RePair My Queries

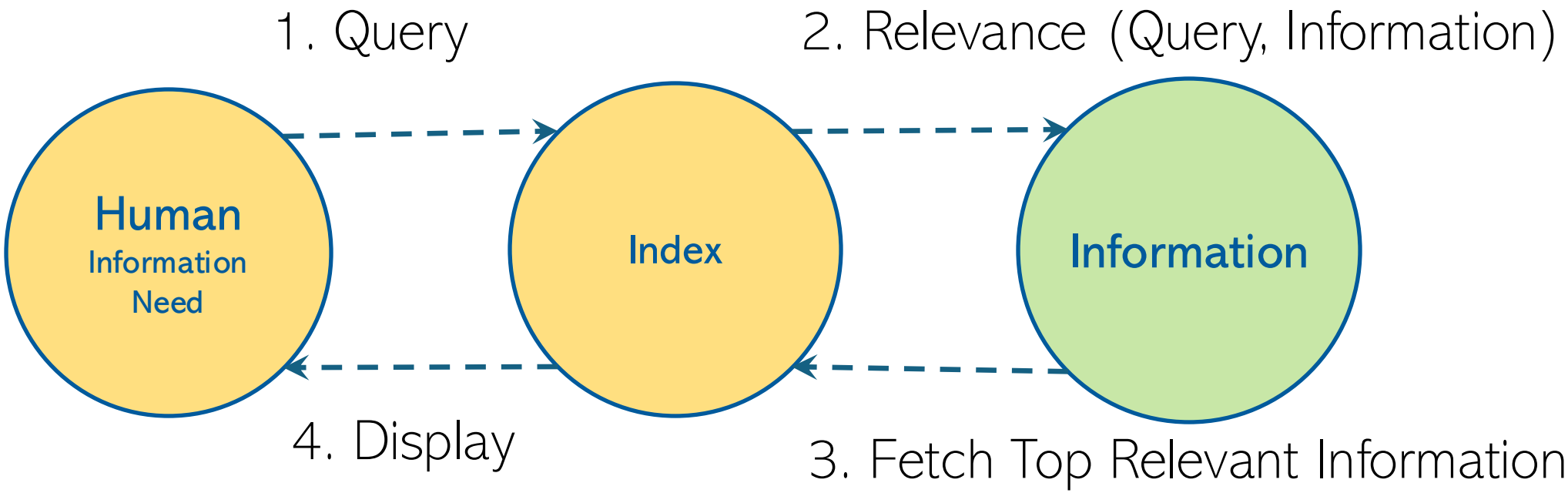
Personalized Query Reformulation via Conditional Transformers

Fani's Lab!

Google Search I'm Feeling Lucky

Google offered in: Français

Information Retrieval: Yesterday



Repair & RePair

q: *"which type of viruses can be directly used for translation"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC3579402
Tinkering with Translation: Protein Synthesis in Virus- ...
by D Walsh · 2013 · Cited by 277 — Exceptionally, retroviruses use **viral reverse transcriptase** to convert their (+)-stranded RNA genomes into a dsDNA form, which then serves as template…

https://www.ncbi.nlm.nih.gov › articles › PMC7158166
Replication and Expression Strategies of Viruses - PMC
by S Rampersad · 2018 · Cited by 71 — Other strategies used by viruses include internal initiation of translation of uncapped RNAs in **picornaviruses** and their relatives, snatching ...

LibreTexts
https://bio.libretexts.org › ... › Microbiology (Bruslind)
22: The Viruses
May 16, 2022 — **Class IV: +ssRNA Viruses with** plus-strand RNA, such as poliovirus, can use their genome directly as mRNA with translation by the host ribosome ...

Wikipedia
https://en.wikipedia.org › wiki › RNA_virus
RNA virus
Positive-sense ssRNA viruses (Group IV) have their genome directly utilized as mRNA, with host ribosomes translating it into a single protein that is modified ...

BCcampus Open Publishing
https://opentextbc.ca › biology › chapter › 12-1-viruses
12.1 Viruses – Concepts of Biology – 1st Canadian Edition
DNA viruses usually use host cell proteins and enzymes to make additional DNA that is used to copy the genome or be transcribed to messenger RNA (mRNA), which ...

q*: *"the gene from cell transcribed positive-sense rna"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC7173417
Introduction to RNA Viruses - PMC
by S Payne · 2017 · Cited by 40 — The genome of a positive-strand RNA virus is an mRNA that is translated, upon entry into **the cells**, to produce proteins needed for transcription and…

https://www.ncbi.nlm.nih.gov › articles › PMC165243
Host Factors in Positive-Strand RNA Virus Genome ...
by P Ahlquist · 2003 · Cited by 606 — All known **positive-strand RNA** viruses carry **genes** for an **RNA**-dependent **RNA** polymerase (RdRp) used in genome replication.

Wikipedia
https://en.wikipedia.org › wiki › Sense_(molecular_bi...
Sense (molecular biology)
Therefore, in positive-sense RNA viruses, the viral RNA genome can be considered viral mRNA, and can be immediately translated **by the host cell**.

Nature Journal
https://www.nature.com › scitable › topicpage › dna-tran...
DNA Transcription | Learn Science at Scitable
In eukaryotes, **RNA** pol I **transcribes** the **genes** that encode most of the ribosomal RNAs (rRNAs), and **RNA** pol III **transcribes** the **genes** for one small rRNA, **plus** ...

LibreTexts
https://bio.libretexts.org › Microbiology › 09:_Viruses
9.8A: Positive-Strand RNA Viruses of Animals
Jan 3, 2021 — Positive-sense viral RNA is similar to mRNA and thus can be immediately translated by **the host cell**.

Repair & RePair

q: *"which type of viruses can be directly used for translation"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC3579402

Tinkering with Translation: Protein Synthesis in Virus- ...

by D Walsh · 2013 · Cited by 277 — Exceptionally, retroviruses use **viral reverse transcriptase** to convert their (+)-stranded RNA genomes into a dsDNA form, which then serves as template…

https://www.ncbi.nlm.nih.gov › articles › PMC7158166

Replication and Expression Strategies of Viruses - PMC

by S Rampersad · 2018 · Cited by 71 — Other strategies used by viruses include internal initiation of translation of uncapped RNAs in **picornaviruses** and their relatives, snatching ...

LibreTexts
https://bio.libretexts.org › ... › Microbiology (Bruslind)

22: The Viruses

May 16, 2022 — **Class IV: +ssRNA Viruses with** plus-strand RNA, such as poliovirus, can use their genome directly as mRNA with translation by the host ribosome ...

Wikipedia
https://en.wikipedia.org › wiki › RNA_virus

RNA virus

Positive-sense ssRNA viruses (Group IV) have their genome directly utilized as mRNA, with host ribosomes translating it into a single protein that is modified ...

BCcampus Open Publishing
https://opentextbc.ca › biology › chapter › 12-1-viruses

12.1 Viruses – Concepts of Biology – 1st Canadian Edition

DNA viruses usually use host cell proteins and enzymes to make additional DNA that is used to copy the genome or be transcribed to messenger RNA (mRNA), which ...

q*: *"the gene from cell transcribed positive-sense rna"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC7173417

Introduction to RNA Viruses - PMC

by S Payne · 2017 · Cited by 40 — The genome of a positive-strand RNA virus is an mRNA that is translated, upon entry into **the cells**, to produce proteins needed for transcription and…

https://www.ncbi.nlm.nih.gov › articles › PMC165243

Host Factors in Positive-Strand RNA Virus Genome ...

by P Ahlquist · 2003 · Cited by 606 — All known **positive-strand RNA** viruses carry **genes** for an **RNA**-dependent **RNA** polymerase (RdRp) used in genome replication.

Wikipedia
https://en.wikipedia.org › wiki › Sense_(molecular_bi...

Sense (molecular biology)

Therefore, in positive-sense RNA viruses, the viral RNA genome can be considered viral mRNA, and can be immediately translated **by the host cell**.

Nature Journal
https://www.nature.com › scitable › topicpage › dna-tran...

DNA Transcription | Learn Science at Scitable

In eukaryotes, **RNA** pol I **transcribes** the **genes** that encode most of the ribosomal RNAs (rRNAs), and **RNA** pol III **transcribes** the **genes** for one small rRNA, **plus** ...

LibreTexts
https://bio.libretexts.org › Microbiology › 09:_Viruses

9.8A: Positive-Strand RNA Viruses of Animals

Jan 3, 2021 — Positive-sense viral RNA is similar to mRNA and thus can be immediately translated by **the host cell**.

