ► Maui ►  
grand wailea resort ►  
surf shop lahaina long



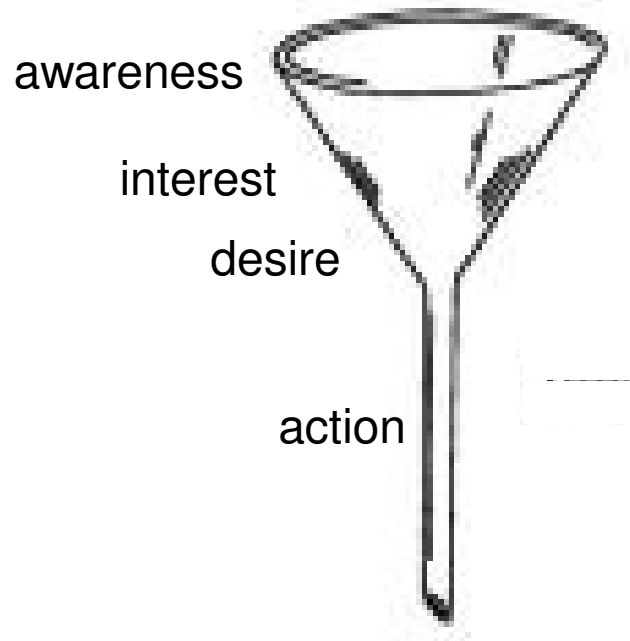
# Click values



vacation	550k imps	\$1.02 cpc
hawaiian flight	225 imps	\$0.62 cpc
mauii	584 imps	\$0.16 cpc
maui	110k imps	\$2.14 cpc
grand wailea resort	3335 imps	\$1.22 cpc
surf shop lahaina longboard	0 imps	\$0.10 cpc

Monthly impression counts and #1 CPCs from Yahoo!, Feb 2006

# Conventional wisdom



*...more **generic search phrases** indicate a searcher is **higher up in the conversion funnel...** generic phrases lay the **groundwork for more specific searches...** thus, **even if the economics of more general phrases don't meet an advertiser's ROI target,** general phrases play an essential role in a search portfolio...*

vacation ► hawaiian flight ► maui ► maui ► grand wailea resort  
camcorder ► digital camcorder ► samsung SC-DC 164  
safe car ► volvo ► volvo s80 ► volvo s80 atlanta dealer

# True?

Often the data say:

# NO.

Let's build a mathematical model to see.



