



Towards Interactive Intelligent Search Assistants

交互式智能信息检索助手的研究

UMassAmherst

College of Information
& Computer Sciences
Center for Intelligent Information Retrieval



Keping Bi (毕可平)

**Institute of Computing Technology
Chinese Academy of Sciences**



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES



- **Research Interests**

- Information Retrieval, Search Personalization,
- Conversational Search, Dense Retrieval

- **Professional Experiences**

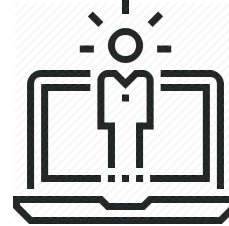
- *Applied Scientist II*
 - Email Search Team, MSAI, **Microsoft**
- *Senior Research Engineer (T5)*
 - LTR Team, Web Search Department, **Baidu**

- **Education**

- *Ph.D.*
 - CIIR, Computer Science, **University of Massachusetts Amherst**
- *M.S.*
 - Machine Intelligence, **Peking University**
- *B.E.*
 - Computer Science, **Nankai University**



Various Information Needs



“I want an **iphone 13 case**”



“What are the **methods to control type ii diabetes?**”

... losing a modest amount of weight by following a reduced-calorie eating plan and being physically active most days of the week.

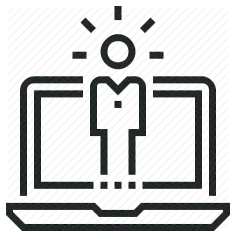
“Any **good places** recommended for **dinner?**”



Araya's Place Vegan Thai Bellevue

4.5 ★★★★★ 1,015 reviews · \$\$

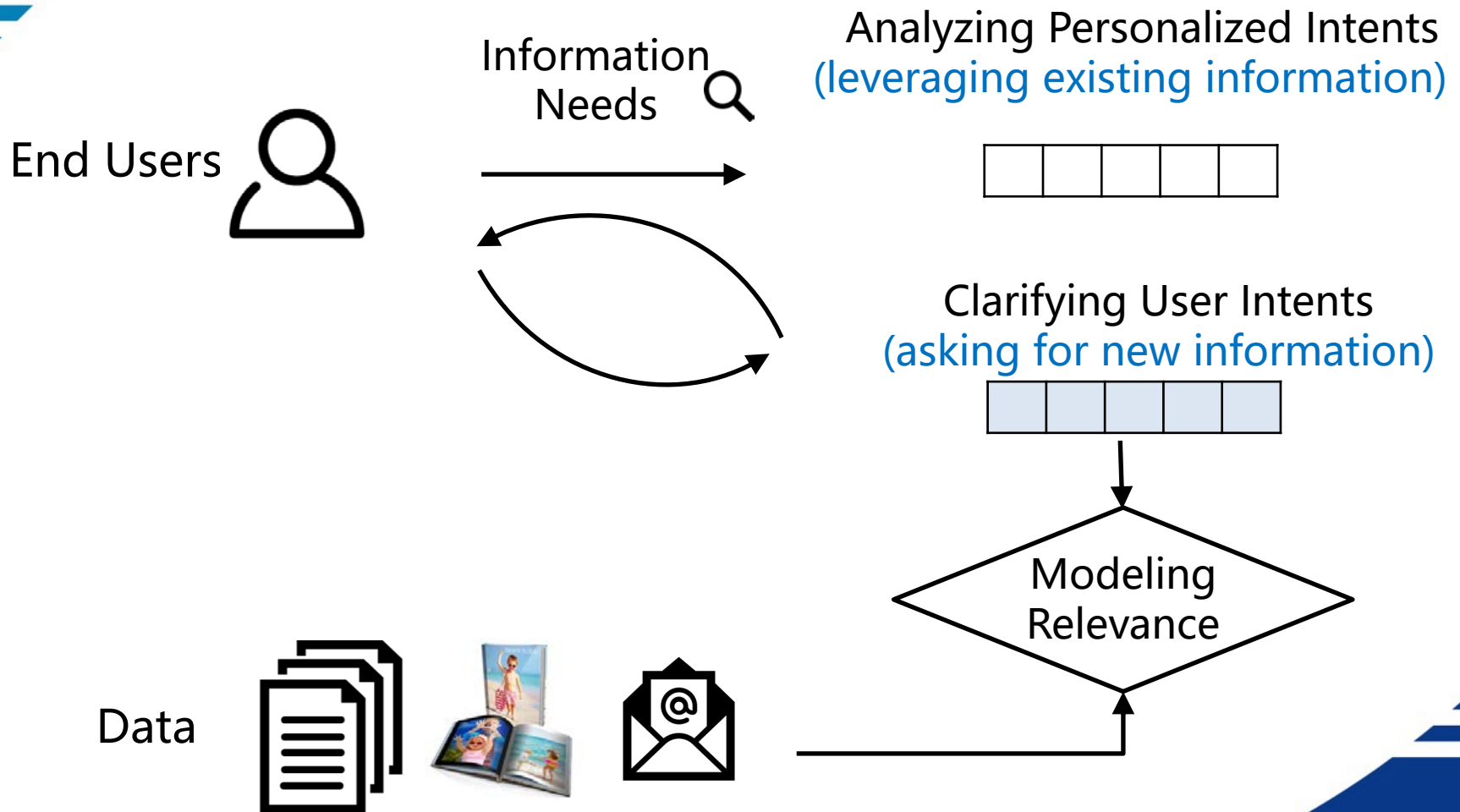
Vegan restaurant



- interact with users
- understand user intents
- provide ideal information



Interactive Search Assistants



Intelligent Search Assistant

Identifying Personalized Intents

- Identify personalized intents based on users' historical behavior data. [SIGIR'17, TOIS'19, SIGIR'20, SIGIR'21, WWW'21]