Repair & RePair

q: *"which type of viruses can be directly used for translation"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC3579402

Tinkering with Translation: Protein Synthesis in Virus- ... ✓

by D Walsh · 2013 · Cited by 277 — Exceptionally, retroviruses use **viral reverse transcriptase** to convert their (+)-stranded RNA genomes into a dsDNA form, which then serves as template…

https://www.ncbi.nlm.nih.gov › articles › PMC7158166

Replication and Expression Strategies of Viruses - PMC ✗

by S Rampersad · 2018 · Cited by 71 — Other strategies used by viruses include internal initiation of translation of uncapped RNAs in **picornaviruses** and their relatives, snatching ...

LibreTexts
https://bio.libretexts.org › ... › Microbiology (Bruslind)

22: The Viruses ✓

May 16, 2022 — **Class IV: +ssRNA Viruses with** plus-strand RNA, such as poliovirus, can use their genome directly as mRNA with translation by the host ribosome ...

Wikipedia
https://en.wikipedia.org › wiki › RNA_virus

RNA virus ✓

Positive-sense ssRNA viruses (Group IV) have their genome directly utilized as mRNA, with host ribosomes translating it into a single protein that is modified ...

BCcampus Open Publishing
https://opentextbc.ca › biology › chapter › 12-1-viruses

12.1 Viruses – Concepts of Biology – 1st Canadian Edition ✗

DNA viruses usually use host cell proteins and enzymes to make additional DNA that is used to copy the genome or be transcribed to messenger RNA (mRNA), which ...

q*: *"the gene from cell transcribed positive-sense rna"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC7173417

Introduction to RNA Viruses - PMC ✓

by S Payne · 2017 · Cited by 40 — The genome of a positive-strand RNA virus is an mRNA that is translated, upon entry into **the cells**, to produce proteins needed for transcription and…

https://www.ncbi.nlm.nih.gov › articles › PMC165243

Host Factors in Positive-Strand RNA Virus Genome ... ✓

by P Ahlquist · 2003 · Cited by 606 — All known **positive-strand RNA** viruses carry **genes** for an **RNA**-dependent **RNA** polymerase (RdRp) used in genome replication.

Wikipedia
https://en.wikipedia.org › wiki › Sense_(molecular_bi...

Sense (molecular biology) ✓

Therefore, in positive-sense RNA viruses, the viral RNA genome can be considered viral mRNA, and can be immediately translated **by the host cell**.

Nature Journal
https://www.nature.com › scitable › topicpage › dna-tran...

DNA Transcription | Learn Science at Scitable ✓

In eukaryotes, **RNA** pol I **transcribes** the **genes** that encode most of the ribosomal RNAs (rRNAs), and **RNA** pol III **transcribes** the **genes** for one small rRNA, **plus** ...

LibreTexts
https://bio.libretexts.org › Microbiology › 09:_Viruses

9.8A: Positive-Strand RNA Viruses of Animals ✓

Jan 3, 2021 — Positive-sense viral RNA is similar to mRNA and thus can be immediately translated by **the host cell**.

Information Retrieval: Yesterday

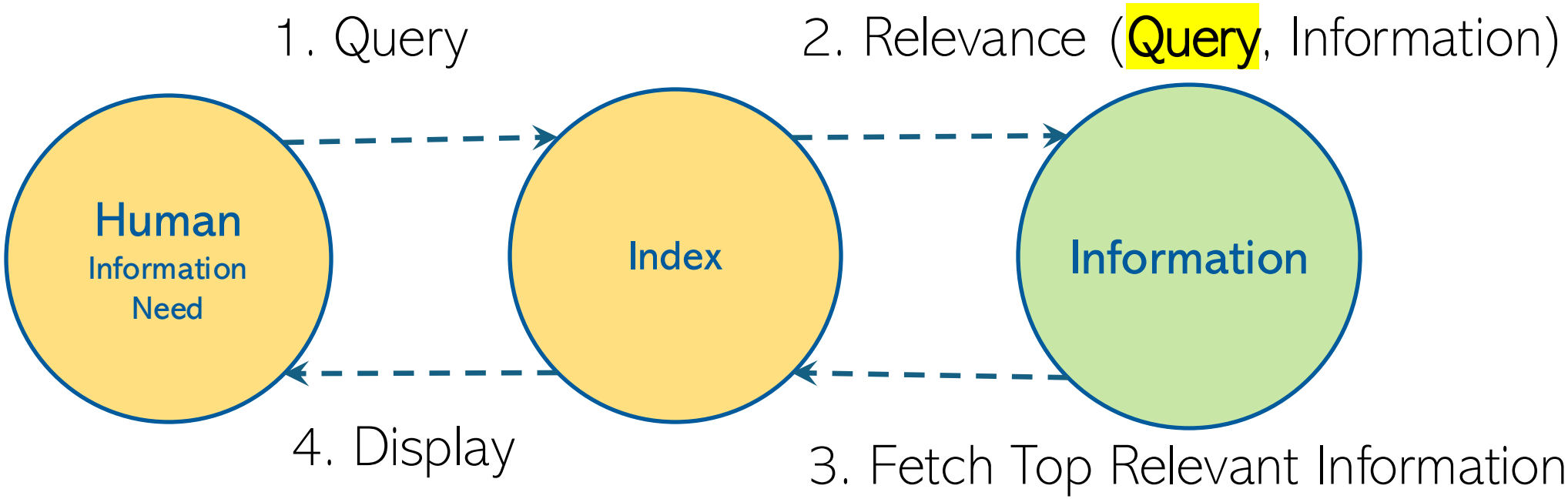


Photo: unknown

Query Formulation by User: A Difficult Task

Unclear/Uncertain Knowledge [Belkin'97]

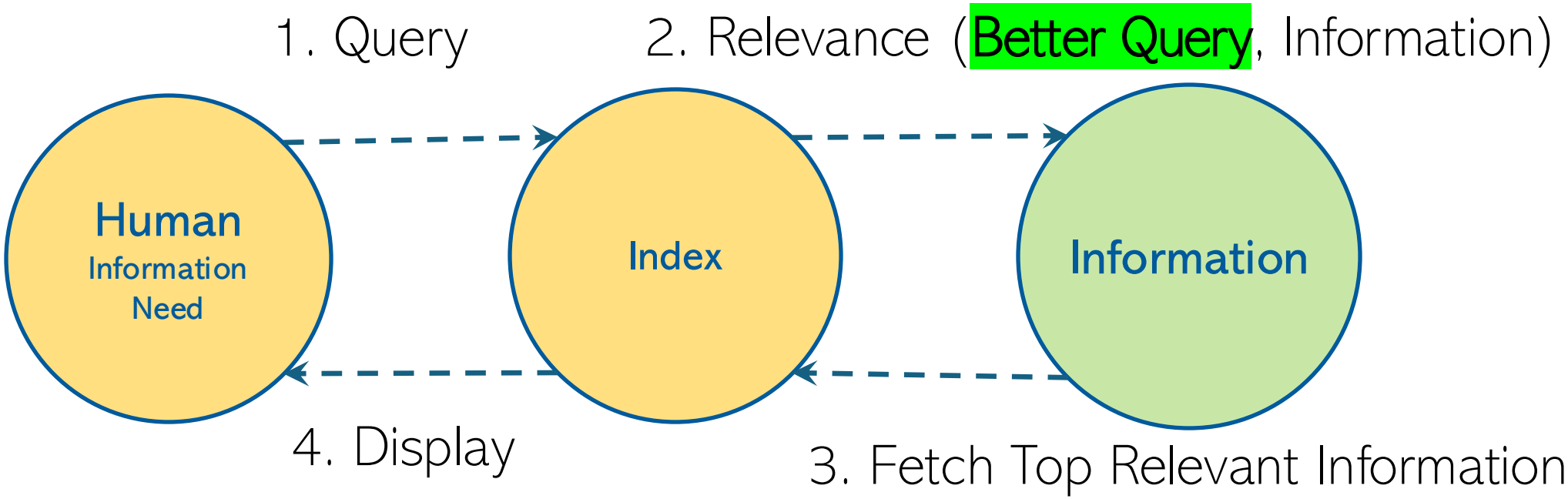
Vocabulary Mismatch [Furnas'87]

"lost shot" vs. "stray bullet"

N. J. **Belkin**, R. N. Oddy, and H. M. Brooks. **1997**. Ask for information retrieval: part I.: background and theory. Readings in information retrieval. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 299–304.

G. W. **Furnas**, T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. The vocabulary problem in human-system communication. Commun. ACM 30, 11 (Nov. **1987**), 964–971.