*three matrix formulas ahead!*



A. A. Markov (1856).

# Markov model

discrete state, discrete time Markov chain  
states  $i=1..n$

transition probabilities  $P_{ij}$

“memoryless” property

$$P_{ij} = P(X_{n+1} = j \mid X_n = i)$$

stationary distribution from eigenvectors

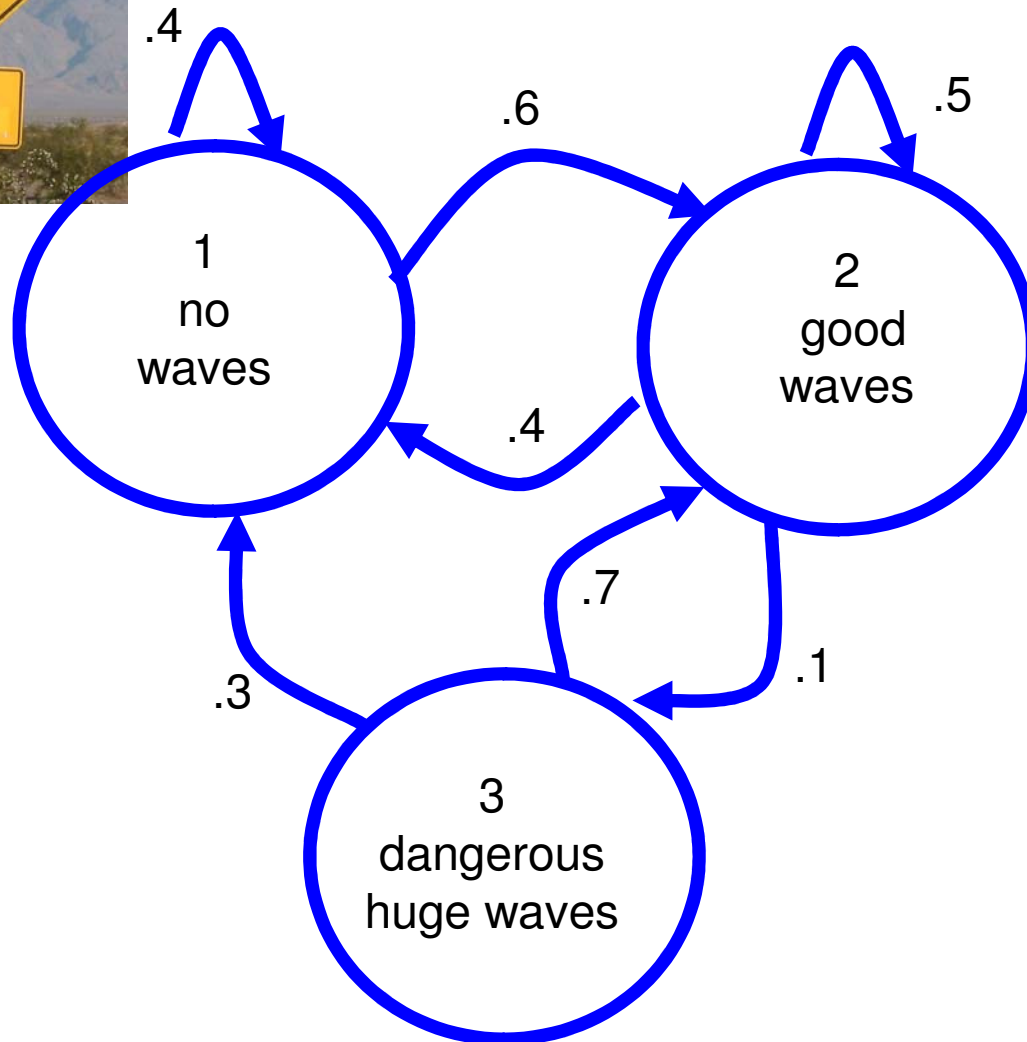
$$\pi^T \mathbf{P} = \pi^T$$

most important for us:

limiting distribution of random walk

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \mathbf{1} \pi^T$$

# Toy example


$$P_{ij}$$

0.4	0.5	0.3
0.6	0.4	0.7
0.0	0.1	0.0

$$\lim_{k \rightarrow \infty} P^k = 1\pi^T$$

$$\Pi_{ij}$$

.445
.505
.050

?@#!

...what the heck does this  
math gibberish have anything  
to do with SEO / SEM?





# SEO / SEM implications

SEO: Google uses markov models for...

PageRank v1

detection of paid links

bad neighborhoods

SEM: RKG uses markov models for...

**click stream analysis**

# SEM click stream complexity

Step 1: Characterize search phrases

1 word phrases

2 word phrases

3 word phrases

4+ word phrases

client brand phrase

SKU

non-dictionary



# SEM click stream complexity

## Step 2: Characterize click streams

vacation ▶ hawaiian flight ▶ maui ▶ maui ▶ grand wailea resort  
1 ▶ 2 ▶ ND ▶ BP ▶ BP

