

The First Click is the Deepest: Assessing Information Scent Predictions for a Personalized Search Engine

Karen Church, Mark T. Keane & Barry Smyth
Adaptive Information Cluster,
University College Dublin

Third Workshop on Empirical Evaluation of Adaptive Systems
Eindhoven, The Netherlands, 23rd August 2004



Introduction

- First-Click Behaviour
 - Enters query, scans results & chooses a link
- Aim: Assess Information Scent Predictions
- Two Information Scent Techniques
 - Cognitive Walkthrough for the Web (CWW)
 - The InfoScent Bloodhound Simulator
- Evaluate in Context of I-SPY



I-SPY

- Collaborative Search
 - Re-ranks results using community-based hits
- I-SPY Experiment
 - 2 sessions, 2 groups (Ns = 45 and 47)
 - 25 Computer Science/AI related questions
- Web Search Data
 - User queries, urls, user clicks, ranking information
 - 811 distinct queries
 - 10,445 unique pages returned
 - 427 unique pages clicked on by users



Cognitive Walkthrough for the Web

- Evaluates Website Usability using LSA
- Latent Semantic Analysis (LSA)
 - Captures word co-occurrence in large text corpora
- Calculate LSA Values for Query-Result Pair
- Issues
 - Missing terms
 - Query & result type used
 - String-matching better candidate



CWW Evaluation

- Common Data Set
 - 132 valid distinct queries
 - 2571 distinct results
- Sub-classified Queries
 - Short queries – terms input by users
 - Long queries – elaboration of question
- Results
 - (1) Title alone, (2) Blurb alone, (3) Title + Blurb
- Six Possible Query-Result Pairs