Information Retrieval: Yesterday



Repair & RePair

q: *"which type of viruses can be directly used for translation"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC3579402

Tinkering with Translation: Protein Synthesis in Virus- ... ✓

by D Walsh · 2013 · Cited by 277 — Exceptionally, retroviruses use **viral reverse transcriptase** to convert their (+)-stranded RNA genomes into a dsDNA form, which then serves as template…

https://www.ncbi.nlm.nih.gov › articles › PMC7158166

Replication and Expression Strategies of Viruses - PMC ✗

by S Rampersad · 2018 · Cited by 71 — Other strategies used by viruses include internal initiation of translation of uncapped RNAs in **picornaviruses** and their relatives, snatching ...

LibreTexts
https://bio.libretexts.org › ... › Microbiology (Bruslind)

22: The Viruses ✓

May 16, 2022 — **Class IV: +ssRNA Viruses with** plus-strand RNA, such as poliovirus, can use their genome directly as mRNA with translation by the host ribosome ...

Wikipedia
https://en.wikipedia.org › wiki › RNA_virus

RNA virus ✓

Positive-sense ssRNA viruses (Group IV) have their genome directly utilized as mRNA, with host ribosomes translating it into a single protein that is modified ...

BCcampus Open Publishing
https://opentextbc.ca › biology › chapter › 12-1-viruses

12.1 Viruses – Concepts of Biology – 1st Canadian Edition ✗

DNA viruses usually use host cell proteins and enzymes to make additional DNA that is used to copy the genome or be transcribed to messenger RNA (mRNA), which ...

q*: *"the gene from cell transcribed positive-sense rna"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC7173417

Introduction to RNA Viruses - PMC ✓

by S Payne · 2017 · Cited by 40 — The genome of a positive-strand RNA virus is an mRNA that is translated, upon entry into **the cells**, to produce proteins needed for transcription and…

https://www.ncbi.nlm.nih.gov › articles › PMC165243

Host Factors in Positive-Strand RNA Virus Genome ... ✓

by P Ahlquist · 2003 · Cited by 606 — All known **positive-strand RNA** viruses carry **genes** for an **RNA**-dependent **RNA** polymerase (RdRp) used in genome replication.

Wikipedia
https://en.wikipedia.org › wiki › Sense_(molecular_bi...

Sense (molecular biology) ✓

Therefore, in positive-sense RNA viruses, the viral RNA genome can be considered viral mRNA, and can be immediately translated **by the host cell**.

Nature Journal
https://www.nature.com › scitable › topicpage › dna-tran...

DNA Transcription | Learn Science at Scitable ✓

In eukaryotes, **RNA** pol I **transcribes** the **genes** that encode most of the ribosomal RNAs (rRNAs), and **RNA** pol III **transcribes** the **genes** for one small rRNA, **plus** ...

LibreTexts
https://bio.libretexts.org › Microbiology › 09:_Viruses

9.8A: Positive-Strand RNA Viruses of Animals ✓

Jan 3, 2021 — Positive-sense viral RNA is similar to mRNA and thus can be immediately translated by **the host cell**.

Repair & RePair

q: *"which type of viruses can be directly used for translation"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC3579402

Tinkering with Translation: Protein Synthesis in Virus- ... ✓

by D Walsh · 2013 · Cited by 277 — Exceptionally, retroviruses use **viral reverse transcriptase** to convert their (+)-stranded RNA genomes into a dsDNA form, which then serves as template…

https://www.ncbi.nlm.nih.gov › articles › PMC7158166

Replication and Expression Strategies of Viruses - PMC ✗

by S Rampersad · 2018 · Cited by 71 — Other strategies used by viruses include internal initiation of translation of uncapped RNAs in **picornaviruses** and their relatives, snatching ...

LibreTexts
https://bio.libretexts.org › ... › Microbiology (Bruslind)

22: The Viruses ✓

May 16, 2022 — **Class IV: +ssRNA Viruses with** plus-strand RNA, such as poliovirus, can use their genome directly as mRNA with translation by the host ribosome ...

Wikipedia
https://en.wikipedia.org › wiki › RNA_virus

RNA virus ✓

Positive-sense ssRNA viruses (Group IV) have their genome directly utilized as mRNA, with host ribosomes translating it into a single protein that is modified ...

BCcampus Open Publishing
https://opentextbc.ca › biology › chapter › 12-1-viruses

12.1 Viruses – Concepts of Biology – 1st Canadian Edition ✗

DNA viruses usually use host cell proteins and enzymes to make additional DNA that is used to copy the genome or be transcribed to messenger RNA (mRNA), which ...

q*: *"the gene from cell transcribed positive-sense rna"*

National Institutes of Health (.gov)
https://www.ncbi.nlm.nih.gov › articles › PMC7173417

Introduction to RNA Viruses - PMC ✓

by S Payne · 2017 · Cited by 40 — The genome of a positive-strand RNA virus is an mRNA that is translated, upon entry into **the cells**, to produce proteins needed for transcription and…

https://www.ncbi.nlm.nih.gov › articles › PMC165243

Host Factors in Positive-Strand RNA Virus Genome ... ✓

by P Ahlquist · 2003 · Cited by 606 — All known **positive-strand RNA** viruses carry **genes** for an **RNA**-dependent **RNA** polymerase (RdRp) used in genome replication.

Wikipedia
https://en.wikipedia.org › wiki › Sense_(molecular_bi...

Sense (molecular biology) ✓

Therefore, in positive-sense RNA viruses, the viral RNA genome can be considered viral mRNA, and can be immediately translated **by the host cell**.

Nature Journal
https://www.nature.com › scitable › topicpage › dna-tran...

DNA Transcription | Learn Science at Scitable ✓

In eukaryotes, **RNA** pol I **transcribes** the **genes** that encode most of the ribosomal RNAs (rRNAs), and **RNA** pol III **transcribes** the **genes** for one small rRNA, **plus** ...

LibreTexts
https://bio.libretexts.org › Microbiology › 09:_Viruses

9.8A: Positive-Strand RNA Viruses of Animals ✓

Jan 3, 2021 — Positive-sense viral RNA is similar to mRNA and thus can be immediately translated by **the host cell**.

Query Refinement, Reformulation, Suggestion, Expansion

Photo by: Julia Wimmerlin

Repair & RePair

q: first submitted query *"which type of viruses can be directly used for translation"*

Search Session & Clickthrough

q*: last submitted query *"the gene from cell transcribed positive-sense rna"*

Repair & RePair

q: first submitted query *"which type of viruses can be directly used for translation"*

Supervised Seq2Seq Neural Models [Dehghani'17, Ahmad'19]

q*: last submitted query *"the gene from cell transcribed positive-sense rna"*

Wasi **Ahmad** et al. Context Attentive Document Ranking and Query Suggestion. SIGIR'19.
Dehghani et al. Learning to Attend, Copy, and Generate for Session-Based Query Suggestion. **CIKM** '17. ACM, New York, NY, USA, 1747–1756.

Repair & RePair

q: first submitted query *"which type of viruses can be directly used for translation"*

Query (Topic) Drift [Tamannaee'20]

q*: last submitted query *"computer viruses"* or *"language translation"*

Tamannaee, Fani, et al. ReQue: A Configurable Workflow and Dataset Collection for Query Refinement. **CIKM** '20

Repair & RePair

q: first submitted query *"which type of viruses can be directly used for translation"*

ReQue: No query drift, guaranteed! [Tamannaee'20]

q*: refined query (ies) *"positive-sense rna"*

"gene from cell transcribed rna"

"the gene from cell transcribed positive-sense rna"

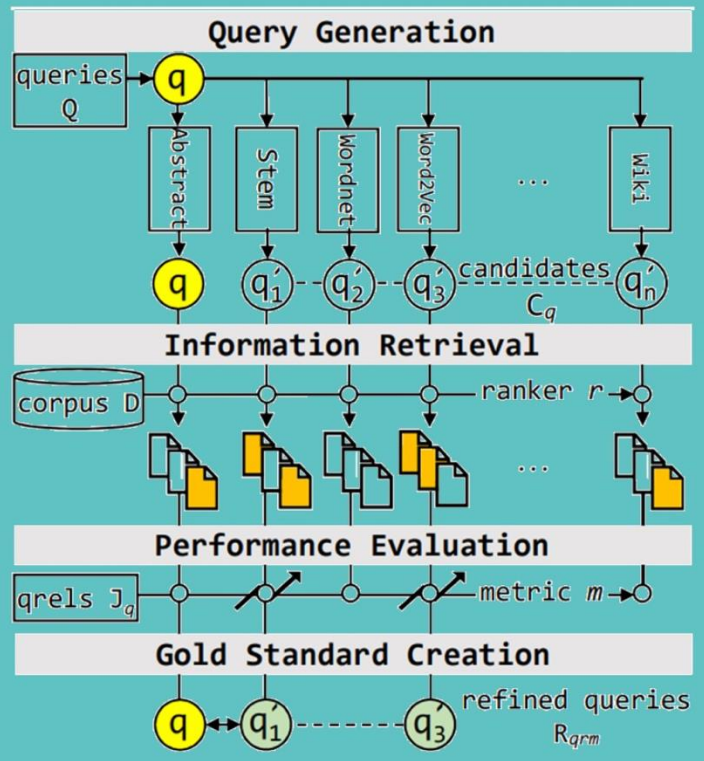
Tamannaee, Fani, et al. ReQue: A Configurable Workflow and Dataset Collection for Query Refinement. CIKM '20

