HUMANE INFORMATION SEEKING:
GOING BEYOND THE IR WAY

JIN YOUNG KIM @ IBM RESEARCH
Information seeking requires a communication.

You need the **freedom** of expression.
You need someone who **understands**.
Information Seeking circa 2012

Search engine accepts **keywords** only.
Search engine doesn’t understand **you**.
Toward Humane Information Seeking

Rich User Modeling
Profile
Context
Behavior

Rich User Interactions
Search
Browsing
Filtering
Challenges in Rich User Interactions

Enabling rich interactions
Evaluating complex interactions
Challenges in Rich User Modeling

Profile

Context

Behavior

Representing the user

Estimating the user model
The HCIR Way:

Interaction History

**USER**

- Action
- Action
- Action

**SYSTEM**

- Response
- Response
- Response

**User Model**

- Profile
- Context
- Behavior

Filtering / Browsing
Relevance Feedback
...

Filtering Conditions
Related Items
...

Providing **personalized results** vs. **rich interactions** are complementary, yet both are needed in most scenarios.
The Rest of Talk...

**Personal Search**
- Improving search and browsing for known-item finding
- Evaluating interactions combining search and browsing

**Book Search**
- Analyzing interactions combining search and filtering

**Web Search**
- User modeling based on reading level and topic
- Proving non-intrusive recommendations for browsing
PERSONAL SEARCH

Retrieval And Evaluation Techniques for Personal Information [Thesis]
Why does Personal Search Matter?

Knowledge workers spend up to 25% of their day looking for information. – IDC Group
Example: Search over Social Media

Evaluating Search in Personal Social Media Collections [WSDM12]
Characteristics of Personal Search

- Many document types
- Unique metadata for each type
- Users mostly do re-finding\(^1\)
- Opportunities for personalization
- Challenges in evaluation

Most of these hold true for enterprise search!
Building a Query Model for Structured Docs [ECIR09,12]

**Field Relevance**

- Different field is important for different query-term

Why don’t we provide field operator or advanced UI?
Estimating the Field Relevance

• If User Provides Feedback
  • Relevant document provides sufficient information

• If No Feedback is Available
  • Combine field-level term statistics from multiple sources

\[
\text{Collection} \quad + \quad \text{Top-k Docs} \quad \cong \quad \text{Relevant Docs}
\]
Retrieval Using the Field Relevance

• Comparison with Previous Work

\[ q_1 \quad q_2 \quad \ldots \quad q_m \]

\[ f_1 \quad w_1 \]

\[ f_2 \quad w_2 \]

\[ \ldots \quad \ldots \]

\[ f_n \quad w_n \]

\[ P(F_1 | q_1) \quad P(F_1 | q_m) \]

\[ P(F_2 | q_1) \quad P(F_2 | q_m) \]

\[ \ldots \quad \ldots \]

\[ P(F_n | q_1) \quad P(F_n | q_m) \]

\[ \text{sum} \]

\[ \text{multiply} \]

• Ranking in the Field Relevance Model

\[
Score(D, Q) = \prod_{i=1}^{m} \sum_{j=1}^{n} \hat{P}(f_j | q_i, R) \times \text{Per-term Field Score} \times \text{Per-term Field Weight}
\]
Evaluating the Field Relevance Model

- **Retrieval Effectiveness** *(Metric: Mean Reciprocal Rank)*

<table>
<thead>
<tr>
<th></th>
<th>DQL</th>
<th>BM25F</th>
<th>MFLM</th>
<th>FRM-C</th>
<th>FRM-T</th>
<th>FRM-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREC</td>
<td>54.2%</td>
<td>59.7%</td>
<td>60.1%</td>
<td>62.4%</td>
<td>66.8%</td>
<td>79.4%</td>
</tr>
<tr>
<td>IMDB</td>
<td>40.8%</td>
<td>52.4%</td>
<td>61.2%</td>
<td>63.7%</td>
<td>65.7%</td>
<td>70.4%</td>
</tr>
<tr>
<td>Monster</td>
<td>42.9%</td>
<td>27.9%</td>
<td>46.0%</td>
<td>54.2%</td>
<td>55.8%</td>
<td>71.6%</td>
</tr>
</tbody>
</table>

Fixed Field Weights | Per-term Field Weights
Summary so far...

- **Query Modeling for Structured Documents**
  - Using the estimated field relevance improves the retrieval
  - User’s feedback can help personalize the field relevance

- **What’s Coming Next**
  - Alternatives to keyword search: associative browsing
  - Evaluating the search and browsing together
What if keyword search is not enough?

