

# LET YOUR USERS DO THE RANKING

Ojas Parekh

1465 - Discrete Math and Complex Systems

**DATACon 2011**



Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.



**Sandia National Laboratories**

# Enterprise Analytics at Sandia

***Improve the ability of the workforce to find the information they need to perform their jobs, wherever it may reside***

From: XXXXXXXX  
Sent: Thursday, September 16, 2010 1:40 PM  
To: LibSearch  
Subject: Help with SearchPoint

My search for "WebCars" came up with this as the first hit:  
[Sandia Data Storage Service \(SDSS\): Private vs. Shared](#)  
...there is absolutely nothing about WebCars there ... ***Why can't SearchPoint ever seem to find anything useful or relevant?*** ...

Regards,  
XXXXXXXXXX

Please let us know if you need assistance with any of the following:

- ~~Adding your website to SearchPoint~~
- ~~Removing your website from SearchPoint~~
- ~~Getting a search box for your website~~
- ~~Adding an acronym to SearchPoint~~
- ~~Adding a word to SearchPoint's spell checker~~

***Replace SearchPoint with something that works***

- Complaints from users of SearchPoint, Sandia's enterprise search solution
- Late 2009: Enterprise Analytics project formed and funded to advance algorithmic technology in search applications
- 2010: Embarked on effort to develop and implement a model to improve ranking of search results based on user behavior





# Using Users

- Enterprise search is hard; usual suspects don't work:  
“Some customers ask to what extent Google universal search technology relies on PageRank. The answer is: very little.”
- Implicit feedback from user search patterns
- Radlinski & Joachims, KDD 2005
- Apply machine learning/optimization to passively collected user feedback





# Aggregate Behavior Model

- Collect historical data: What are the queries? Which search results are clicked?
- Apply machine learning/optimization to discover preferred results
- Adjust the ranking of search results for queries based upon user preference
- New data (user behavior) incorporated daily
- Results improve:
  - **Immediately upon implementation, and**
  - **Naturally and automatically over time as a function of user actions**





# How do Users Search?

*How do I get reimbursed for lunch when interviewing a job candidate?*

- Initial query may not be the “right” query
- Results are used (repeatedly) to refine the query
- Generates a query chain

interview meal	
1.	<a href="#">Result</a> snippet for result 1
2.	<a href="#">Result</a> snippet for result 2
3.	<a href="#">Result</a> snippet for result 3
4.	<a href="#">Result</a> snippet for result 4
5.	<a href="#">Result</a> snippet for result 5
6.	<a href="#">Result</a> snippet for result 6
7.	<a href="#">Result</a> snippet for result 7
8.	<a href="#">Result</a> snippet for result 8

**Query:** interview meal  
**User clicked on:** #5

interview host	
1.	<a href="#">Result</a> snippet for result 1
2.	<a href="#">Result</a> snippet for result 2
3.	<a href="#">Result</a> snippet for result 3
4.	<a href="#">Result</a> snippet for result 4
5.	<a href="#">Result</a> snippet for result 5
6.	<a href="#">Result</a> snippet for result 6
7.	<a href="#">Result</a> snippet for result 7
8.	<a href="#">Result</a> snippet for result 8

**Query:** interview host  
**User clicked on:** none

interview host meal	
1.	<a href="#">Result</a> snippet for result 1
2.	<a href="#">Result</a> snippet for result 2
3.	<a href="#">Result</a> snippet for result 3
4.	<a href="#">Result</a> snippet for result 4
5.	<a href="#">Result</a> snippet for result 5
6.	<a href="#">Result</a> snippet for result 6
7.	<a href="#">Result</a> snippet for result 7
8.	<a href="#">Result</a> snippet for result 8

**Query:** interview host meal  
**User clicked on:** #1

Outlined region: What user looked at (eye tracking studies)



# How do Users Search?

*How do I get reimbursed for lunch when interviewing a job candidate?*

Results from the SRN/SRN Web

SearchPoint  
SON/SRN Web Results

Search:  Search

☐ Expand your search to include words with the same meaning

Advanced Look Up Acronyms...