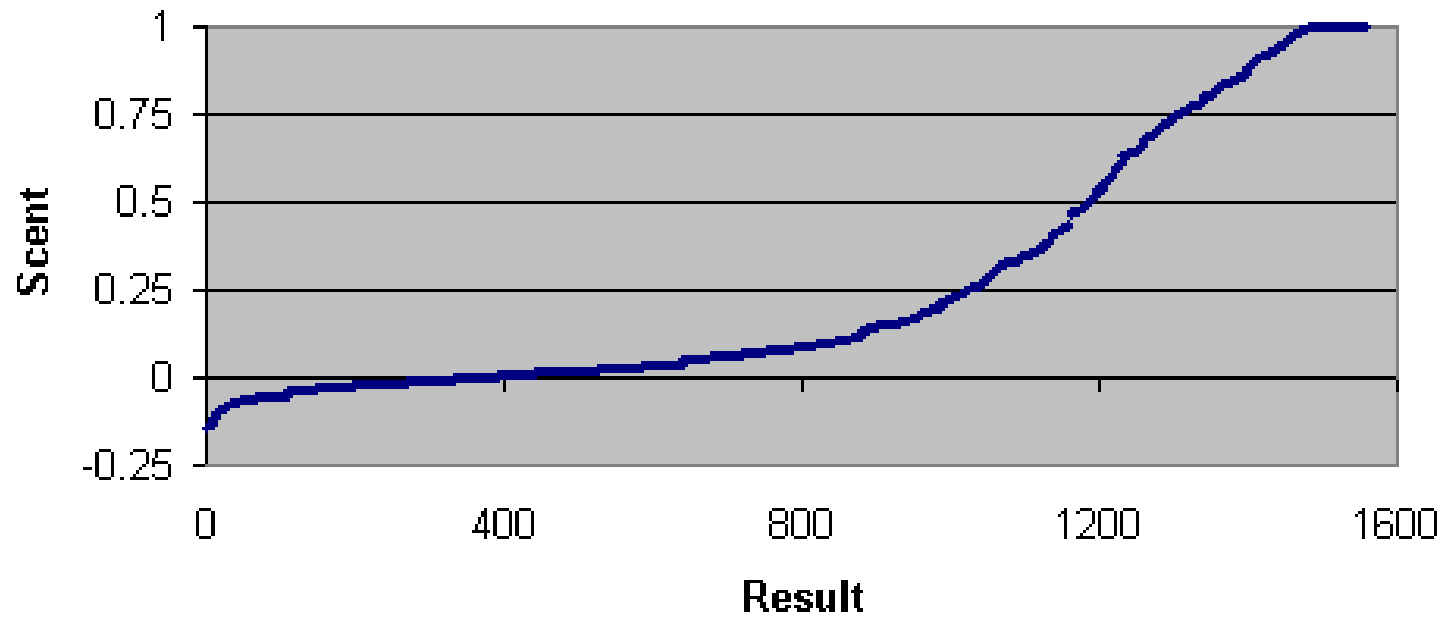


CWW Results

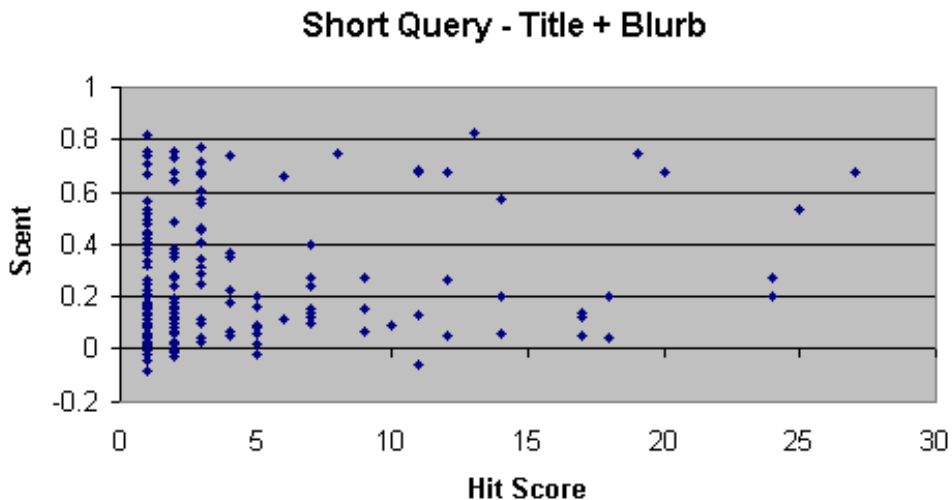
- Distribution: Scent Values Low
 - Mean 0.26
 - Standard Deviation 0.35
- Correlations of Scent x Hits Very Low
- “Best Case” is Not Good !
- May be Just String Matching