Search first, then browse through documents!
Building the Associative Browsing Model

[CIKM10,11]

1. Document Collection

2. Link Extraction

3. Link Refinement

Click-based Training
Evaluation Challenges for Personal Search
[CIKM09, SIGIR10, CIKM11]

• **Previous Work**
  - Each based on its own user study
  - No comparative evaluation was performed yet

• **Building Simulated Collections**
  - Crawl CS department webpages, docs and calendars
  - Recruit department people for user study

• **Collecting User Logs**
  - *DocTrack*: a human-computation search game
  - *Probabilistic User Model*: a method for user simulation
DocTrack Game

System

Randomly choose two candidate documents

Skim though documents (15 seconds each)

Randomly pick one target document

Use keyword search to find the document

Generate a ranked list for keyword search

User

Target Item

Find It!

Item Not Found! You have 8 trials left.

You can skip to the next item or move back to the previous page.
Probabilistic User Modeling

<table>
<thead>
<tr>
<th>Evaluation Type</th>
<th>Total</th>
<th>Browsing used</th>
<th>Successful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>63,260</td>
<td>9,410 (14.8%)</td>
<td>3,957 (42.0%)</td>
</tr>
<tr>
<td>User Study</td>
<td>290</td>
<td>42 (14.5%)</td>
<td>15 (35.7%)</td>
</tr>
</tbody>
</table>

Probabilistic user model trained on log data from user study.
Parameterization of the User Model

Query Generation for Search
- Preference for specific field

Link Selection for Browsing
- Breadth-first vs. depth-first

Evaluate the system under various assumptions of user, system and the combination of both
BOOK SEARCH

Understanding Book Search Behavior on the Web

[Submitted to SIGIR12]
Why does Book Search Matter?
Understanding Book Search on the Web

- **OpenLibrary**
  - User-contributed online digital library
  - DataSet: 8M records from web server log
Comparison of Navigational Behavior

- Users entering directly show different behaviors from users entering via web search engines

Users entering the site directly

Users entering via Google
Comparison of Search Behavior

Rich interaction reduces the query lengths
Summary so far...

- **Rich User Interactions for Book Search**
  - Combination of external and internal search engines
  - Combination of search, advanced UI, and filtering

- **Analysis using User Modeling**
  - Model both navigation and search behavior
  - Characterize and compare different user groups

- **What Still Keeps Me Busy...**
  - Evaluating the Field Relevance Model for book search
  - Build a predictive model of task-level search success [1]

[1] Beyond DCG: User Behavior as a Predictor of a Successful Search [Hassan10]
Web Search

Characterizing Web Content, User Interests, and Search Behavior by Reading Level and Topic

[WSDM12]
Myths on Web Search

• **Web search is a solved problem**
  - Maybe true for navigational queries, yet not for tail queries\(^1\)

• **Search results are already personalized**
  - Lots of localization efforts (e.g., query: pizza)
  - Little personalization at individual user level

• **Personalization will solve everything**
  - Not enough evidence in many cases
  - Users do deviate from their profile

\(^1\) Web search solved? All result rankings the same?

Need for rich user modeling and interaction!
User Modeling by Reading Level and Topic

• **Reading Level and Topic**
  • Reading Level: proficiency (comprehensibility)
  • Topic: topical areas of interests

• **Profile Construction**

• **Profile Applications**
  • Improving personalized search ranking
  • Enabling expert content recommendation
**Reading level distribution varies across major topical categories**

<table>
<thead>
<tr>
<th>Category</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
<th>R8</th>
<th>R9</th>
<th>R10</th>
<th>R11</th>
<th>R12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.17</td>
<td>0.10</td>
<td>0.15</td>
<td>0.04</td>
<td>0.02</td>
<td>0.03</td>
<td>0.20</td>
<td>0.27</td>
</tr>
<tr>
<td>Health</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.18</td>
<td>0.08</td>
<td>0.13</td>
<td>0.04</td>
<td>0.04</td>
<td>0.10</td>
<td>0.27</td>
<td>0.11</td>
</tr>
<tr>
<td>Science</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.23</td>
<td>0.09</td>
<td>0.07</td>
<td>0.02</td>
<td>0.01</td>
<td>0.08</td>
<td>0.27</td>
<td>0.17</td>
</tr>
<tr>
<td>Computers</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.24</td>
<td>0.19</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.32</td>
<td>0.12</td>
</tr>
<tr>
<td>Business</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.22</td>
<td>0.16</td>
<td>0.09</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
<td>0.26</td>
<td>0.12</td>
</tr>
<tr>
<td>Society</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.23</td>
<td>0.07</td>
<td>0.35</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.22</td>
<td>0.06</td>
</tr>
<tr>
<td>Adult</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.28</td>
<td>0.26</td>
<td>0.14</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
<td>0.13</td>
<td>0.06</td>
</tr>
<tr>
<td>Kids and Teens</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.23</td>
<td>0.26</td>
<td>0.13</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>Games</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
<td>0.36</td>
<td>0.10</td>
<td>0.11</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Recreation</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.11</td>
<td>0.44</td>
<td>0.19</td>
<td>0.08</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>Arts</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.08</td>
<td>0.40</td>
<td>0.27</td>
<td>0.10</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>Home</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.19</td>
<td>0.41</td>
<td>0.14</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>News</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.41</td>
<td>0.33</td>
<td>0.14</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Shopping</td>
<td>0.00</td>
<td>0.00</td>
<td>0.22</td>
<td>0.29</td>
<td>0.24</td>
<td>0.09</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.07</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Sports</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.56</td>
<td>0.11</td>
<td>0.10</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Profile matching can predict user’s preference over search results