Help Contact Us FAQs

SON/SRN Web Results

Corporate Policy System Technical Library Catalog Technical Library Report Search Sandia OrgFinder Sandia Yellow Pages

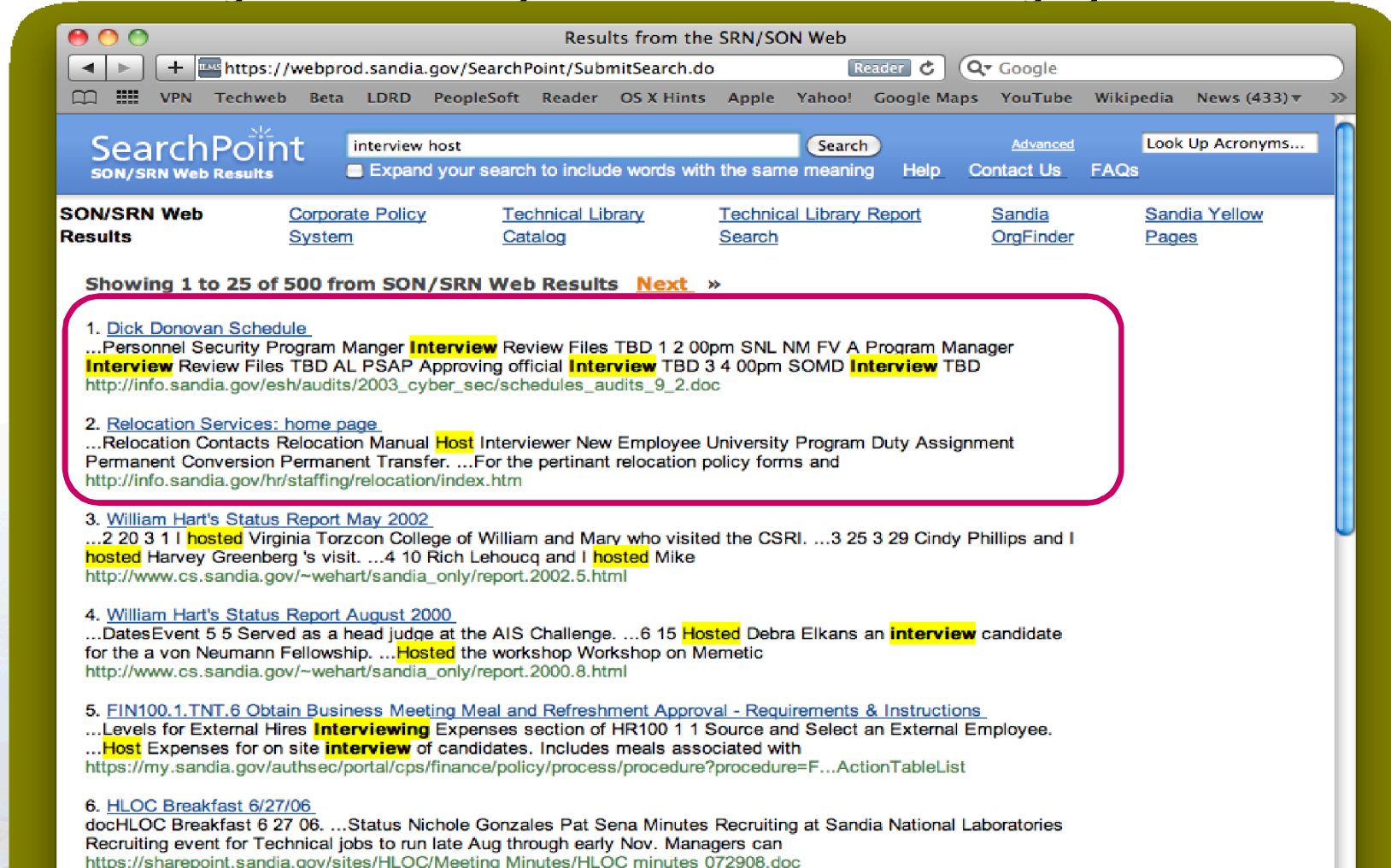
Showing 1 to 25 of 143 from SON/SRN Web Results **Next** >>