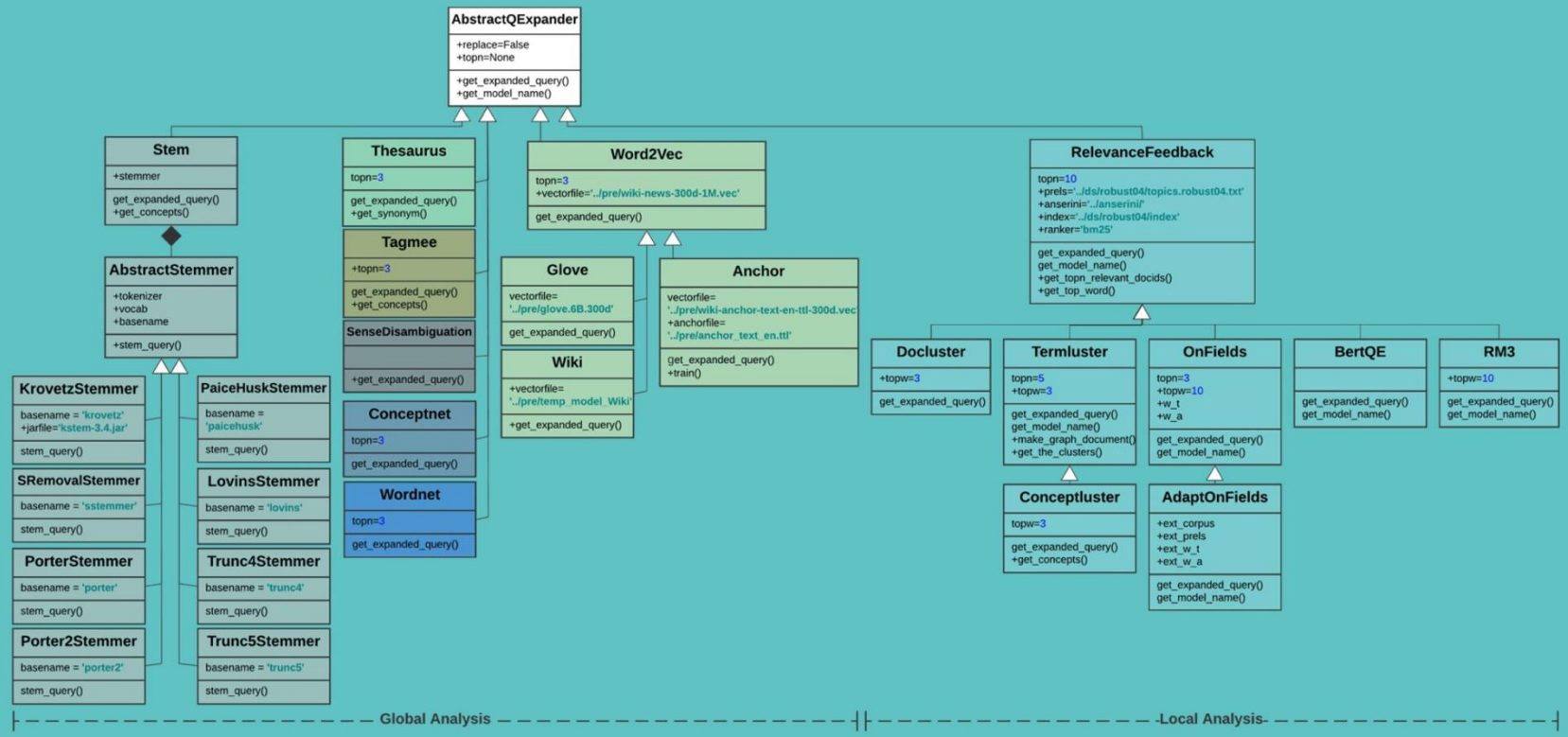
An Extensible Toolkit of Query Refinement Methods and Gold Standard Dataset Generation

H. Fani[1], N. Arabzadeh[2], M. Tamannaee[2], F. Zarrinkalam[2], J. Samouh[2], S. Paydar[2], E. Bagheri[2]

[1]University of Windsor, Canada. [2]Ryerson University, Canada.

Laboratory for Systems, Software and Semantics (LS3)





ReQue

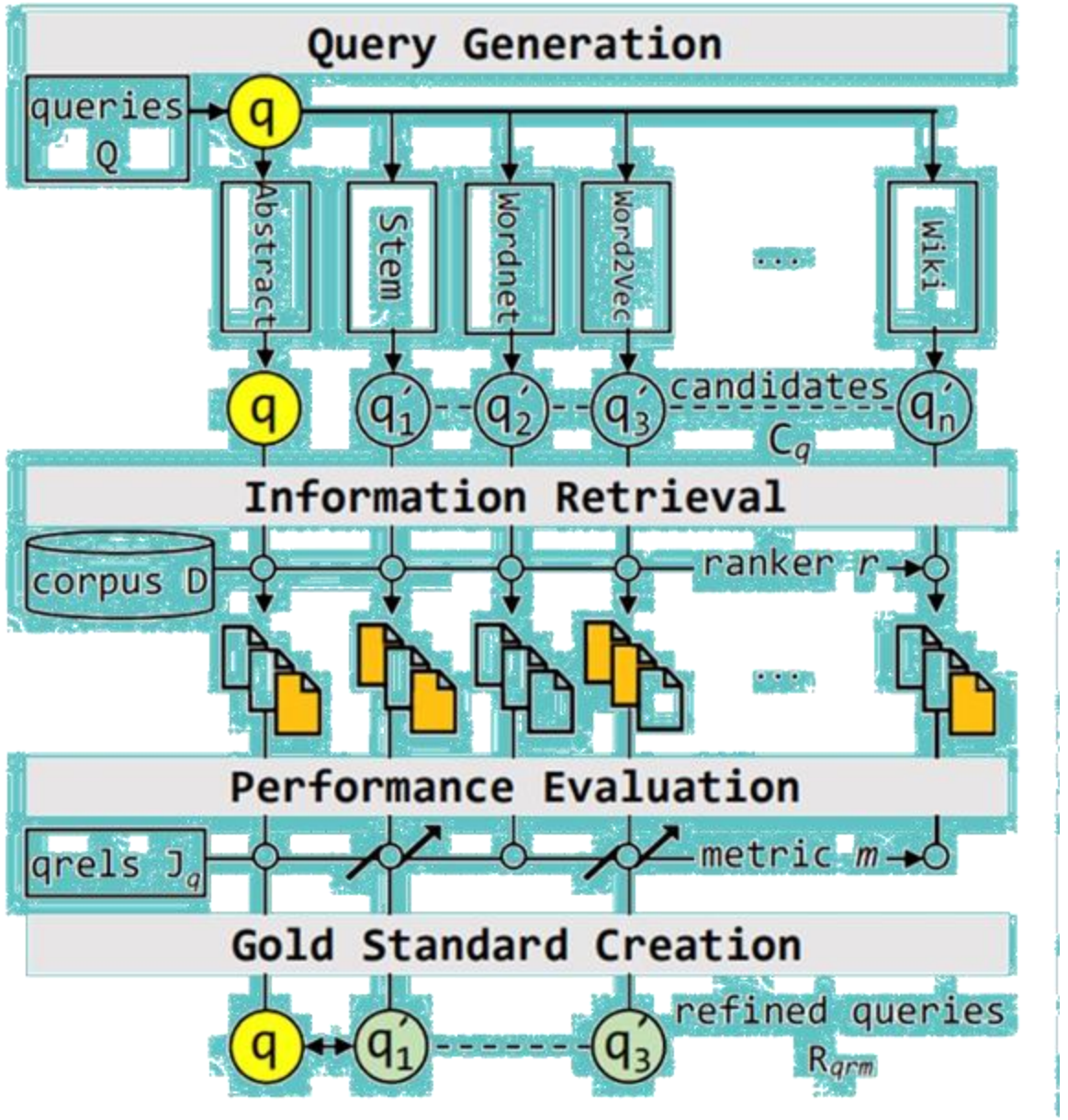
github.com/hosseinfani/ReQue

Repair & RePair

q: first submitted query

ReQue

q*: refined query (ies)