Clarifying User Intents

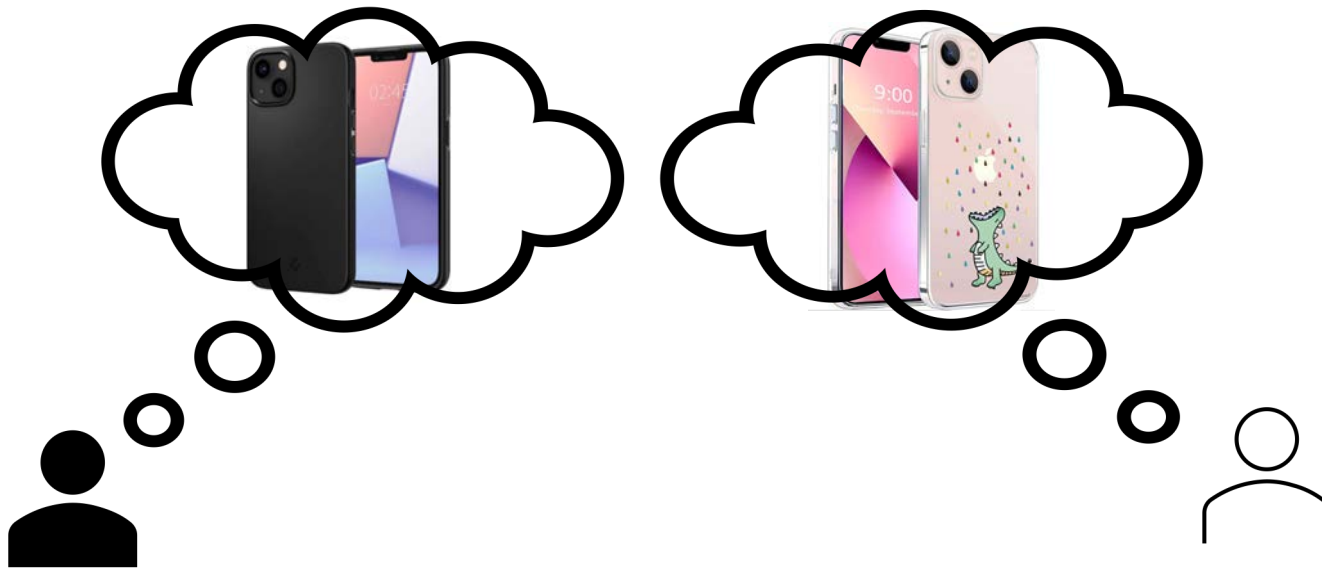
- Ask clarifying questions to understand their information needs. [WCIS'19, ECIR'19, ICTIR'21]

Modeling Relevance

- Analyze the relevance of candidate information to user intents. [eCommerce'19, CIKM'19a, CIKM'19b]

Part I: Personalized Intent Identification

[SIGIR'17, TOIS'19, SIGIR'20, SIGIR'21, WWW'21]



"I want an **iphone 13 case**"

Is Personalization Always Helpful?

"Toothpaste"



Sensodyne Pronamel Fresh Breath Enamel Toothpaste for Sensitive Teeth, Pack of 3

★★★★★ ~ 168

\$16⁹⁸ (\$1.42/Ounce)

I have sensitive teeth.



Colgate Max Fresh Whitening Toothpaste with Breath Strips, Clean Mint - 6 ounce (4 Pack)

★★★★★ ~ 1,992

\$7⁹⁸ (\$0.33/Ounce) \$13.96

I don't care!



Dynamic Personalization is Needed!



Part I: Personalized Intent Identification

[SIGIR'17, TOIS'19, SIGIR'20, SIGIR'21, WWW'21]