1. [SF 6432-BM:BUSINESS MEETING with Meal/Refreshment are Provided](#)  
...Business **Meal** attendance form SF 6432 BM business meeting meeting with **meals** SF 4600 BM. ...Business **Meals** Checklist Prior written management approval is required for serving any business **meal** and  
[https://sharepoint.sandia.gov/sites/05056\\_Operations/IO\\_Visitors/Document Templates/4600bm.doc](https://sharepoint.sandia.gov/sites/05056_Operations/IO_Visitors/Document Templates/4600bm.doc)
2. [Prior written management approval is always required for serving business meals and/or refreshments](#)  
...Business **Meals** Checklist Prior written management approval is required for serving any business **meal** and or refreshments at Sandia. ...Recruiting other targeted recruiting efforts Management  
[https://sharepoint.sandia.gov/sites/05056\\_Operations/IO\\_Visitors/Business Meals/BMM\\_C...ed Nov 2008.doc](https://sharepoint.sandia.gov/sites/05056_Operations/IO_Visitors/Business Meals/BMM_C...ed Nov 2008.doc)
3. [Microsoft Word - 4600bm.doc](#)  
...This documentation includes A completed Business Meeting **Meal** Form SF 4600 BM for all meetings where **meals** and or refreshments ...Providing Business Meeting **Meals** The meeting must cross a normal  
<http://info.sandia.gov/corpdata/corppforms/4600bm.pdf>
4. [No title](#)  
...37 Phone **Interviews** Phone **Interviewing** Success Three Types of Telephone **Interviews** Telephone **Interview** Preparation Phone Personality Matching Technique Open and Available Technique Stand and  
[https://sharepoint.sandia.gov/sites/Student-Networking-Site/Shared Documents/Tips\\_for...nterviewing.pdf](https://sharepoint.sandia.gov/sites/Student-Networking-Site/Shared Documents/Tips_for...nterviewing.pdf)
5. [Relocation Services: Host Interviewer](#)  
...Local transportation rental car if travel is by air **Meals** and lodging en route and for duration of **interview**. ...**Meals** and lodging are limited to the maximum per diem rate. ...Up to four  
[http://info.sandia.gov/hr/staffing/relocation/host\\_interviewer.htm](http://info.sandia.gov/hr/staffing/relocation/host_interviewer.htm)
6. [Business Meeting Meals - Changes for FY05](#)  
...Business **Meals** Business Meeting **Meals** BIT BIT Meeting cfo finan. ...Reports Update of Financial Manual Business Expenses 18 1 concerning Business Meeting **Meals** Implementation of Business Meeting  
[http://cfo.sandia.gov/finan/oracle/bit-minutesFY05/mar05\\_business\\_meals.ppt](http://cfo.sandia.gov/finan/oracle/bit-minutesFY05/mar05_business_meals.ppt)



# How do Users Search?

*How do I get reimbursed for lunch when interviewing a job candidate?*



Results from the SRN/SON Web

SearchPoint  
SON/SRN Web Results

Search: interview host

Expand your search to include words with the same meaning

Advanced Look Up Acronyms...

Help Contact Us FAQs

SON/SRN Web Results

Corporate Policy System Technical Library Catalog Technical Library Report Search Sandia OrgFinder Sandia Yellow Pages