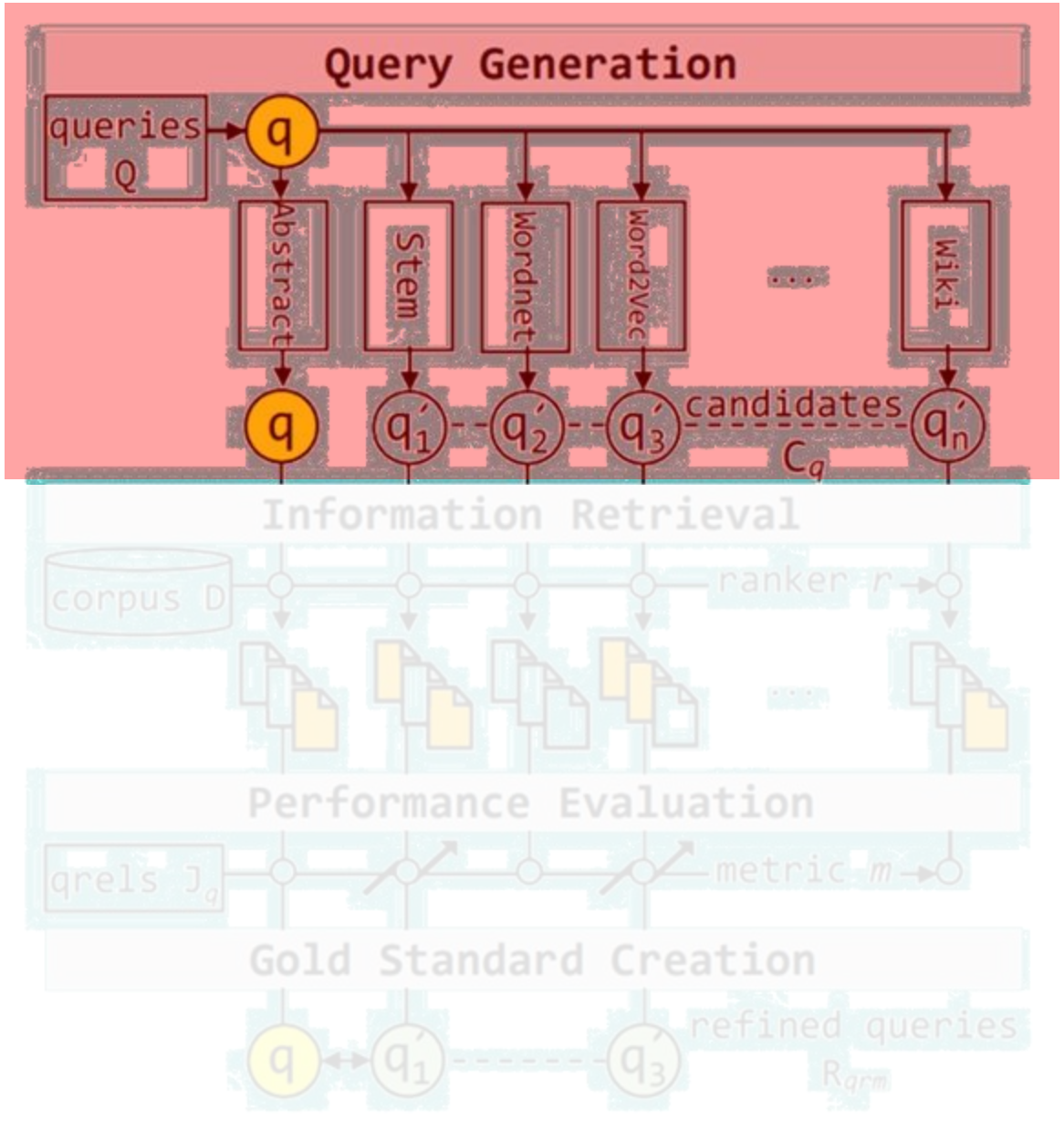
Tamannaee, Fani, et al. ReQue: A Configurable Workflow and Dataset Collection for Query Refinement. CIKM '20

Repair & RePair

q: first submitted query

ReQue: Time Consuming.

q*: refined query (ies)



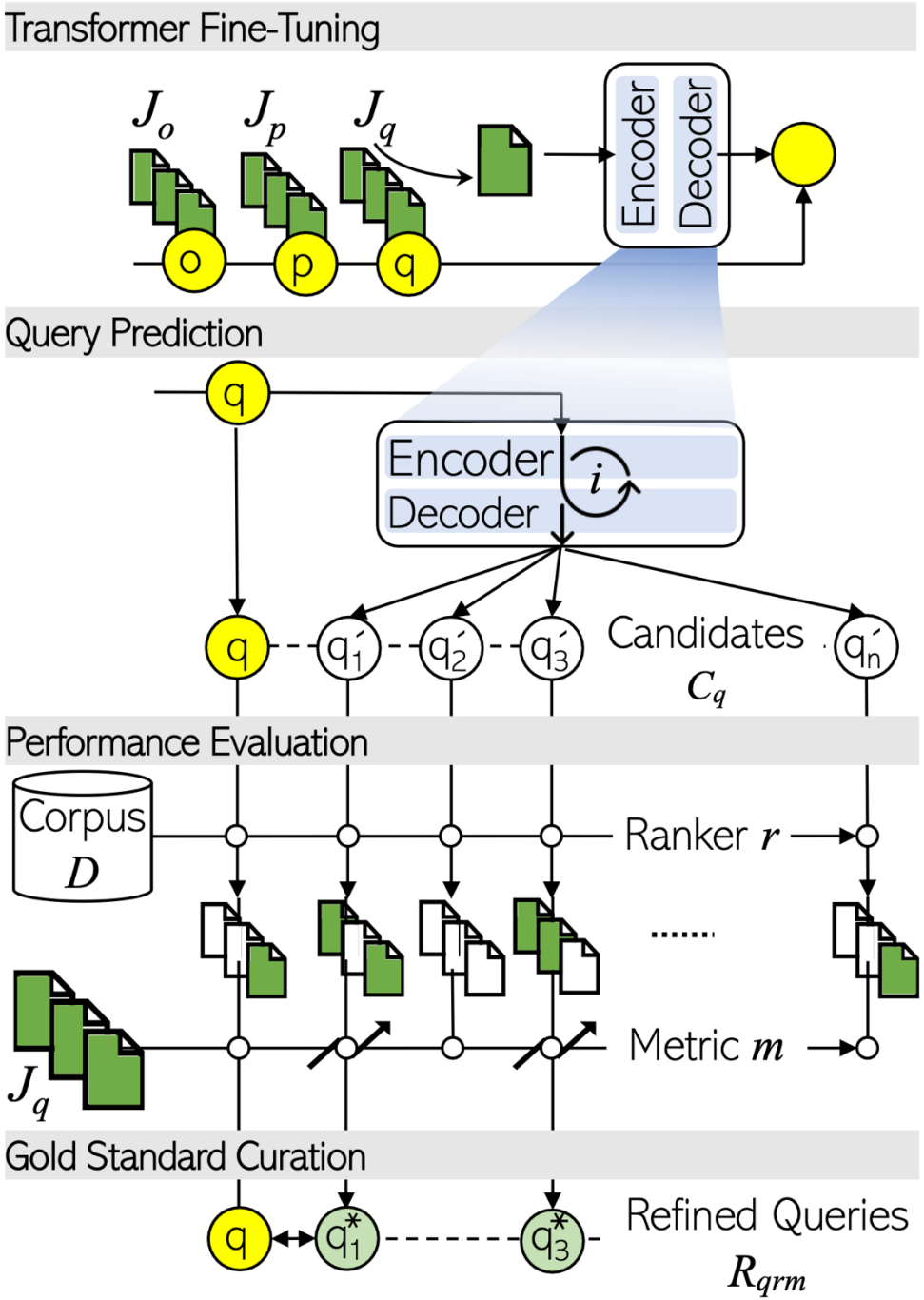
Tamannaee, Fani, et al. ReQue: A Configurable Workflow and Dataset Collection for Query Refinement. CIKM '20

Repair & RePair

q: first submitted query

RePair: Transformer-based.

q*: refined query (ies)



Arabzadeh et al. Matches Made in Heaven: Toolkit and Large-Scale Datasets for Supervised Query Reformulation. CIKM '21.