CWW Scent Distribution

Scent Distribution for Short Query-Title



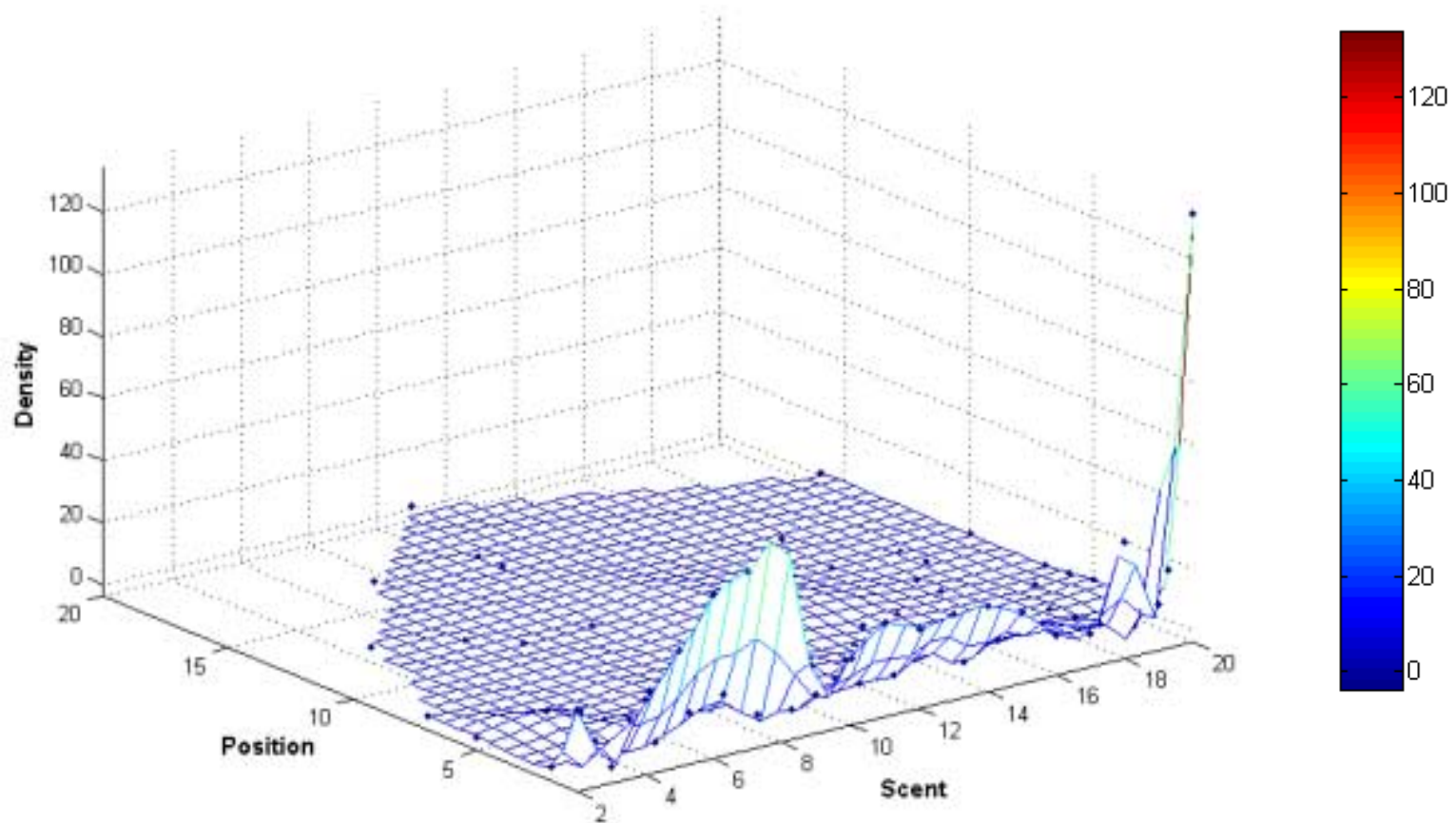
CWW Correlations



Result Type	Correlation
Short Query – Title	0.10
Short Query – Blurb	0.11
Short Query – Title + Blurb	0.13

- Common Hit Data
 - Links chosen by users, all of whom entered an identical query with same question in mind
 - 110 distinct queries
 - 1218 urls
- Weak Correlations
 - All ≤ 0.13

CWW Visualization





“Best Case” is Not Good !

- Removed Unsupported Query Terms
 - Weak correlation, all ≤ 0.15
- ‘Best-Fit’ Correlations
 - Pick most highly correlated case from the query-result matrix
 - The query-result pairing could be any of the 6 possibilities (short query & title, long query & blurb, etc)
 - Moderate correlation, 0.5



Does CWW = String Matching ?

- Tversky's Contrast Model of Similarity
 - $S(a, b) = \theta f(A \cap B) - \beta f(A - B) - \alpha f(B - A)$
- Correlations between Scent Scores & String Matches
- Moderate/High Correlation 0.6
- String Matching is a Better Candidate



The InfoScent Bloodhound Simulator

- Analyzes a Website and Generates a Usability Report
- WUFIS: Web User Flow by Information Scent
 - Predictive modeling algorithm
 - Information Retrieval techniques and Spreading Activation
- Generate Proximal Scent Matrix
- Issue with the Domain Used



Bloodhound Evaluation

- Same Common Data Set
 - 132 valid distinct queries
 - 2571 results
- Three Different Domains
 - Entire domain
 - Hits domain
 - Query domain



