# Personalizing Search: Promise & Pitfalls

Author: Jay Patel, Chief Science Officer, ChoiceStream*

Notes prepared for remarks to *Search Engine Meeting*, Boston, 23 April 2007

*ChoiceStream develops personalization solutions for online entertainment services, online portals, search services, eRetailers, and TV/video providers. ChoiceStream's patent-pending technologies helps firms increase customer retention and boost sales by reliably surfacing the most relevant content and products to each consumer. For additional information, please visit www.choicestream.com.

**Table of Contents**

0 Introduction ........................................ 1
1 How ChoiceStream Personalizes Search ... 1
  1.1 PageRank Personalization ................ 3
2 Four complications.................................... 3
  2.1 Stated preference v Action ............... 3
  2.2 Non-random sampling ...................... 4
  2.3 Explaining v Influencing .................. 4
  2.4 Chimera of T.O.E. ............................ 4
3 User Experience ......................................... 5
  3.1 Less clutter, usable control ............... 5
  3.2 Low latency ...................................... 5
4 Conclusion ................................................. 5

## 0 Introduction

In the 1990s, on-line search developed to surface relevant content from the exponentially growing world-wide-web. Its initial efforts employed mechanical matching between a user's search query and indexed words of web-pages. It accommodated compound and fuzzy queries as well as manually created taxonomies.

By the early 2000s, search development focused on improved sorting of the results from these mechanical matches. Google's PageRank method emerged as the winner. And what a spectacular commercial success that proved to be once it was married to sponsored-link auctions – Google's market capitalization reaches $150Bn!

Circa 2007, the challenging frontier for search is to broach the chasm to personalized relevance. Search results meaningful to searcher A are often different from those for B. New solutions include the deploying of query-relevant controls that allow the user to reorder results to fit personal intentions and interests.

The new search paradigm aims not only to allow the individual to actively shape results but also to infer the individual's tastes. With reliable inference of interests and intentions in hand, search results can be personalized passively and ubiquitously. Reliable passive personalization has been proven in domains such as video and music.

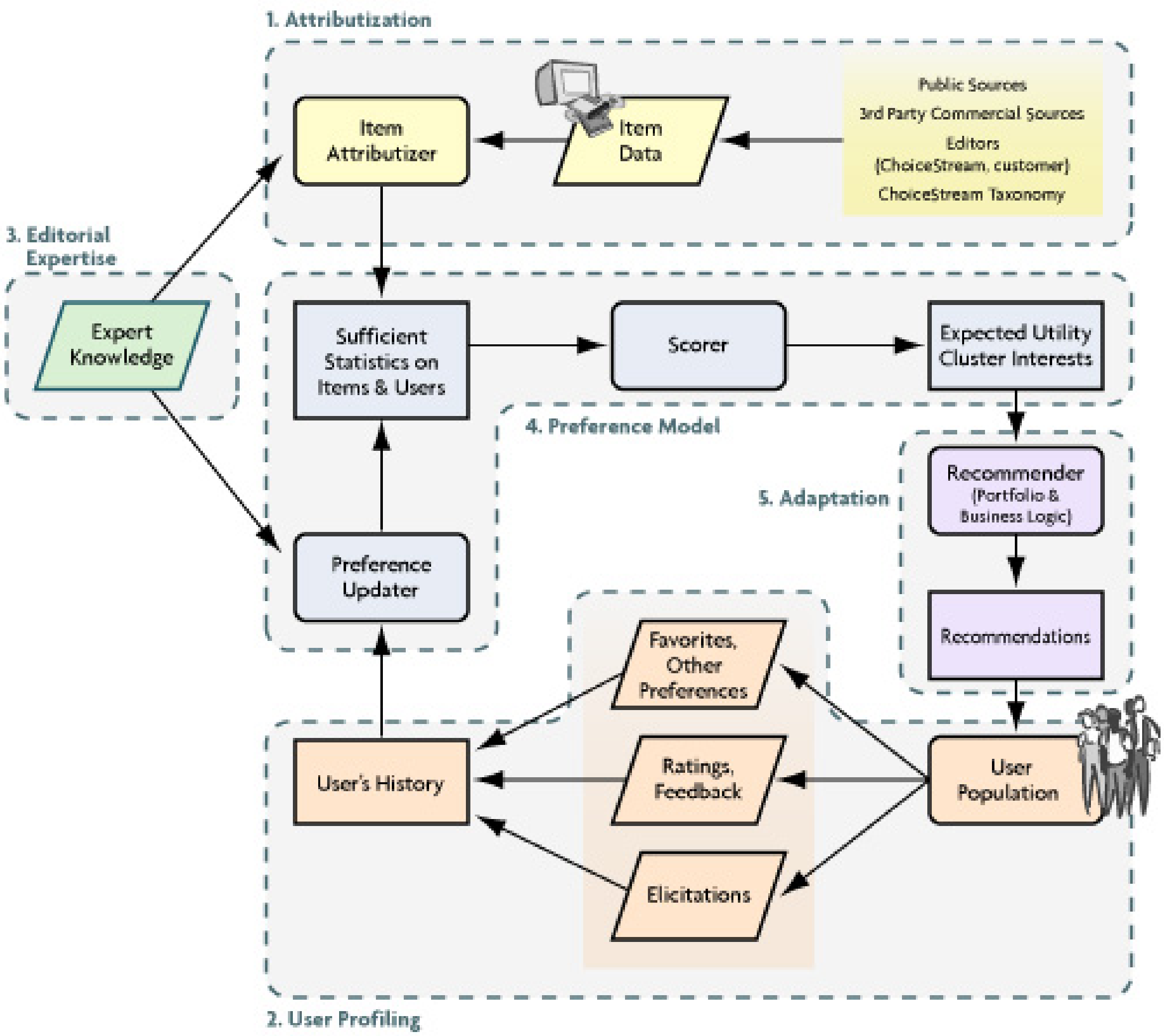
I share lessons from ChoiceStream's development toward this new paradigm, drawing on general search as well as domain-specific search and eCommerce. My remarks touch upon three topics:

1. How ChoiceStream personalizes search
2. Practical complications
3. Aspects of user experience

## 1 How ChoiceStream Personalizes Search

ChoiceStream personalizes search by extending its RealRelevance™ platform that has successfully personalized entertainment recommendations [movies, TV, music], targeted sponsored links, and informed shopping [cross-sells/up-sells/promotions].

**Figure. Structure of ChoiceStream's RealRelevance™ Platform**




ChoiceStream's general framework, dubbed ABCM for "Attributized Bayesian Choice Modeling," is currently patent-pending. Its components in the context of personalizing search are delineated below, where the numbering cross-references the blocks in the figure above:

1. **Attribute-based understanding of Web pages** (each page is auto-classified according to categories, topics, etc., and is scored on subtle attributes such as site functions, styles, and demographic appeal). Additional attributes exploit:
   - **'Global' relevance rank of the standard search engine** (Google's rank, for example, would be a baseline measure of value for an anonymous user)
   - **Nature of advertisers sponsoring keyword** (infers characteristics of sponsored-link web-pages associated with a query)
2. **User profiling** (domain-specific preferences, demographics, cohort's click-through behaviors, intentions from stack of user's recent searches, interests from recent visits/clicks)

3. **Editorial and third-party information (open directory project [dmoz], folksonomies [wisdom of crowds])** (taxonomies, when available, are usable for webpages; they also provide estimation and validation data for attributization writ large)
4. **Preference scorer** (model structure and parameter values are calibrated to feedback from actual usage/activity of search; challenges include choicesets, non-random sampling, cohort identification, and possible divergence between stated preferences and actions)
5. **Adaptation: UI, contextual sliders, portfolio mix** (supports both passive and user-managed personalization; balances portfolio of results, and displays context-relevant interactive controls)

The 'output' of ChoiceStream's personalized search is a ranking or targeting by predicted relevance for:

- **Search results**, showing preferred styles, meanings, categories with favorites placed on top. These are often presented under a "My Results" tab so as not to displace standard results.
- **Editorial sites**, choosing the most likely to be interesting subset from a pool of host-selected recommended links.
- **Clusters/taxonomy-nodes**, showing domains of likely interest in context of query
- **Sponsored links**, maximizing expected revenue by estimating likelihood of a click-through for a particular user on a set of links.
- **Advertising**, achieving greater relevance and click-through by surfacing ads based on the user's interests as well as the user's current search activity.
- **Targeted paid inclusion**, making paid listings more valuable by targeting inclusions to best prospects and thereby minimizing 'wasted eyeballs'.

### 1.1 PageRank Personalization

In contrast to ChoiceStream's approach, personalization within the Google-pioneered page-rank framework proceeds by modifying the implicit random-jump vectors that underlie its computation of webpage importances. For example, a user interested in sports gets modeled as having interest in intermittent jumps to sites/pages like http://www.espn.com, while a user interested in current events gets modeled as having interest in intermittent jumps to pages like http://www.nytimes.com. In practice, personalizing the random-jump vectors operates with topics or hosts rather than individual pages (– see discussion and references in: Kamvar, S., Haveliwala, T., Manning, C., Golub, G., 2003, "Exploiting the Block Structure of the Web for Computing PageRank", working paper, Stanford University).

Unsurprisingly, a different starting point leads to a quite different personalization approach. We don't expect one approach to dominate everywhere but rather that that one approach will prove better than another in a given context or domain. Any purely analytical procedure to rank the competing personalization approaches will make so many simplifying assumptions that it won't be trustworthy. The contextually relevant winners (including possibly combining approaches) get discovered by experiment and by careful assessment in the field rather than by belief and assertion.