Narayanan and Fani. 2023. RePair: An Extensible Toolkit to Generate Large-Scale Datasets for Query Refinement via Transformers. CIKM '23.

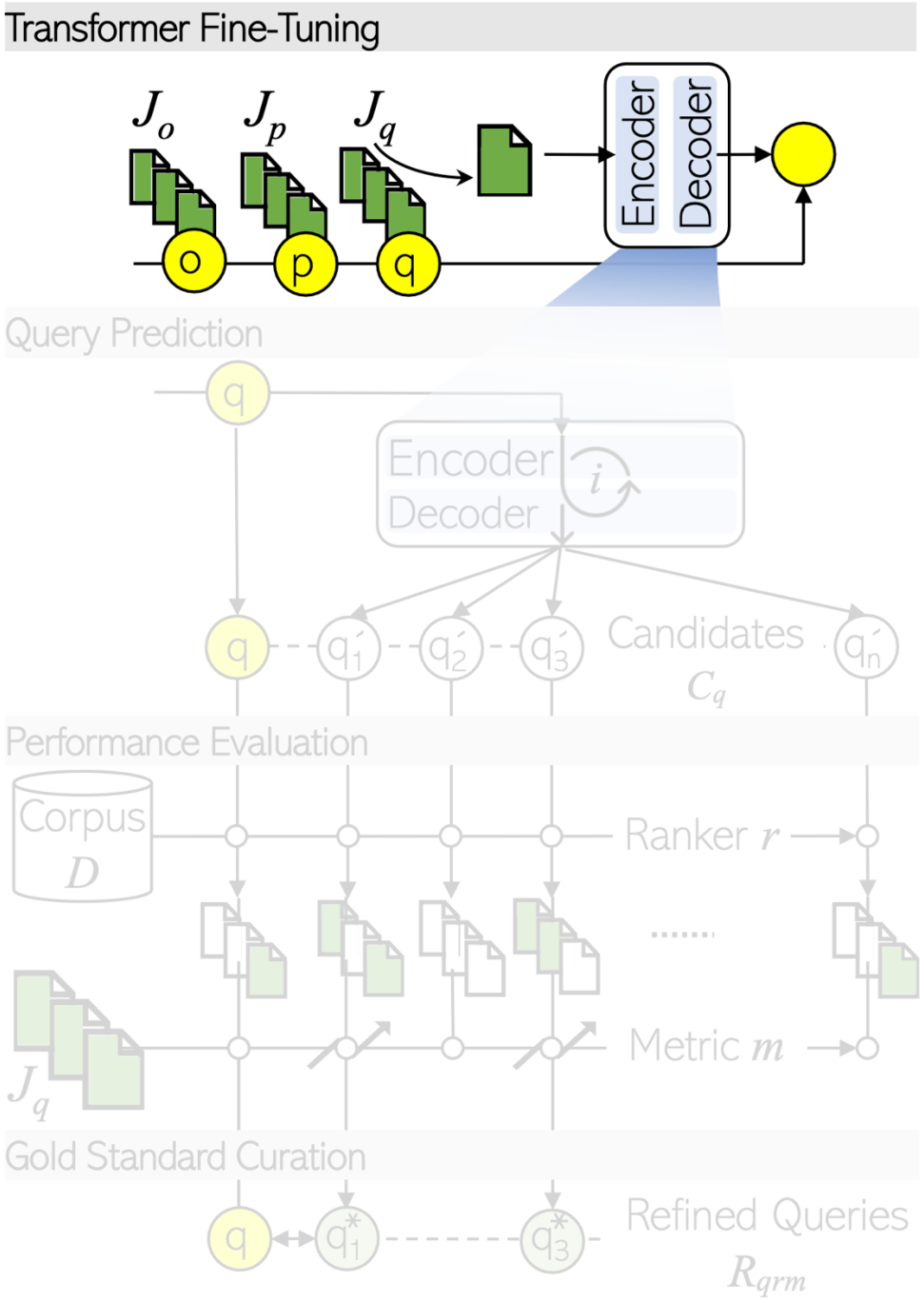
Repair & RePair

q: first submitted query

↓

J_q: relevant docs (rels)

$$P(q \mid J_q)$$



Arabzadeh et al. Matches Made in Heaven: Toolkit and Large-Scale Datasets for Supervised Query Reformulation. CIKM '21.

Narayanan and Fani. 2023. RePair: An Extensible Toolkit to Generate Large-Scale Datasets for Query Refinement via Transformers. CIKM '23.

Repair & RePair

q: first submitted query

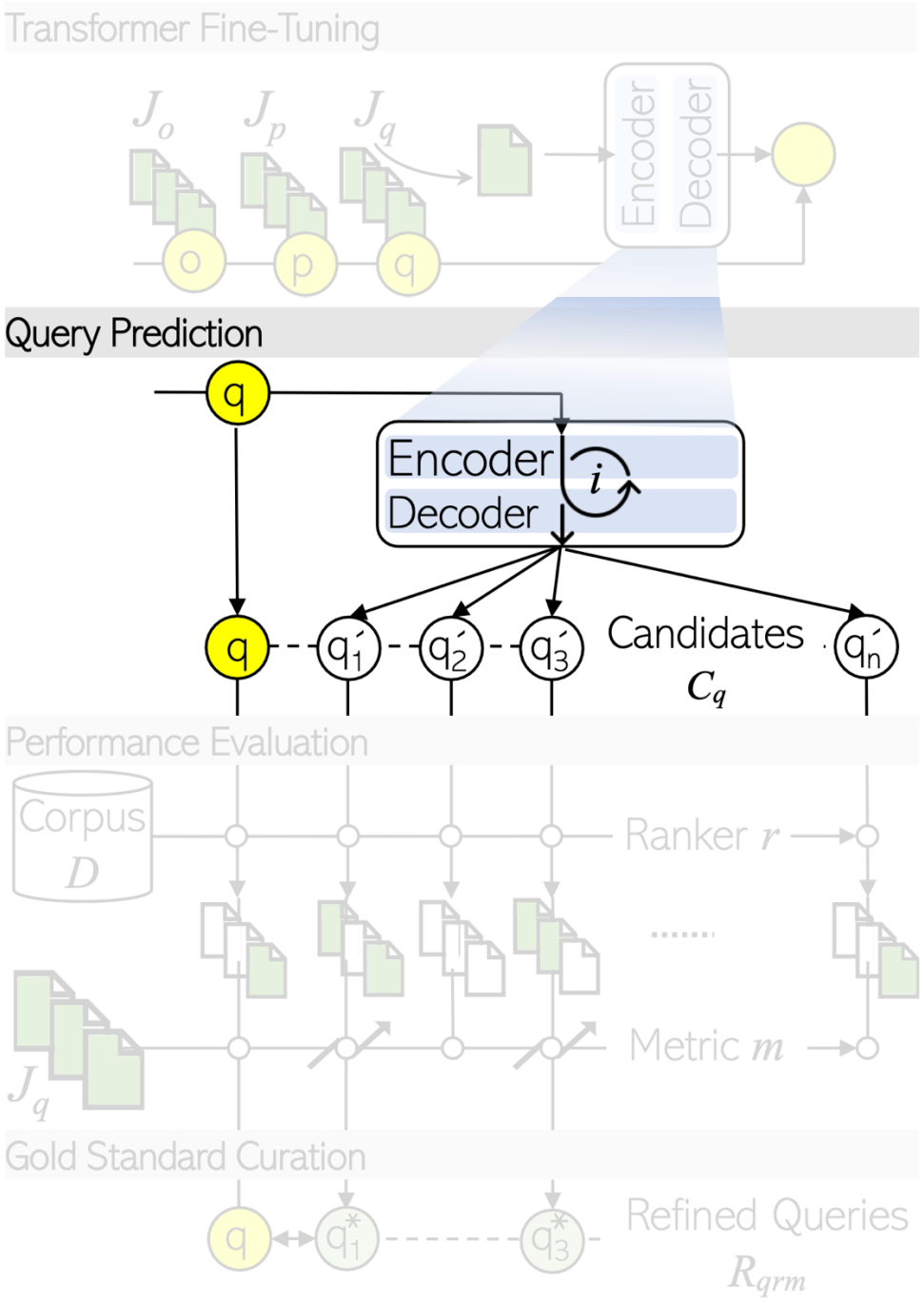
↓

J_q: relevant docs (rels)

↓

q*: refined query (ies)

$$P(q^* \mid q)$$



Repair & RePair

q: first submitted query *"which type of viruses can be directly used for translation"*