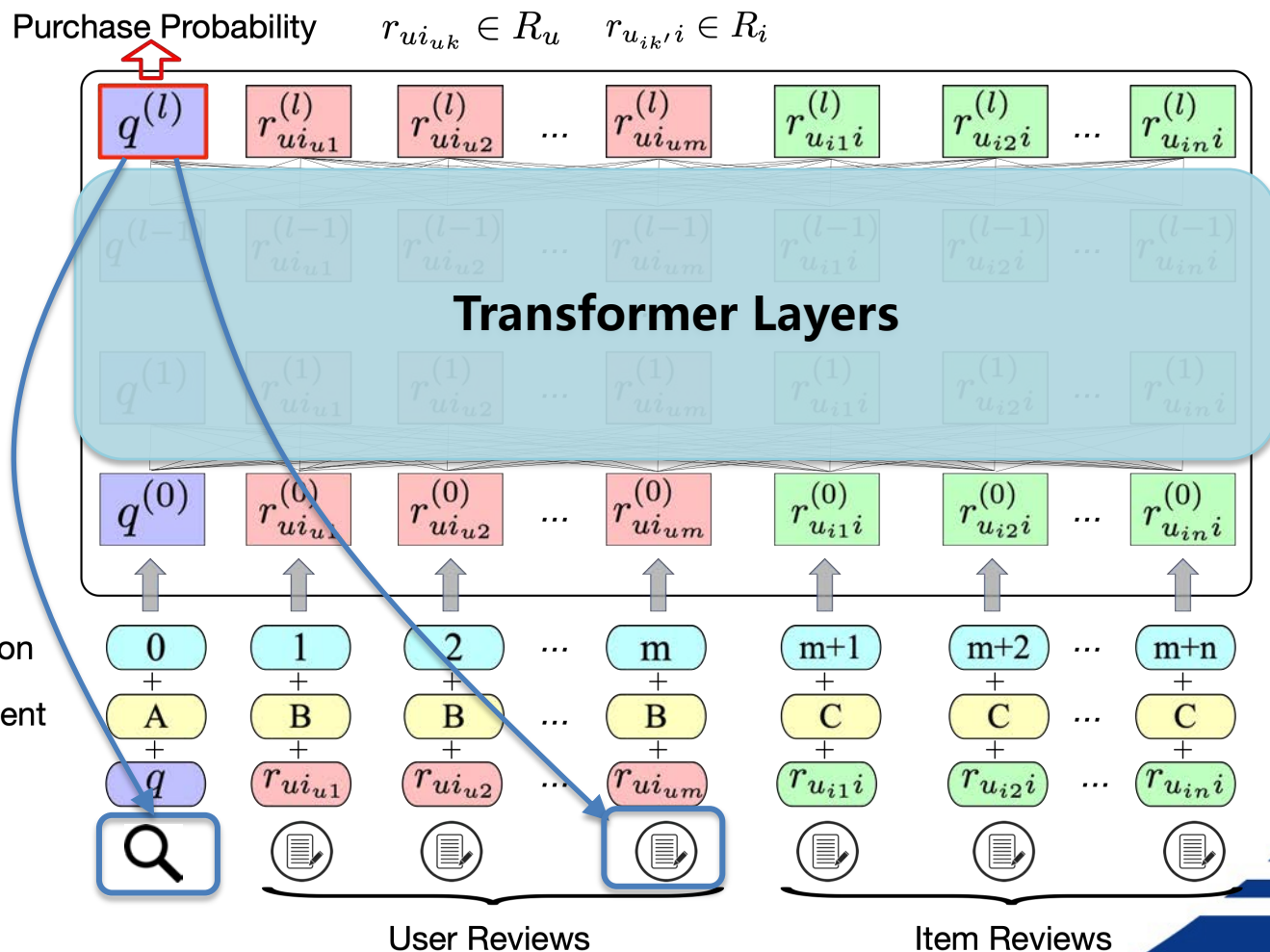


Long-term behaviors indicate user preferences

Dynamic Personalization

[SIGIR'20, SIGIR'21, WWW'21]

Attention Allocation



u:user

q:query

i:item

R_u : m reviews associated with u

$R_u = (r_{ui_{u1}}, r_{ui_{u2}}, \dots, r_{ui_{um}})$

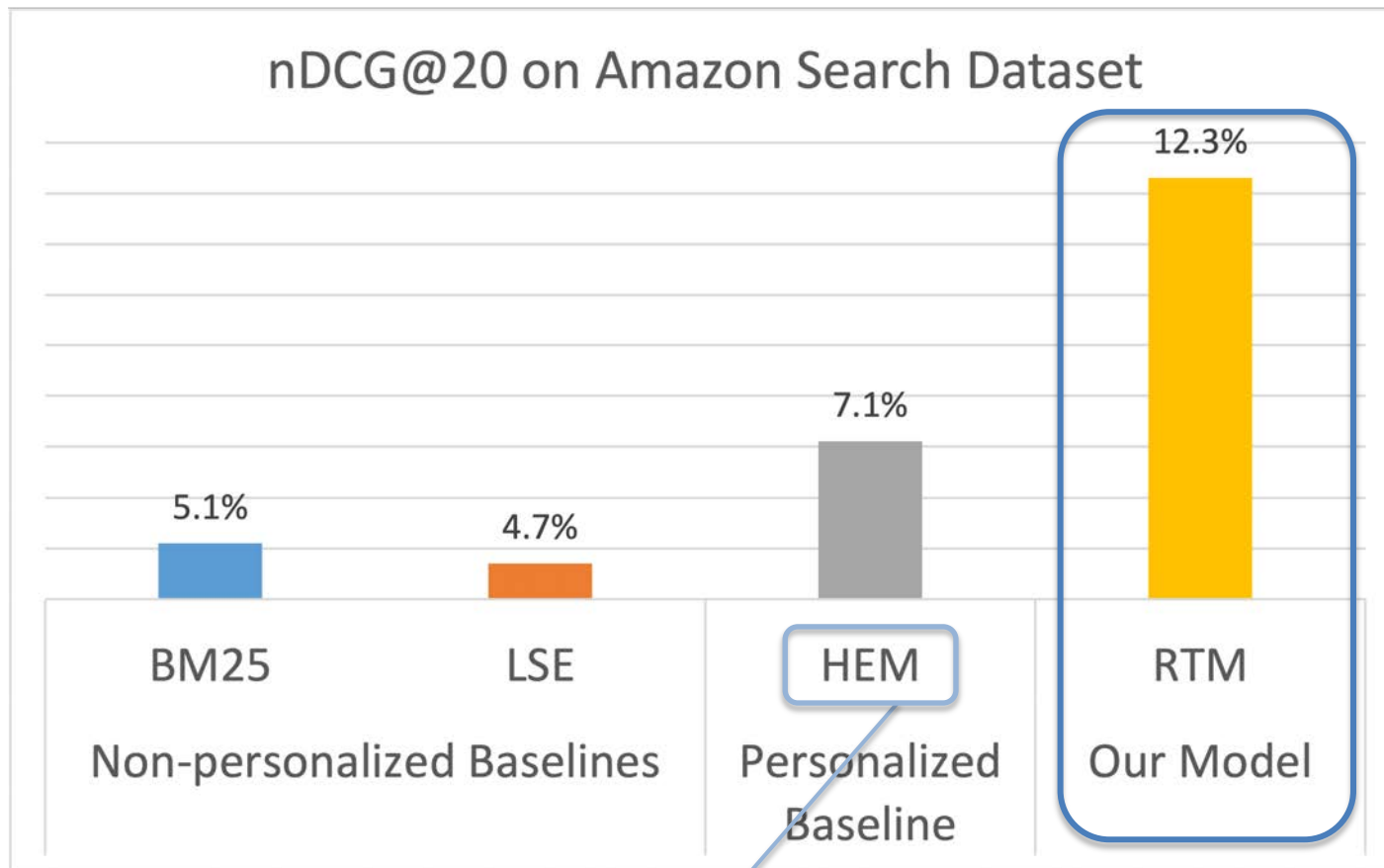
R_i : reviews associated with i

$R_i = (r_{u_{i1}i}, r_{u_{i2}i}, \dots, r_{u_{in}i})$



Part I: Personalized Intent Identification

[SIGIR'17, SIGIR'20, SIGIR'21]