**Metric**
- % of user’s preferences predicted by profile matching

**Results**
- By the degree of **focus** in user profile
- By the **distance metric** between user and website

<table>
<thead>
<tr>
<th>User Group</th>
<th>#Clicks</th>
<th>$KL_R(u,s)$</th>
<th>$KL_T(u,s)$</th>
<th>$KL_{RLT}(u,s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑Focused</td>
<td>5,960</td>
<td>59.23%</td>
<td>60.79%</td>
<td>65.27%</td>
</tr>
<tr>
<td>↓Diverse</td>
<td>147,195</td>
<td>52.25%</td>
<td>54.20%</td>
<td>54.41%</td>
</tr>
<tr>
<td>↓Diverse</td>
<td>197,733</td>
<td>52.75%</td>
<td>53.36%</td>
<td>53.63%</td>
</tr>
</tbody>
</table>
Comparing Expert vs. Non-expert URLs

- Expert vs. Non-expert URLs taken from [White’09]
Enabling Browsing for Web Search

- **SurfCanyon®**
  - Recommend results based on clicks

Initial results indicate that recommendations are useful for shopping domain.
Looking Onward
Summary: Rich User Interactions

- **Combining Search and Browsing for Personal Search**
  - Associative browsing complements search for known-item finding

- **Combining Search and Filtering for Book Search**
  - Rich interactions reduce user efforts for keyword search

- **Non-intrusive Browsing for Web Search**
  - Providing suggestions for browsing is beneficial for shopping task
Summary: Rich User Modeling

- **Query (user) modeling improves ranking quality**
  - Estimation is possible without past interactions
  - User feedback improves effectiveness even more

- **User Modeling improves evaluation / analysis**
  - Prob. user model allows the evaluation of personal search
  - Prob. user model explains complex book search behavior

- **Enriched representation has additional values**
Where’s the Future of Information Seeking?

Thank you! Any Questions?

@ct4socialsoft

Computer Human Interaction - it's not just about front-end usability; it's about designing a purposeful overall integrated client experience.
Selected Publications

• **Structured Document Retrieval**
  - A Probabilistic Retrieval Model for Semi-structured Data [ECIR09]
  - A Field Relevance Model for Structured Document Retrieval [ECIR11]

• **Personal Search**
  - Retrieval Experiments using Pseudo-Desktop Collections [CIKM09]
  - Ranking using Multiple Document Types in Desktop Search [SIGIR10]
  - Building a Semantic Representation for Personal Information [CIKM10]
  - Evaluating an Associative Browsing Model for Personal Info. [CIKM11]
  - Evaluating Search in Personal Social Media Collections [WSDM12]

• **Web / Book Search**
  - Characterizing Web Content, User Interests, and Search Behavior by Reading Level and Topic [WSDM12]
  - Understanding Book Search Behavior on the Web [In submission to SIGIR12]

More at @jin4ir, or cs.umass.edu/~jykim
Optional Slides
Bonus: My Self-tracking Efforts

- **Life-optimization Project** (2002~2006)
- **LiFiDeA Project** (2011-2012)
Topic and reading level characterize websites in each category.

Interesting divergence for the case of users.
The Great Divide: IR vs. RecSys

**IR**
- Query / Document
- Provide relevant info.
- Reactive (given query)
- SIGIR / CIKM / WSDM

**RecSys**
- User / Item
- Support decision making
- Proactive (push item)
- RecSys / KDD / UMAP

- Both requires similarity / matching score
- Personalized search involves user modeling
- Most RecSys also involves keyword search
- Both are parts of user’s info seeking process
Criteria for Choosing IR vs. RecSys

- Confidence in predicting user’s preference
- Availability of matching items to recommend

- User’s willingness to express information needs
- Lack of evidence about the user himself
The Great Divide: IR vs. CHI

IR
- Query / Document
- Relevant Results
- Ranking / Suggestions
- Feature Engineering
- Batch Evaluation (TREC)
- SIGIR / CIKM / WSDM

CHI
- User / System
- User Value / Satisfaction
- Interface / Visualization
- Human-centered Design
- User Study
- CHI / UIST / CSCW

Can we learn from each other?