Query-driven, but user-less

q*: refined query *"the gene from cell transcribed positive-sense rna"*

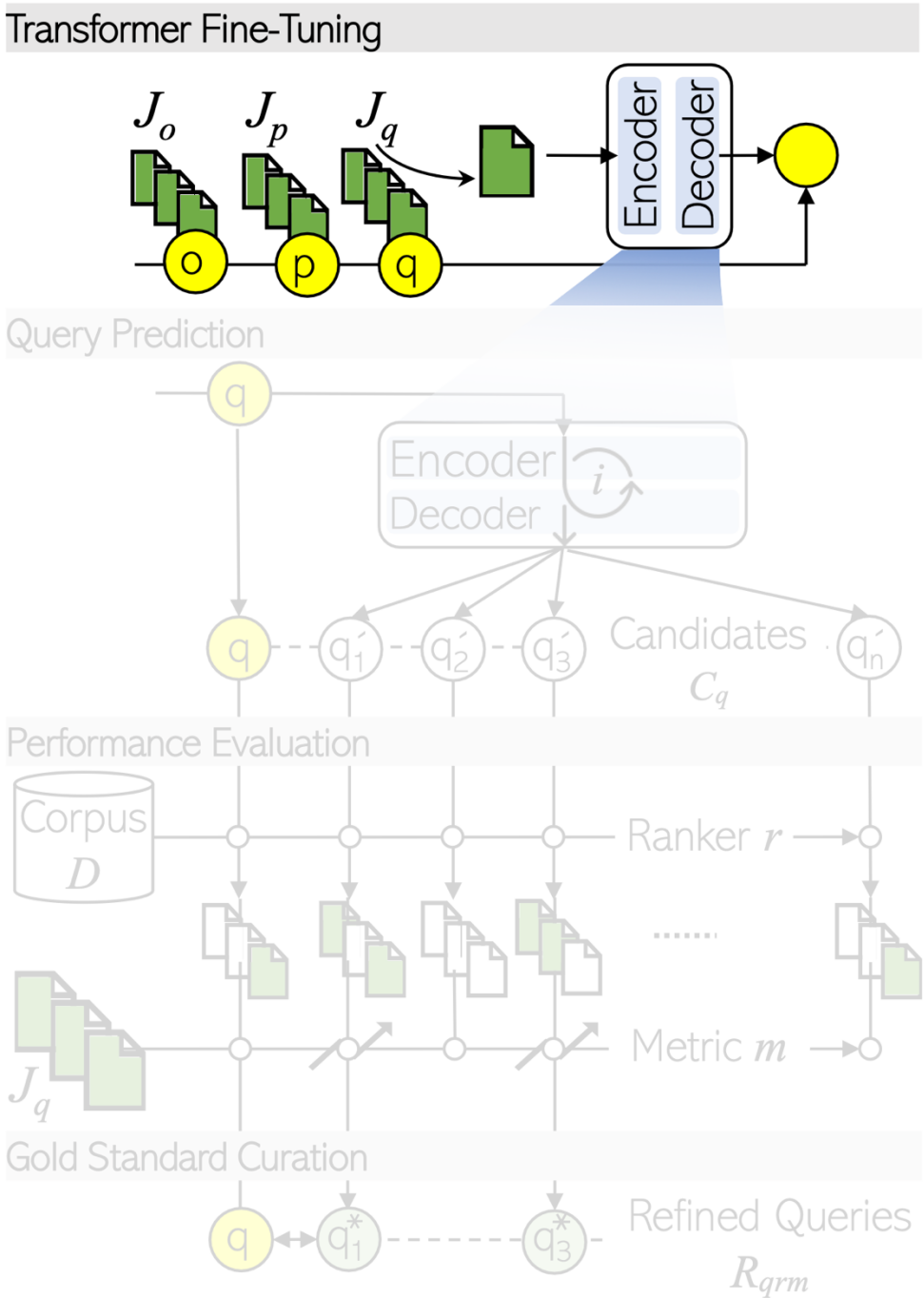
Repair & RePair My Query

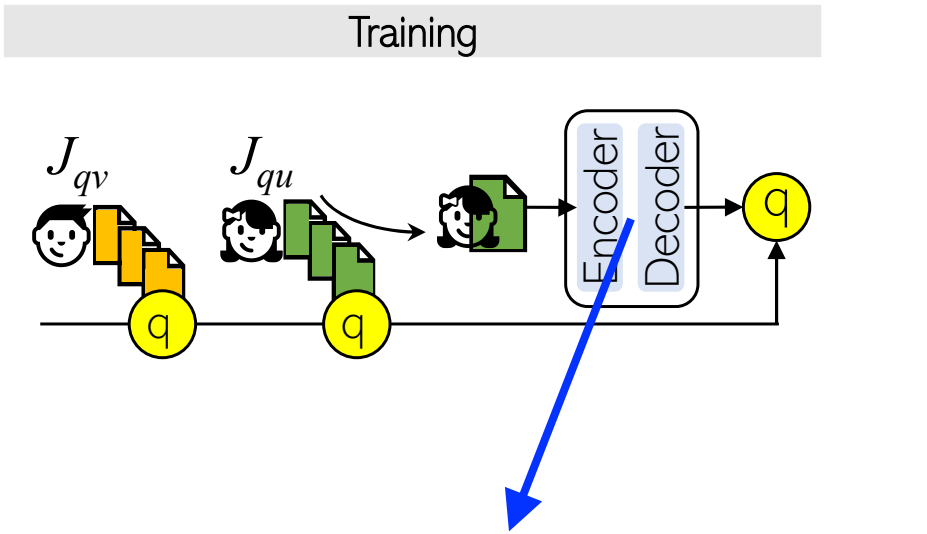
q: first submitted query *"which type of viruses can be directly used for translation"*