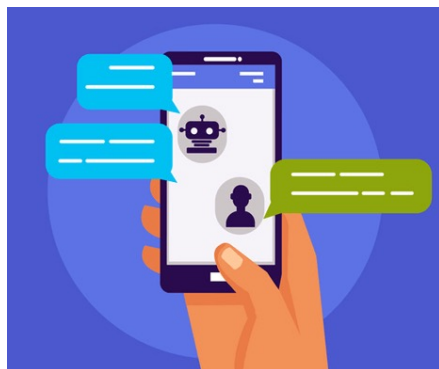
*The higher,
the better*

Static personalization

Part II: Clarifying User Intents

[WCIS'19, ECIR'19, ICTIR'21]

- Existing information may not be enough to identify user intents
- Interactions are feasible through conversations



Proactively asking clarifying questions

Asking Clarifying Questions

Query: *diversity*

Intent Description: *How is workplace diversity achieved and managed?*



Are you looking for a **definition** of diversity?

No. That's not what I want.



Are you looking for **educational materials** about diversity?



Yes, I need materials on achieving workplace diversity



Would you like the **legal definition** of diversity?

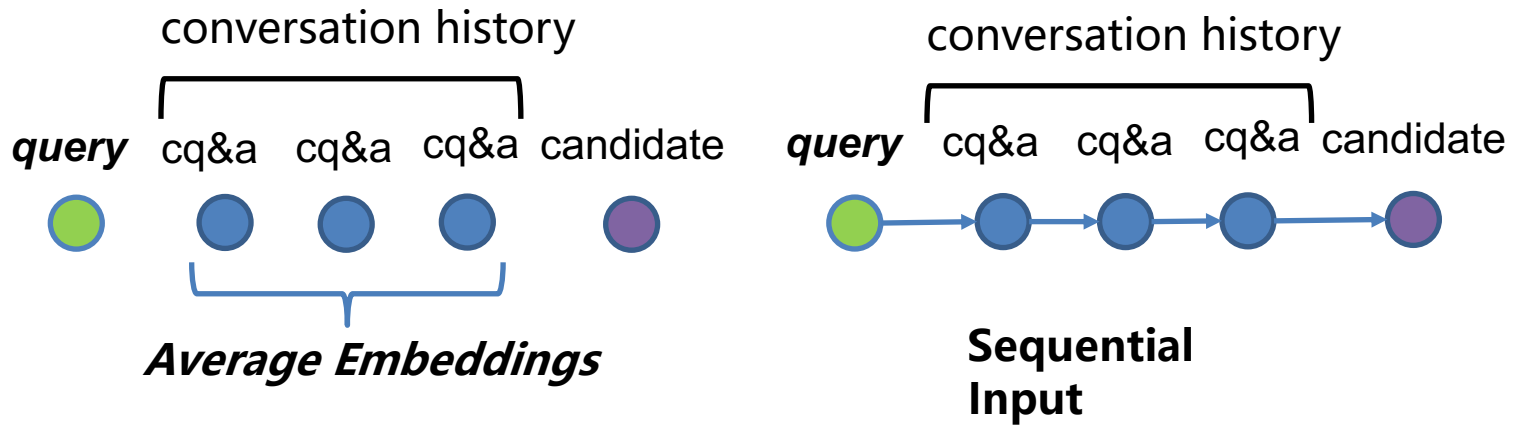


No.



Continue Asking

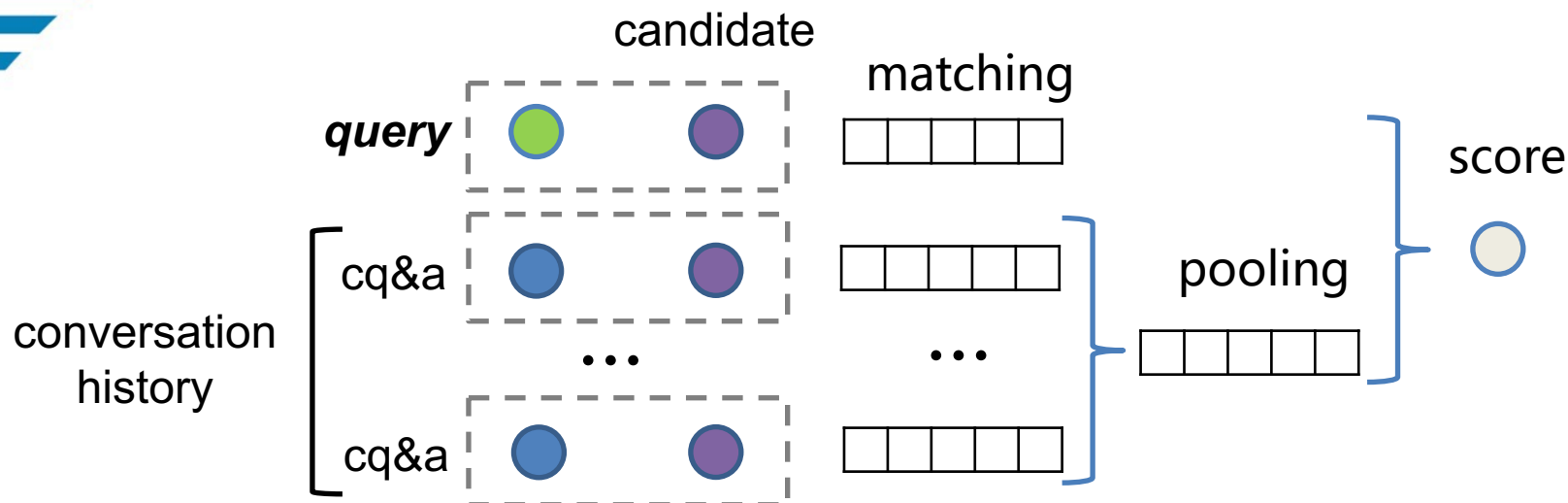
Previous Studies



Do not differentiate answers that are confirmation or denial

-  *query*
-  *cq &a: clarifying question and answer in each conversation turn*
-  *candidate clarifying question*