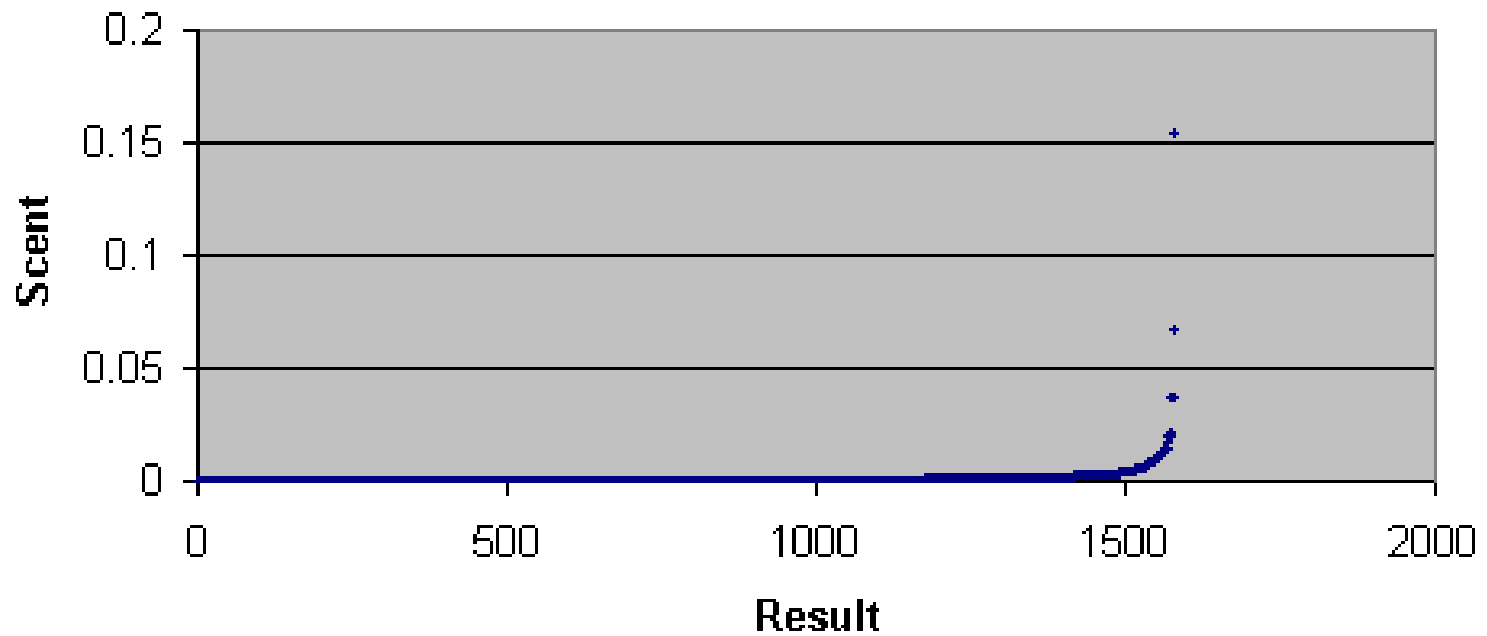
Bloodhound Results

- Distribution: Scent Values Very Low
 - Mean 0.001
 - Standard Deviation 0.005
- Correlations of Scent x Hits Negligible
- Specialized Domains Don't Work