## 2 Four complications

I turn to four complications that complicate personalization of search in practice:

1. Stated preference v Action
2. Non-random sampling
3. Explaining v Influencing
4. Chimera of T.O.E.

### 2.1 Stated preference v Action

To illustrate, consider personalizing search for music. In elicitation, users happily announce considerable interest for such features as complex or critically-acclaimed music, alternative music, discovery, etc. In practice, they much prefer to consume easy listening, familiar songs, and mainstream artists. Not attending to this wedge between stated interests and behavior can lead to over-reliance of search results on elicitation that the user will experience as unsatisfactory. This wedge is reliably known to marketers and policy makers, and manifests itself in many domains. In our experiments and analyses, we verify that it applies to search too! Fortunately, the biases can be substantially corrected for and so elicitation information remains valuable.

*Bottom line: Elicitations indicate interest, but are not simple anchors.*

## 2.2 Non-random sampling

Consider the illustrative case of personalizing search for movies. Sites like Yahoo!Movies, Blockbuster, and Netflix gather user's ratings and employ them to personalize displays of movie content. With great fanfare in 2006, Netflix announced a $1M challenge to world-wide researchers to improve predictions of hold-out ratings given a set of estimation ratings from real users at their site. Will advances in this task per se enable Netflix to significantly improve personalizing movies? Only a wee bit!

My skepticism can be readily understood by the following analogous task: find a good book for me. You have my ratings on some books that I have read in the recent past, which ratings prove to be perfectly correlated with the books' ratings at Amazon. If you only focus on modeling my ratings, you will recommend books that are highest rated at Amazon. But you will overlook the fact that I only chose to rate/read psychometrics and data-mining books. So most of the books that you will find for me, which will be outside psychometrics and data-mining albeit highest rated at Amazon, will be poor choices. In fact, a model that ignores my ratings but accurately models what I chose to rate will do much better – it will find psychometrics and data-mining books that I have not yet read.

For books and video (and other entertainment domains), our experience is that ratings of folks are far from randomly distributed across the choicesets. Returning to the Netflix challenge, assuming that we only rely on ratings data, the better single task is to predict what a person chooses to rate. Even better, of course, is to do both: predict accurately what items each subscriber chooses to rate from the large set of available items, and predict the subscriber's ratings conditional upon the few items that are chosen to rate. We find that models which do both are indeed superior performers in practice.

*Bottom line: Good modeling accounts for the information-generating process.*

## 2.3 Explaining v Influencing

Personalized search modeling draws on analyses of data gathered on past search activity. Often, only information on clicked items is logged but details of all the results that were presented but not acted upon are missed. And, often, all historical results have been generated by one core algorithm. In such cases, a new model that fits the logged data well may be unreliable for predicting what happens when search results in the future are, say, powered by that very model.

Essentially, we need to account for the context in which users made their choices, and users must be subjected to different types of choicesets such that their tastes are adequately explored. These considerations are well understood by pharmacological researchers whose gold standard is wide-ranging double-blind randomized trials with contending treatments. The cost of applying this gold-standard approach, fortunately, is significantly less for search developers than for our pharmaceutical brethren. Yet we often shy away from the necessary investments, relying on hope and prayer that superior analysis will somehow compensate for flawed/incomplete experiments.

*Bottom line: Progress requires randomized trials that adequately span tastes and settings.*

## 2.4 Chimera of T.O.E.

Researchers know to generalize. But there can be too much of this good thing. We must resist the chimera of a theory of everything (T.O.E.).

Good scientists understand this, as do good physicians and good engineers. Just as personalization springs from our understanding that one size does not fit all, so should our approaches recognize that different domains and contexts will often be better served by different approaches.

So, personalization solutions that power general search-solution providers (Google, Yahoo!, MSN, AOL) will differ in subtle ways from those with domain-specialized needs. (The latter include music-to-download [iTunes, eMusic, Kazaa], video-to-download [Movielink, YouTube, Comcast OnDemand], community-to-socialize [MySpace, Facebook, LinkedIn, Bebo], products-to-buy [Amazon, Overstock, Victoria's Secret Direct].)

*Bottom line: No T.O.E., not a problem: understand domains, conquer contexts.*

# 3 User Experience

Usability matters. So too does allowing for non-inferable intentions of the user that can trump the medium-term-persistent inferable interests. Taken together, these imply that personalizing search must solve the challenges of user interface, user control, and user experience.

## 3.1 Less clutter, usable control

Google's design, validated by many others through usability testing, shows that users value a clean uncluttered page of search results. When introducing personalization controls, our usability studies also find that "less is usually more". For example, a single slider to manage ordering of results often proves more successful than multiple sliders that offer richer control but sacrifice intuitive handling. Similarly, the two ends of each slider must be easily understandable. This often leads to the best poles for a slider being "more" v "fewer".

## 3.2 Low latency

Usability tests reveal that significant dissatisfaction arises if search results take much more than 100 milliseconds to surface. So, personalization strategies must ensure that they can deliver efficient large-scale performance. Good personalization schemes will be amenable to efficient caching. And, they will address the inevitable cases when a search query does not hit the cache.

# 4 Conclusion

ChoiceStream personalizes search by building on its proven RealRelevance™ platform. It attributizes content; it models user's interests and intentions by domain; and it attends to critical complexities of sampling, context, and usability. We heed Wyatt Earp: "Fast is fine, but accuracy is everything."