Asking Clarifying Questions

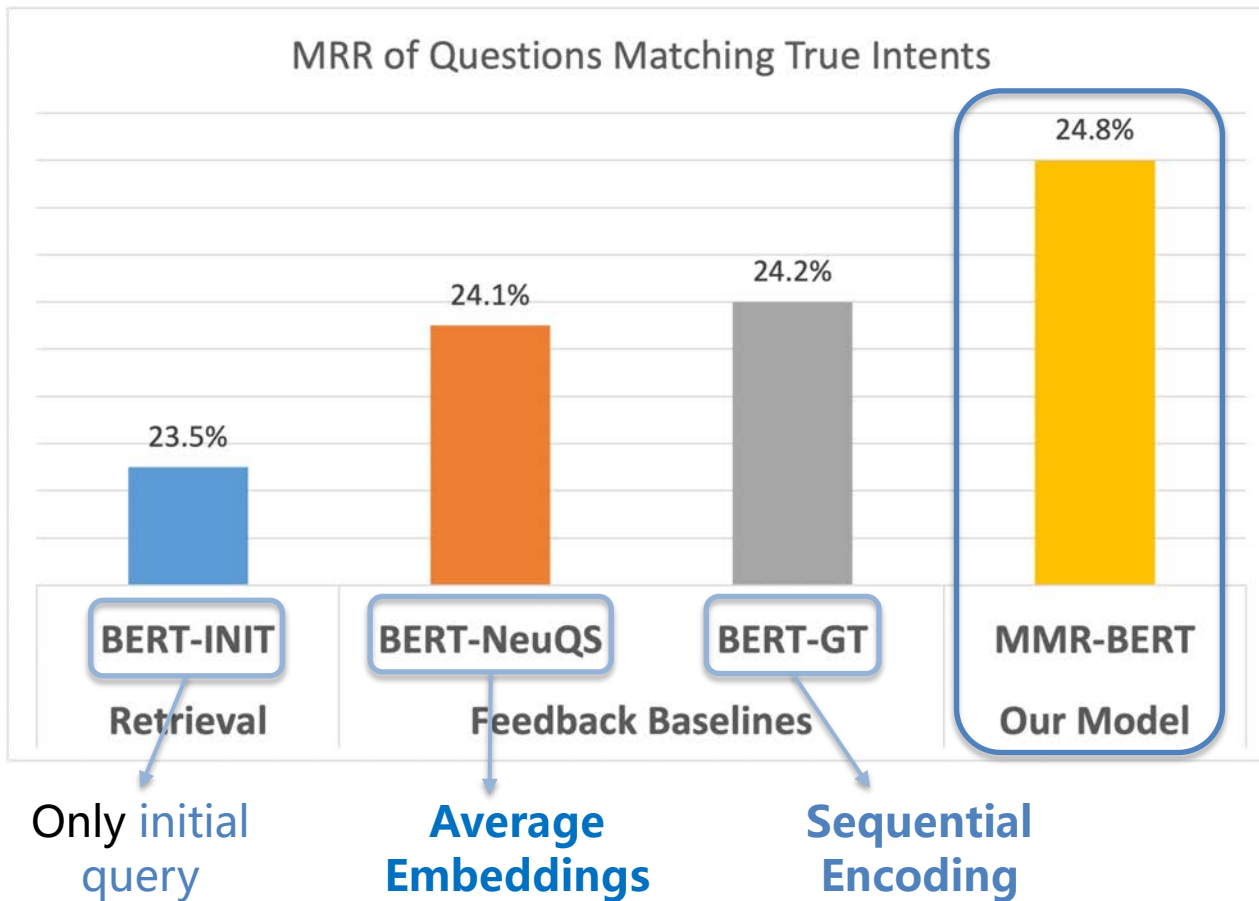


Capture confirmation and denial by individually matching

-  *query*
-  *cq &a: clarifying question and answer in each conversation turn*
-  *candidate clarifying question*