Showing 1 to 25 of 500 from SON/SRN Web Results [Next](#) >>

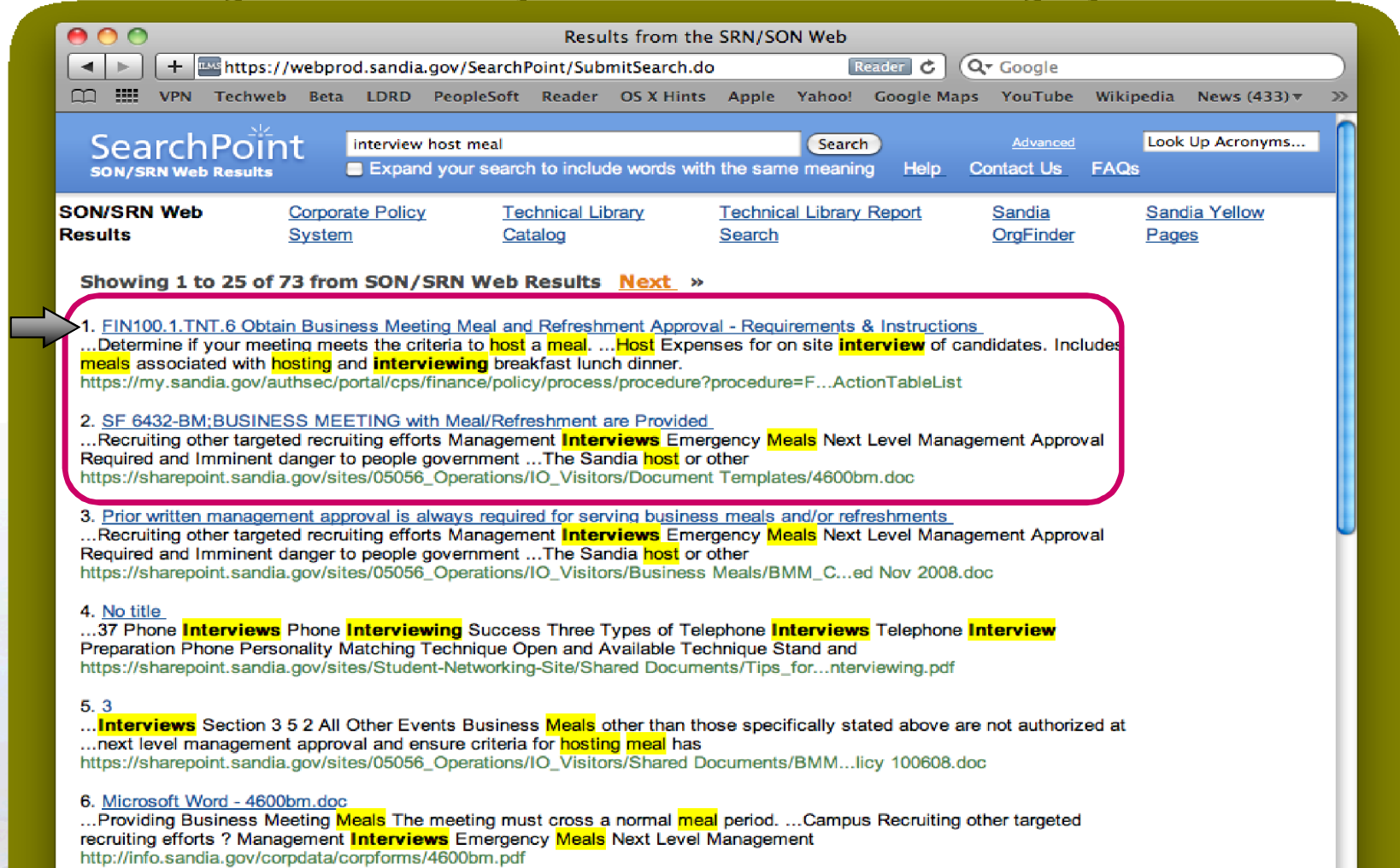
- [Dick Donovan Schedule](#)  
...Personnel Security Program Manger **Interview** Review Files TBD 1 2 00pm SNL NM FV A Program Manager  
**Interview** Review Files TBD AL PSAP Approving official **Interview** TBD 3 4 00pm SOMD **Interview** TBD  
[http://info.sandia.gov/esh/audits/2003\\_cyber\\_sec/schedules\\_audits\\_9\\_2.doc](http://info.sandia.gov/esh/audits/2003_cyber_sec/schedules_audits_9_2.doc)
- [Relocation Services: home page](#)  
...Relocation Contacts Relocation Manual **Host** Interviewer New Employee University Program Duty Assignment  
Permanent Conversion Permanent Transfer. ...For the pertinent relocation policy forms and  
<http://info.sandia.gov/hr/staffing/relocation/index.htm>
- [William Hart's Status Report May 2002](#)  
...2 20 3 1 I **hosted** Virginia Torzcon College of William and Mary who visited the CSRI. ...3 25 3 29 Cindy Phillips and I  
**hosted** Harvey Greenberg 's visit. ...4 10 Rich Lehoucq and I **hosted** Mike  
[http://www.cs.sandia.gov/~wehart/sandia\\_only/report.2002.5.html](http://www.cs.sandia.gov/~wehart/sandia_only/report.2002.5.html)
- [William Hart's Status Report August 2000](#)  
...DatesEvent 5 5 Served as a head judge at the AIS Challenge. ...6 15 **Hosted** Debra Elkans an **interview** candidate  
for the a von Neumann Fellowship. ...**Hosted** the workshop Workshop on Memetic  
[http://www.cs.sandia.gov/~wehart/sandia\\_only/report.2000.8.html](http://www.cs.sandia.gov/~wehart/sandia_only/report.2000.8.html)
- [FIN100.1.TNT.6 Obtain Business Meeting Meal and Refreshment Approval - Requirements & Instructions](#)  
...Levels for External Hires **Interviewing** Expenses section of HR100 1 1 Source and Select an External Employee.  
...**Host** Expenses for on site **interview** of candidates. Includes meals associated with  
<https://my.sandia.gov/authsec/portal/cps/finance/policy/process/procedure?procedure=F...ActionTableList>
- [HLOC Breakfast 6/27/06](#)  
doCHLOC Breakfast 6 27 06. ...Status Nichole Gonzales Pat Sena Minutes Recruiting at Sandia National Laboratories  
Recruiting event for Technical jobs to run late Aug through early Nov. Managers can  
<https://sharepoint.sandia.gov/sites/HLOC/Meeting Minutes/HLOC minutes 072908.doc>





# How do Users Search?

*How do I get reimbursed for lunch when interviewing a job candidate?*



Results from the SRN/SON Web

https://webprod.sandia.gov/SearchPoint/SubmitSearch.do

SearchPoint  
SON/SRN Web Results

interview host meal

Search

Expand your search to include words with the same meaning

Help Contact Us FAQs

SON/SRN Web Results

Corporate Policy System Technical Library Catalog Technical Library Report Search Sandia OrgFinder Sandia Yellow Pages