camcorder ▶ digital camcorder ▶ samsung SC-DC 164  
1 ▶ 2 ▶ SKU

safe car ▶ volvo ▶ volvo s80 ▶ volvo s80 atlanta dealer  
2 ▶ BP ▶ SKU ▶ SKU

Repeat 500,000 times



# SEM click stream complexity

## Step 3: Estimate transition probabilities, $p_{ij}$

### Dataset:

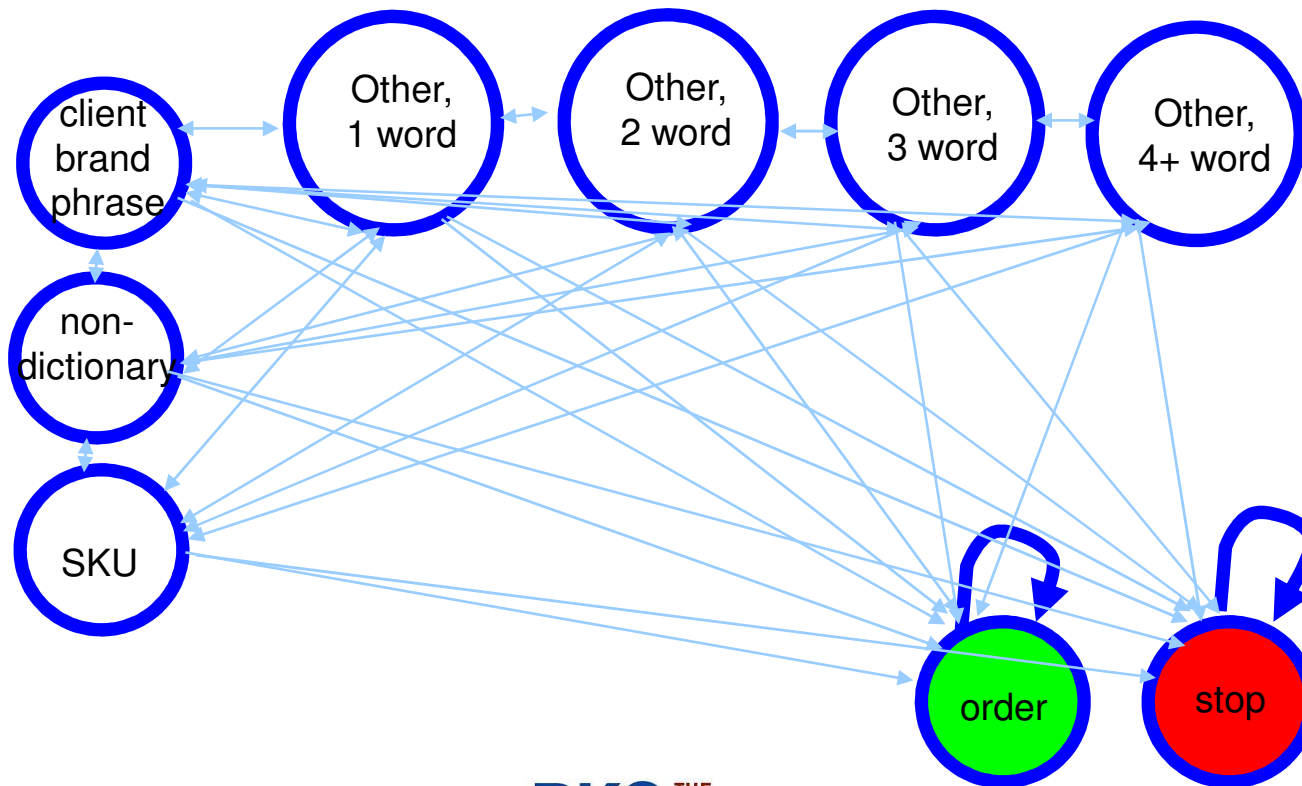
- random sample of 500K clickstreams (695,446 clicks) across 2005
- Google and Yahoo paid search only
- 50+ typical retailers (highly specialized niche retailers excluded)
- B2C only
- order conversions only
- 30 day cookie window
- clickstreams by searcher by retailer



# SEM click stream complexity

Step 4: compute stationary probabilities

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \mathbf{1}\pi^T$$



# 4 conclusions





# **1. Many click streams are short.**

mean length = 1.4 clicks

significant long tail, however

## **Marketing significance:**

Economics, optimization, & bidding by phrase (rather than by click stream) is an extremely good approximation. Can safely ignore pundits who claim otherwise.



## **2. Many click-stream are redundant.**

$P_{ij}$  heavily loaded on the diagonal.  
12% of click streams with >1 click  
have duplicates.

### **Marketing significance:**

Browser auto-completion plays a large  
role in click-streams and navigation.  
Good copy and good landing pages,  
tailored to the ad, are critically important.



### **3. Your brand matters. A lot.**

Conversion typically 200% to 300% higher on brand phrases.

#### **Marketing significance:**

Break out results for phrases involving your brand. Evaluate your in-house team or your SEM agency on their success driving non-brand sales.



## 4. Brand for brand, not ROI.

The “funnel benefit” of generic keywords is, for many advertisers, quite small.

### **Marketing significance:**

Choosing to spend large sums on branding through generic keyword buys is a valid strategy. Realize, however, that such ad buys are about branding, not conversion.

# Mahalo!



Want to learn more? [rimmkaufman.com/ses-feb-06](http://rimmkaufman.com/ses-feb-06)

- this powerpoint presentation
- links to academic research papers for background
- link to free software for markov analysis

Alan Rimm-Kaufman (alan@rimmkaufman.com)