Asking Clarifying Questions

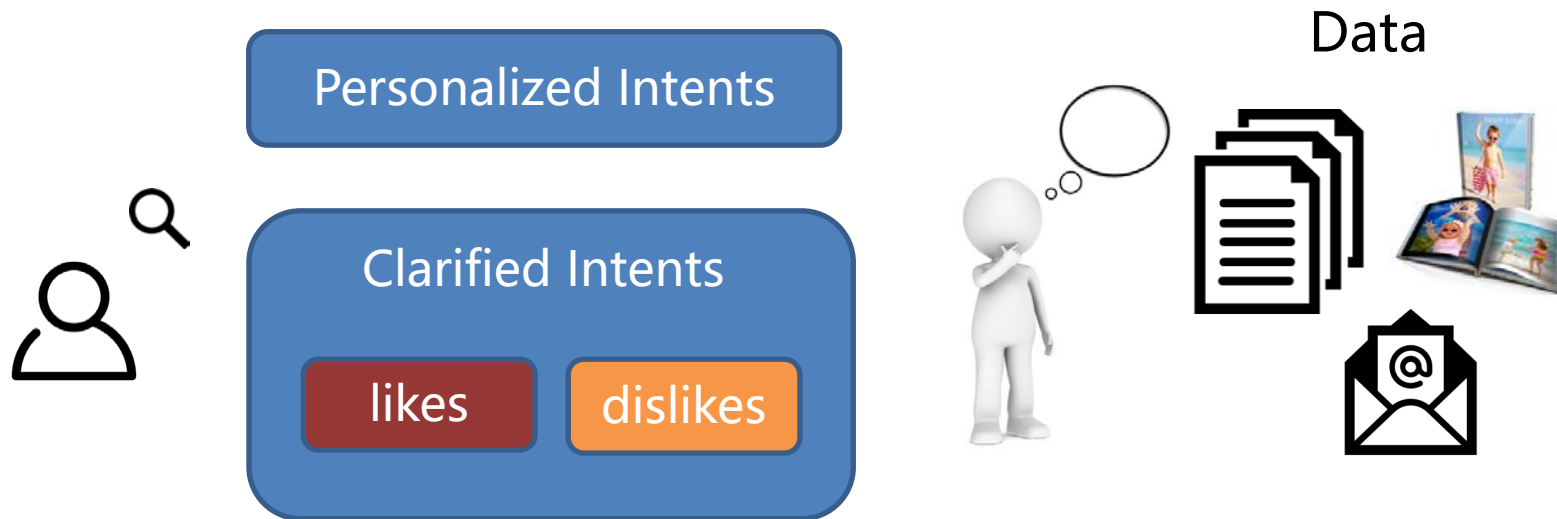
MRR (inverse rank of the clarifying questions that cover a user's true intents)



*The higher,
the better*

Part III: Relevance Modeling

[eCommerce'19, CIKM'19a, CIKM'19b]



**How do we measure information relevance
with respect to user intents?**

Relevance Modeling on Fine-grained Intents

"I want an **iphone 13 case**"



cover:
dark



material:
plastic



plastic:
soft



I want a **case** for **iPhone 13**.



Sure. Do you like these options?



I **don't like** these ones.



Do you like the **cover** to be **dark**?



No. I don't think so.



Do you like the **material** to be **plastic**?



Yes. That would be good!



Do you like the **plastic** to be **soft**?



I **don't** have any **preferences** on that.



Got it! What about these options?



Great! I want to buy the third one.

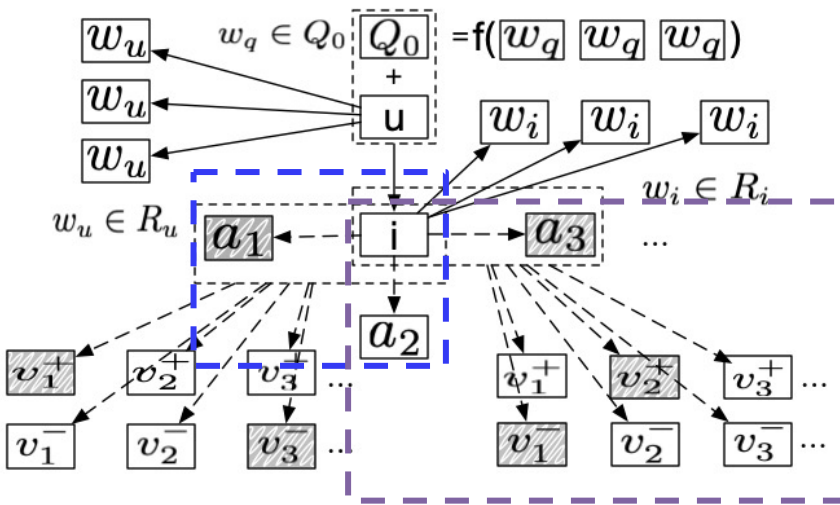


Cool. I'll place the order for you.

A new paradigm to collect fine-grained intents

Relevance Modeling on Fine-grained Intents

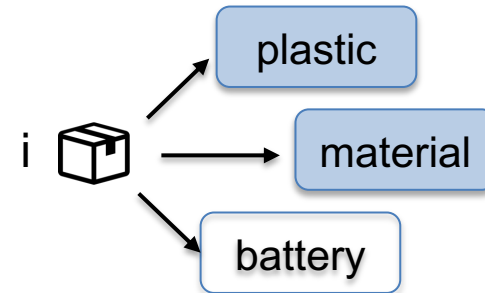
 occurrence v^+ positive values
 nonoccurrence v^- negative values



“likes” MB model

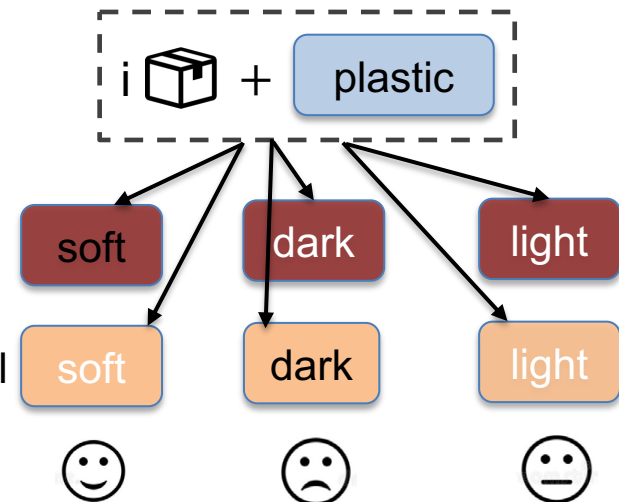
Aspect Generation Model

(Multivariate Bernoulli (MB) Distribution)



Value Generation Model

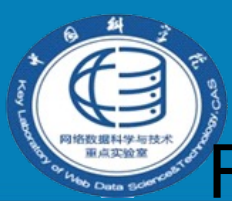
(Two Independent (MB) Models for Positive and Negative Values)