Showing 1 to 25 of 73 from SON/SRN Web Results [Next](#) >>

1. [FIN100.1.TNT.6 Obtain Business Meeting Meal and Refreshment Approval - Requirements & Instructions](#)  
...Determine if your meeting meets the criteria to **host** a **meal**. ...Host Expenses for on site **interview** of candidates. Includes **meals** associated with **hosting** and **interviewing** breakfast lunch dinner.  
<https://my.sandia.gov/authsec/portal/cps/finance/policy/process/procedure?procedure=F...ActionTableList>
2. [SF 6432-BM;BUSINESS MEETING with Meal/Refreshment are Provided](#)  
...Recruiting other targeted recruiting efforts Management **Interviews** Emergency **Meals** Next Level Management Approval Required and Imminent danger to people government ...The Sandia **host** or other  
[https://sharepoint.sandia.gov/sites/05056\\_Operations/IO\\_Visitors/Document Templates/4600bm.doc](https://sharepoint.sandia.gov/sites/05056_Operations/IO_Visitors/Document Templates/4600bm.doc)
3. [Prior written management approval is always required for serving business meals and/or refreshments](#)  
...Recruiting other targeted recruiting efforts Management **Interviews** Emergency **Meals** Next Level Management Approval Required and Imminent danger to people government ...The Sandia **host** or other  
[https://sharepoint.sandia.gov/sites/05056\\_Operations/IO\\_Visitors/Business Meals/BMM\\_C...ed Nov 2008.doc](https://sharepoint.sandia.gov/sites/05056_Operations/IO_Visitors/Business Meals/BMM_C...ed Nov 2008.doc)
4. [No title](#)  
...37 Phone **Interviews** Phone **Interviewing** Success Three Types of Telephone **Interviews** Telephone **Interview** Preparation Phone Personality Matching Technique Open and Available Technique Stand and  
[https://sharepoint.sandia.gov/sites/Student-Networking-Site/Shared Documents/Tips\\_for...nterviewing.pdf](https://sharepoint.sandia.gov/sites/Student-Networking-Site/Shared Documents/Tips_for...nterviewing.pdf)
5. [3](#)  
...**Interviews** Section 3 5 2 All Other Events Business **Meals** other than those specifically stated above are not authorized at ...next level management approval and ensure criteria for **hosting meal** has  
[https://sharepoint.sandia.gov/sites/05056\\_Operations/IO\\_Visitors/Shared Documents/BMM...licy 100608.doc](https://sharepoint.sandia.gov/sites/05056_Operations/IO_Visitors/Shared Documents/BMM...licy 100608.doc)
6. [Microsoft Word - 4600bm.doc](#)  
...Providing Business Meeting **Meals** The meeting must cross a normal **meal** period. ...Campus Recruiting other targeted recruiting efforts ? Management **Interviews** Emergency **Meals** Next Level Management  
<http://info.sandia.gov/corpdata/corpforms/4600bm.pdf>





# Inferring User Preferences

interview meal	
1.	<a href="#">Result</a> snippet for result 1
2.	<a href="#">Result</a> snippet for result 2
3.	<a href="#">Result</a> snippet for result 3
4.	<a href="#">Result</a> snippet for result 4
5.	<a href="#">Result</a> snippet for result 5
6.	<a href="#">Result</a> snippet for result 6
7.	<a href="#">Result</a> snippet for result 7
8.	<a href="#">Result</a> snippet for result 8

**Query:** interview meal  
**User clicked on:** #5

**Inference:**  
#5 preferred over #1-4 and #6

interview host	
1.	<a href="#">Result</a> snippet for result 1
2.	<a href="#">Result</a> snippet for result 2
3.	<a href="#">Result</a> snippet for result 3
4.	<a href="#">Result</a> snippet for result 4
5.	<a href="#">Result</a> snippet for result 5
6.	<a href="#">Result</a> snippet for result 6
7.	<a href="#">Result</a> snippet for result 7
8.	<a href="#">Result</a> snippet for result 8

**Query:** interview host  
**User clicked on:** none

**Inference:**  
none

interview host meal	
1.	<a href="#">Result</a> snippet for result 1
2.	<a href="#">Result</a> snippet for result 2
3.	<a href="#">Result</a> snippet for result 3
4.	<a href="#">Result</a> snippet for result 4
5.	<a href="#">Result</a> snippet for result 5
6.	<a href="#">Result</a> snippet for result 6
7.	<a href="#">Result</a> snippet for result 7
8.	<a href="#">Result</a> snippet for result 8

**Query:** interview host meal  
**User clicked on:** #1

**Inference:**  
#1 preferred over #2

**Inference from Entire Query Chain:**  
#1 (from last query) preferred over #5 (from first query)





# Reranking

- Harvest preferences from query chains
- For each query, seek reordering of docs
- Goal: satisfy “important” preferences
  - May be impossible to satisfy all
  - May be undesirable to satisfy all