Bloodhound Scent Distribution

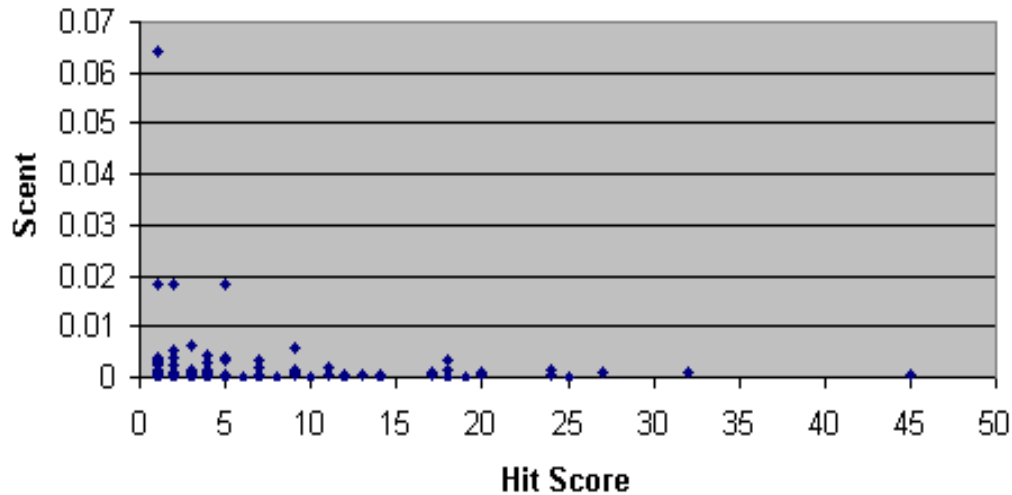
Scent Distribution - Entire Domain



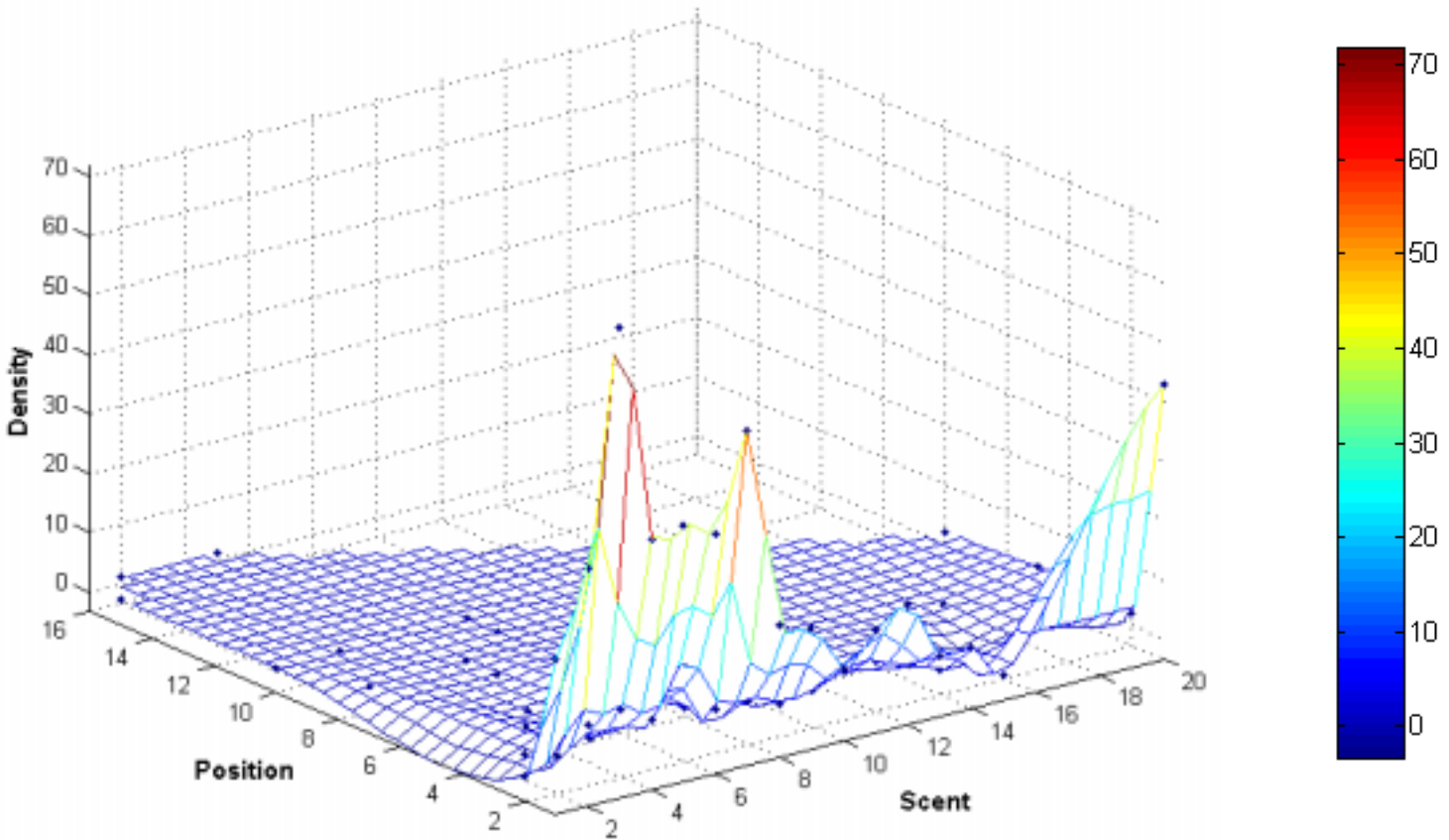
Bloodhound Correlations

- Poor Correlation
 - - 0.047
 - Negative
 - Negligible

Hit Score x Scent - Entire Domain



Bloodhound Visualization





Specialized Domains Don't Work !