“dislikes” MB model

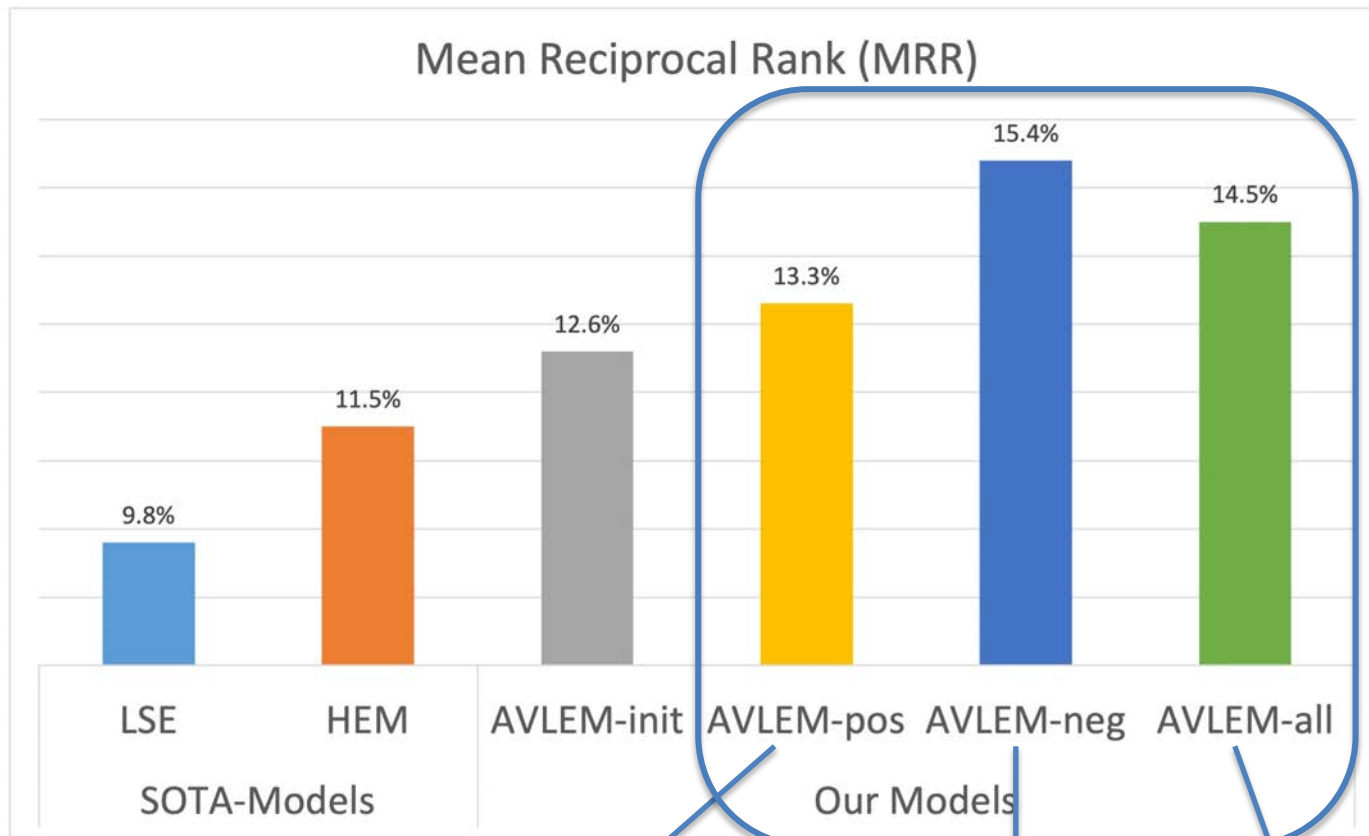
Q_0 : query
 u : user
 i : item
 a : aspect

w_u : word written by u
 w_i : word written for i
 v^+ : value with positive feedback
 v^- : value with negative feedback