# Support Vector Machine Model

- Model preferences with underlying feature space
- Learn vector  $w$ , weighting of feature space
- For each preference  $r_i >_q r_j$ , impose constraint:

$$\varphi(r_i, q) \bullet w > \varphi(r_j, q) \bullet w$$

vector indicating  
relationship in  
feature space

(e.g. common terms)

vector weights  
each feature



# Search Log Data

**Search ID**

**Search Time**

**Click Time**

**User ID**

**Query Text**

**Clicked URL**

**URL Position**

**Number of Results**

**URL #1**

**URL #2**

⋮

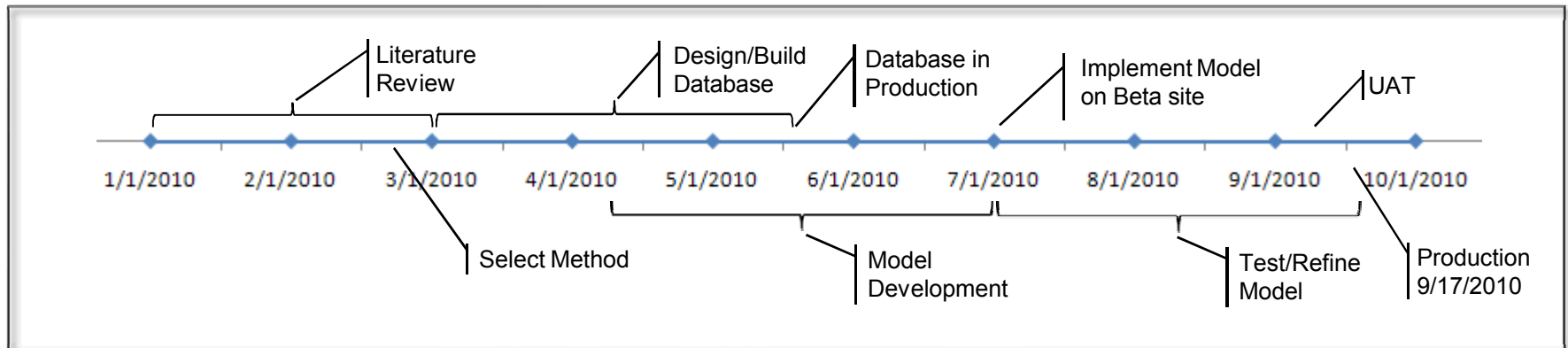
- One row of data each time the user clicks on a result
- If the user clicks on 3 URLs after entering a query, there will be 3 rows with differing values in Click Time, Clicked URL, and URL Position; remaining columns will have the same values for each of the 3 rows
- If the user did not click on any of the returned results, there will be one row for which Clicked URL will be null and URL Position will be zero.





# Develop & Deploy Model

## Model Development Timeline



## Re-Ranking Process (nightly):

- Identify query chains
- Harvest preferences
- Learn  $w$  using SVM solver
- For each query, rank its results by value  $\varphi(r_i, q) \bullet w$

## Search Process:

- User executes search query
- Do we have behavior associated with the query?
- Yes: present re-ranked results interleaved w/SearchPoint results
- No: present search engine results