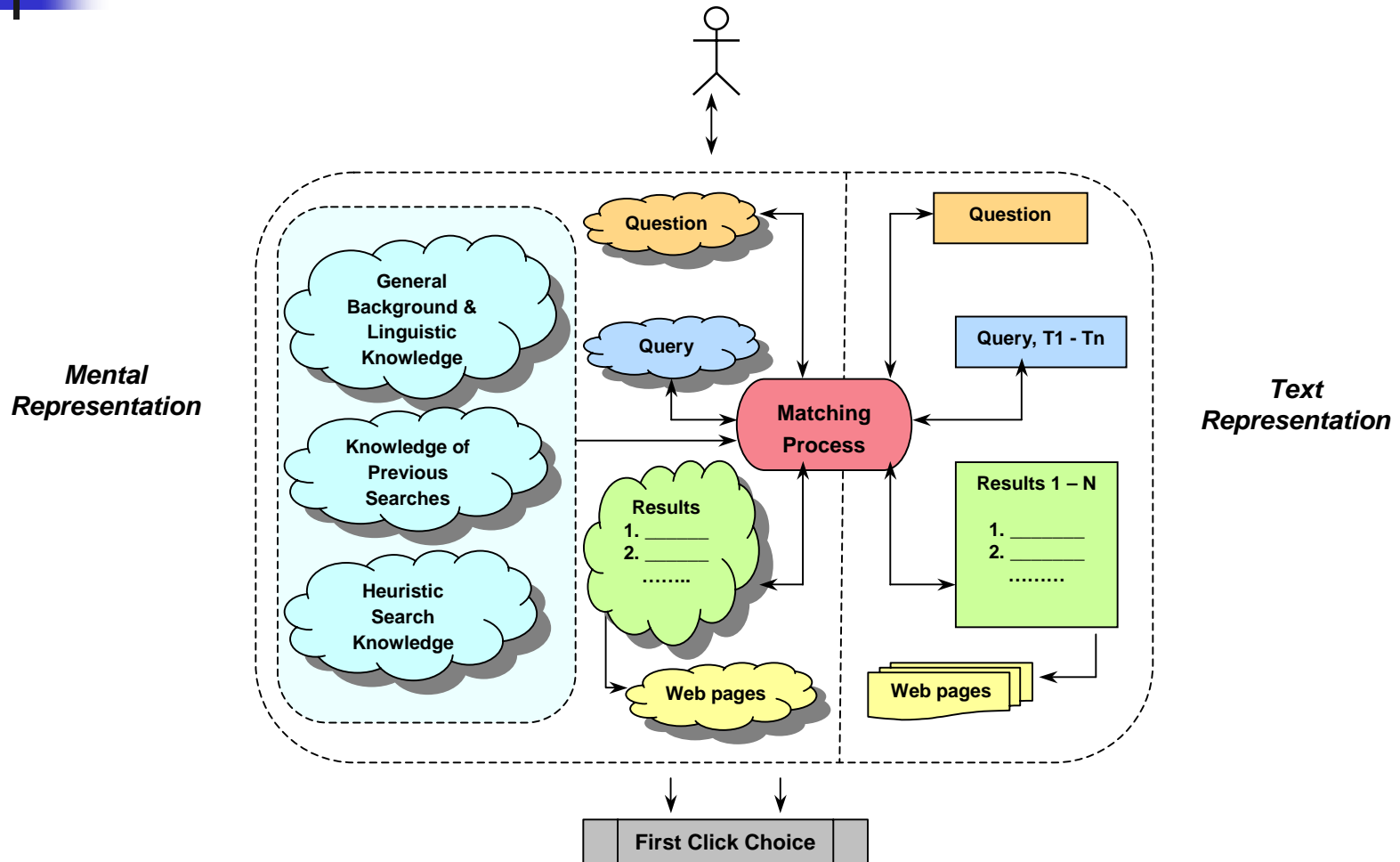
- Hits Domain

- Only web pages selected by users
- Scent distribution slightly improved : Mean 0.31, Standard Deviation 0.11
- Correlation of scent x hits still weak (- 0.11)

- Query Domain

- Only pages returned to a given query
- Chose 4 queries at random
- Majority of scent values < 0.2
- Almost no correlation between scent & hits

First-Click Framework





First-Click Framework

- Problem: Do Not Have Good Techniques for Characterizing User Knowledge
- Need:
 - Cognitive model of First-Click behaviour
 - Good proxies for user knowledge
- Current Proxies of User Knowledge
 - Link-structure analysis : Google
 - Community-based hits analysis : I-SPY
 - Corpus analysis : LSA
 - Term-frequency analysis : Bloodhound



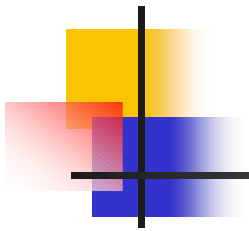
Current Work

- New Predictive Model of First-Click Behaviour
- Working Back from the Data
 - Position is important
 - Specific components of a result are important
 - Term order is important
- Simple String-Matching Model
 - Textual analysis
 - Position probabilities



Conclusions

- Current Scent Predictions Don't Work
 - CWW
 - Limited vocabulary
 - Appears to succeed by string matching
 - Bloodhound
 - No correlation between scent & hits even with a specialized domain
- On a More Positive Note:
 - General framework of 'First-Click' behavior
 - New predictive model



Thank You

Questions ?