Relevance Modeling on Fine-grained Intents

MRR (inverse rank of the ideal items) on Amazon Search Dataset



*The higher,
the better*

likes only

dislikes only

likes & dislikes



Intelligent Search Assistant

**Intelligent
Search
Assistant**

*Identifying
Personalized Intents*

Clarifying User Intents

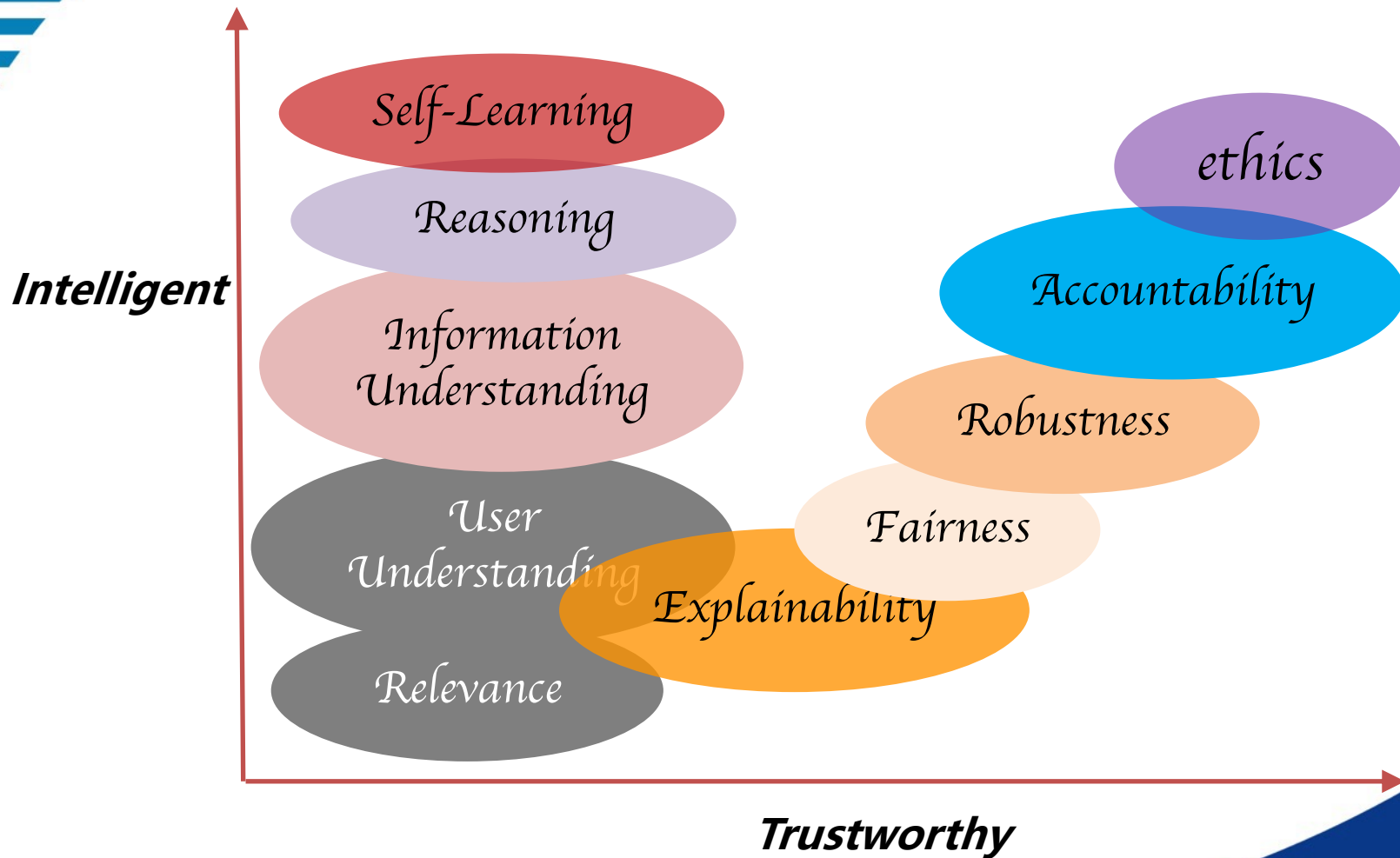
Modeling Relevance

*User
Understanding*

Relevance



Next Generation of Search Assistants





Next Generation of Search Assistants

- Towards Intelligent Information Understanding
 - Heterogeneous Information Representation
 - Expert Response

[TOIS'19,
Eval4NLP'20,
EACL'21]



What do you think of
Russia-Ukraine war?



According to *[DocID/Provider]*, XXX
Whereas *[DocID/Provider]* claims XXX.
Refer to *[VideoID]* for the discussion
between ...
Overall, *I suggest that* XXX.

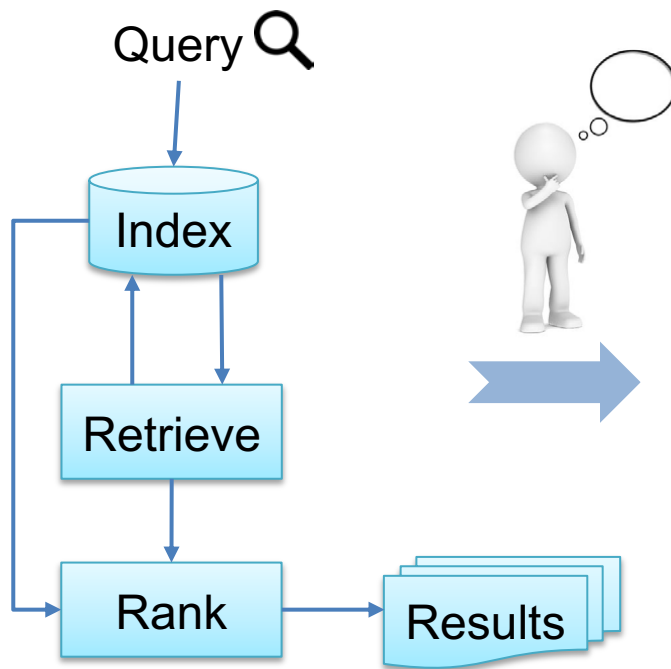
Rich Structure
Multi-modal
Multi/Cross-Lingual

Concise Answer with
References
Expert Advices



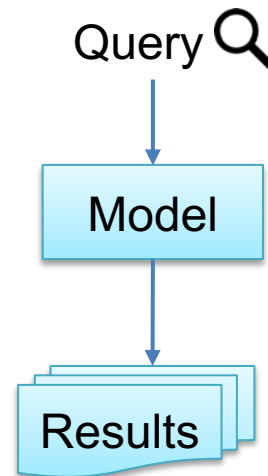
Next Generation of Search Assistants

- Towards Intelligent Model-based Retrieval
 - A Corpus Model Instead of Index



Index-Retrieve-then-Rank

Like Human Brain



Unified-Retrieve-and-Rank

Multi-modal/lingual

Response Generation

Multi-task Learning

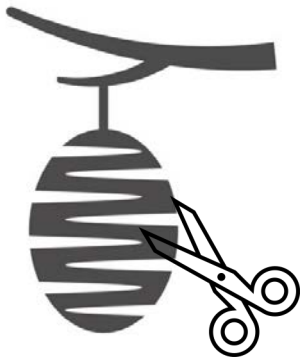
Model Interpretability

Efficiency & Scalability



Next Generation of Search Assistants

- Towards Trustworthy Information Assistants [SIGIR'18]
 - Provide Fair/Unbiased Information
 - Robust to Attack
 - Respect User Privacy



Information Cocoon

Fairness



Disinformation &
Clicks

Robustness



Sensitive Information
Records

Privacy



It's an honor to join the family 😊
Thanks!