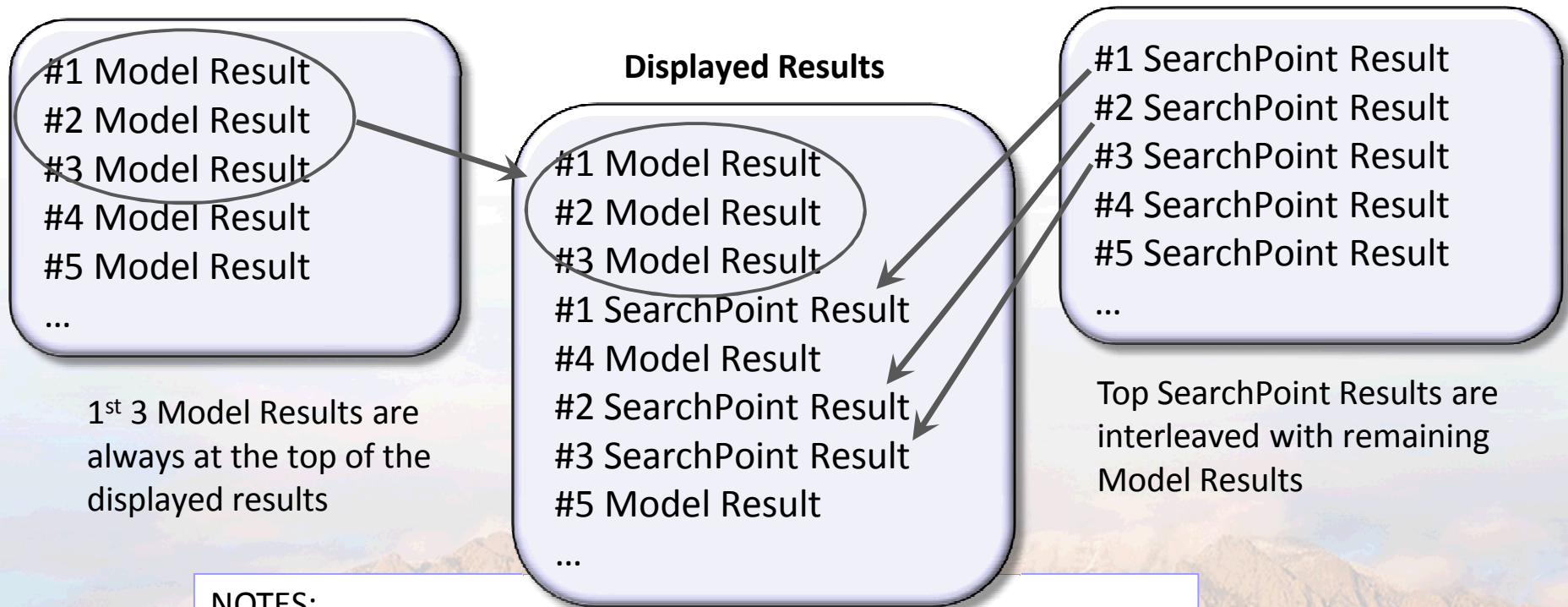


# What about “new” search results? (e.g. new web pages)

**FACT:** Users rarely look past the 1<sup>st</sup> page of results

**ISSUE:** If the 1<sup>st</sup> page of results contains ONLY results from the model, users will not have the chance to click on new search results

**SOLUTION:** Discover timing and methods for injecting new results (new web pages, new topics) into the re-ranked results



## NOTES:

- If Best Bets exist, they are at the top of the SearchPoint Results.
- Displayed Results are de-duped.



# Aggregate Behavior Modeling Example

## Query: retirement

### Raw Search Engine Results (Retirement 101 is #9)

1. [Reimbursement Voucher Form](#)  
snippet for result 1
2. [401\(k\) Login](#)  
snippet for result 2
3. [All Sandia Summary Plan Desc](#)  
snippet for result 3
4. [401\(k\) Savings Plans](#)  
snippet for result 4
5. [Payroll Services](#)  
snippet for result 5
6. [Retirement Help](#)  
snippet for result 6
7. [Pension Plan & Savings Plan M](#)  
snippet for result 7

### Raw Results Re-Ranked by Model (Retirement 101 is #1)

1. [Retirement 101 Homepage](#)  
snippet for result 1
2. [Retirement Memory Book](#)  
snippet for result 2
3. [Sandia 401\(k\) Savings Plans](#)  
snippet for result 3
4. [Benefits Homepage](#)  
snippet for result 4
5. [Pension Security Plan \(RSP\)](#)  
snippet for result 5
6. [Retirement Income Plan \(RI](#)  
snippet for result 6
7. [Sandia National Laboratorie](#)  
snippet for result 7

- Just prior to executing this query, Sandia had announced a change to the employee pension plan.
- These 2 results moved higher shortly after the announcement and remained highly ranked for several weeks.
- Now, many months after the announcement, interest has waned, and these 2 results are typically ranking around #20.

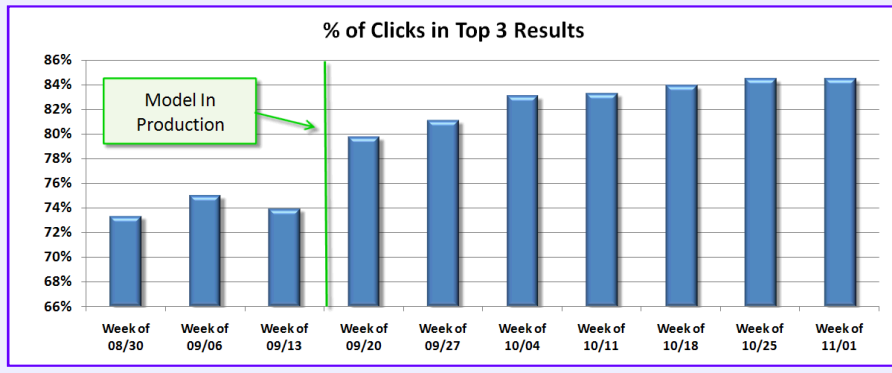
**NOTE:** The Model Results above are “raw results”. They do not include injected SearchPoint results that are in final results as displayed to the user.



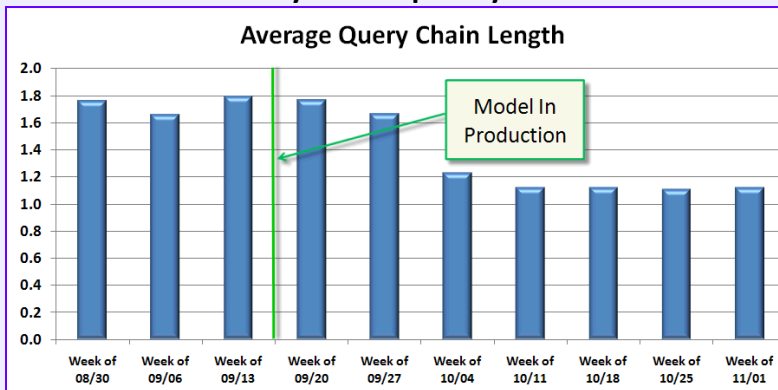
# Aggregate Behavior Improvement Metrics

If the model's ranking of search results truly places the more relevant results higher in the result set, we expect:

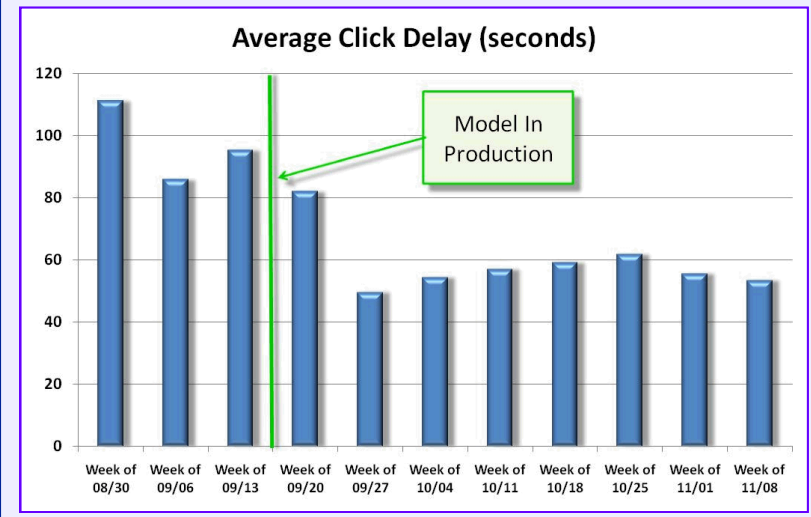
... users will click more frequently on results near the top of the results list



... the user will not find it necessary to modify the query terms



... the user will more quickly find, and click on, results that the user sees as relevant







# And ... complaints are now at an all time low

The “grumpy email” at the beginning of this presentation was sent **one day** before the Aggregate Behavior Model went into production.

Two business days after the first email, the same gentleman sent this:

**From: XXXXXX**  
**Sent: Monday, September 20, 2010 12:20 PM**  
**To: Spomer, Judith E**  
**Cc: LibSearch**  
**Subject: RE: Help with SearchPoint**

Greetings all:

It works much better now. And thank you for putting up with my grumpy e-mail.

Best regards,

XXXXXX



# What are we really solving?



1

2

3

4

1

3

3

2

2

2

X

1



?

?

?

?

Alice

Bob

Carlos





# (Partial) Rank Aggregation

- Given:  $n$  items,  $m$  (partial) rankings over items
- Optional: a priority/weight for each ranking
- Produce a single consensus ranking optimizing objective
- Many reasonable objectives



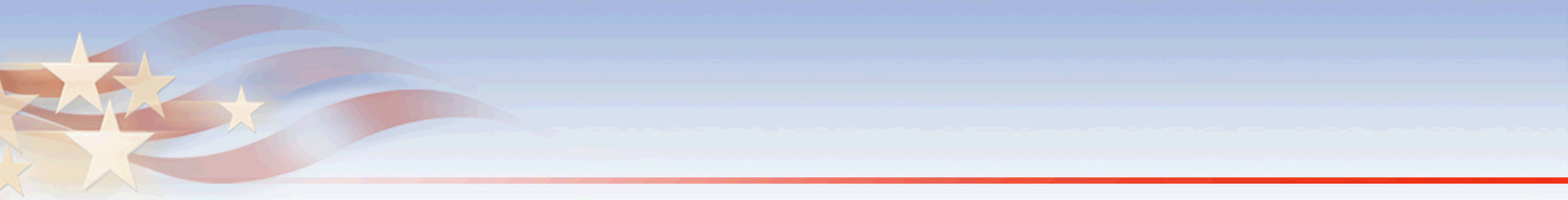


# Challenges

- Malicious users can skew results
- Do we have a good model for web search?
- Algorithms with a more precise performance guarantee?
- Algorithmic identification of poor ranking?







# Questions?