User-based (subjective relevance), query-driven

q^*_{Yoges}: refined query *"positive-sense rna"*

$q^*_{Hossein}$: refined query *"protein synthesis in virus-infected cells"*

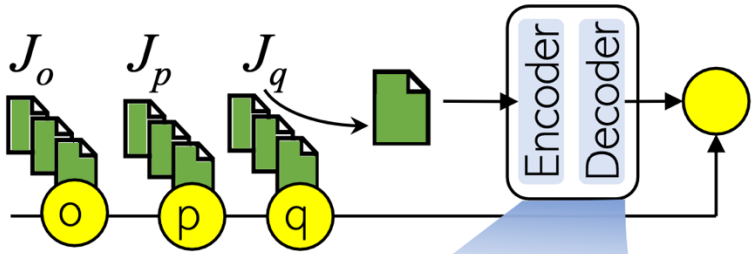




$$P(q \mid X=\text{uid}, J_q^{\text{uid}})$$

Conditional
Language Model

Transformer Fine-Tuning

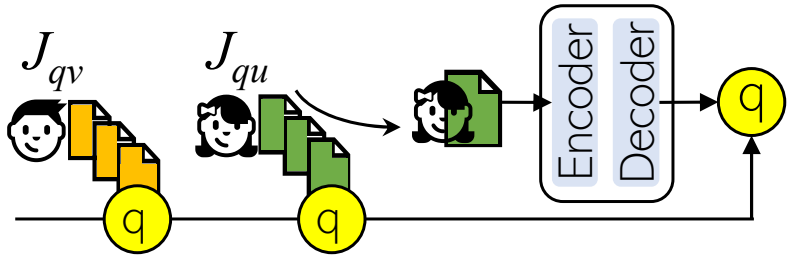


Query Prediction

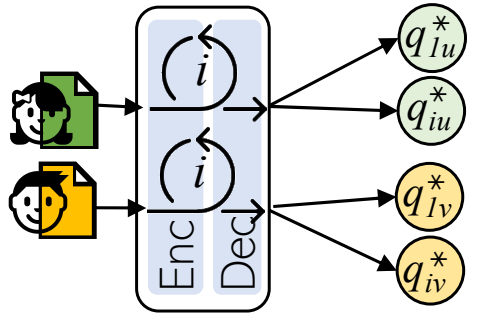
Performance Evaluation

Gold Standard Curation

Training



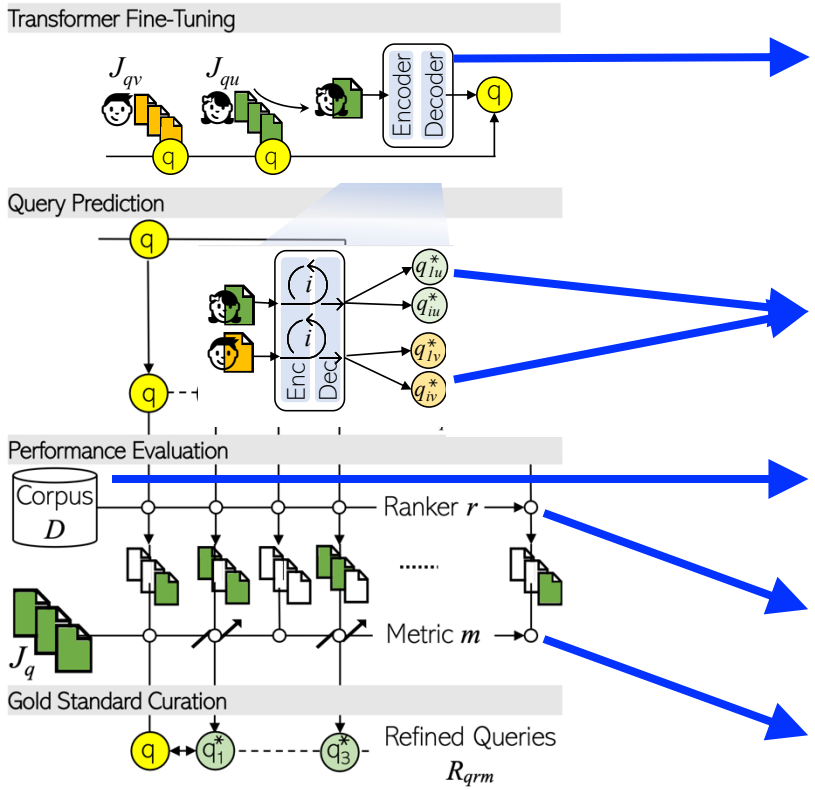
Inference



$$P(q^*_{uid} \mid X=uid, J_q^{uid})$$

5 MB hard drive being shipped by IBM - 1956.

Experiment RePair My Query



T5-base: 220M, Google Cloud,
+4,000 epoch
256 batch size
LR 0.001

Top-10 sampling [Fan et al.' 18]

AOL: UserId, Queries, Clicked URLs, 650K users
ir-dataset.com

Sparse: BM25
Dense: Colbert

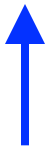
Mean Average Precision (MAP)
Mean Reciprocal Rank (MRR)

Fan, et al. Hierarchical Neural Story Generation. ACL 2018

Evaluation

RQ: Does personalized query refinement via conditional transformers improve the search experience for each user compared to when users are overlooked?

q	user-less q^*	map	uid	clicked url	user intent	user-based q^*	map
		0.142	05183534	www.tanforless.com	airbrush	*'tanforless com'*	1.000
'tanning bed lotions'	*'sunless tanning bed'*	0.014	13149841	www.suntanning.com	salons	*'tanning salons'*	0.083
		0.000	09469379	www.bodyconcept.com	bodybuilding	*'fitness tech'*	0.500



	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q_u^*	3.0543	2.0527	3.4778	3.0717
bm25.mrr$(\cdot : \mathcal{J}_{qu})$				
avg mrr(q)	0.0297		0.0364	
avg mrr(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
#mrr$(q^*) >$ mrr(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531	**1,132,741**	1,037,103	**1,347,612**
bm25.map$(\cdot : \mathcal{J}_{qu})$				
avg map(q)	0.0241		0.0308	
avg map(q^*)	0.4702	**0.5693**	0.4783	**0.5562**
#map$(q^*) >$ map(q)	2,583,023	**4,491,856**	2,421,347	**4,884,799**
%map$(q^*) >$ map(q)	58%	**61%**	52%	**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr$(\cdot : \mathcal{J}_{qu})$				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
#mrr$(q^*) >$ mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr$(q^*) >$ mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map$(\cdot : \mathcal{J}_{qu})$				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
#map$(q^*) >$ map(q)	2,334	**3,355**	2,715	**3,321**
%map$(q^*) >$ map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

Evaluation

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q_u^*	3.0543	2.0527	3.4778	3.0717
bm25.mrr$(\cdot : \mathcal{J}_{qu})$				
avg mrr(q)	0.0297		0.0364	
avg mrr(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
#mrr$(q^*) >$ mrr(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531			**1,347,612**
bm25.map$(\cdot : \mathcal{J}_{qu})$				
avg map(q)	0.		0308	
avg map(q^*)	0.4702			**0.5562**
#map$(q^*) >$ map(q)	2,583,023			**4,884,799**
%map$(q^*) >$ map(q)	58%			**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr$(\cdot : \mathcal{J}_{qu})$				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
#mrr$(q^*) >$ mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr$(q^*) >$ mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map$(\cdot : \mathcal{J}_{qu})$				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
#map$(q^*) >$ map(q)	2,334	**3,355**	2,715	**3,321**
%map$(q^*) >$ map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

While a webpage has **url**, it may be missing title. In **aol.title**, we filter out queries whose relevant webpages have no title.

Stats

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q^*_u	3.0543	2.0527	3.4778	3.0717
bm25.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0297		0.0364	
avg mrr(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
#mrr(q^*) > mrr(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531	**1,132,741**	1,037,103	**1,347,612**
bm25.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0241		0.0308	
avg map(q^*)	0.4702	**0.5693**	0.4783	**0.5562**
#map(q^*) > map(q)	2,583,023	**4,491,856**	2,421,347	**4,884,799**
%map(q^*) > map(q)	58%	**61%**	52%	**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
#mrr(q^*) > mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr(q^*) > mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
#map(q^*) > map(q)	2,334	**3,355**	2,715	**3,321**
%map(q^*) > map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

BM25.MRR

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q^*_u	3.0543	2.0527	3.4778	3.0717
bm25.mrr($\cdot : \mathcal{J}_{qu}$)				
avg **mrr**(q)	0.0297		0.0364	
avg **mrr**(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
$\#$**mrr**(q^*) > **mrr**(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531	**1,132,741**	1,037,103	**1,347,612**
bm25.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0241		0.0308	
avg map(q^*)	0.4702	**0.5693**	0.4783	**0.5562**
$\#$map(q^*) > map(q)	2,583,023	**4,491,856**	2,421,347	**4,884,799**
%map(q^*) > map(q)	58%	**61%**	52%	**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
$\#$mrr(q^*) > mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr(q^*) > mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
$\#$map(q^*) > map(q)	2,334	**3,355**	2,715	**3,321**
%map(q^*) > map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

BM25.MRR

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q^*_u	3.0543	2.0527	3.4778	3.0717
bm25.mrr($\cdot : \mathcal{J}_{qu}$)				
avg **mrr**(q)	0.0297		0.0364	
avg **mrr**(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
$\#$**mrr**(q^*) > **mrr**(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531	**1,132,741**	1,037,103	**1,347,612**
bm25.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0241		0.0308	
avg map(q^*)	0.4702	**0.5693**	0.4783	**0.5562**
$\#$map(q^*) > map(q)	2,583,023	**4,491,856**	2,421,347	**4,884,799**
%map(q^*) > map(q)	58%	**61%**	52%	**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
$\#$mrr(q^*) > mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr(q^*) > mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
$\#$map(q^*) > map(q)	2,334	**3,355**	2,715	**3,321**
%map(q^*) > map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

BM25.MRR

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q^*_u	3.0543	2.0527	3.4778	3.0717
bm25.mrr($\cdot : \mathcal{J}_{qu}$)				
avg **mrr**(q)	0.0297		0.0364	
avg **mrr**(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
$\#$**mrr**(q^*) > **mrr**(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531	**1,132,741**	1,037,103	**1,347,612**
bm25.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0241		0.0308	
avg map(q^*)	0.4702	**0.5693**	0.4783	**0.5562**
$\#$map(q^*) > map(q)	2,583,023	**4,491,856**	2,421,347	**4,884,799**
%map(q^*) > map(q)	58%	**61%**	52%	**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
$\#$mrr(q^*) > mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr(q^*) > mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
$\#$map(q^*) > map(q)	2,334	**3,355**	2,715	**3,321**
%map(q^*) > map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

BM25.MRR

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q^*_u	3.0543	2.0527	3.4778	3.0717
bm25.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0297		0.0364	
avg mrr(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
#mrr(q^*) > mrr(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531	**1,132,741**	1,037,103	**1,347,612**
bm25.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0241		0.0308	
avg map(q^*)	0.4702	**0.5693**	0.4783	**0.5562**
#map(q^*) > map(q)	2,583,023	**4,491,856**	2,421,347	**4,884,799**
%map(q^*) > map(q)	58%	**61%**	52%	**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
#mrr(q^*) > mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr(q^*) > mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
#map(q^*) > map(q)	2,334	**3,355**	2,715	**3,321**
%map(q^*) > map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

MAX MRR 1.00

BM25.MRR

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q^*_u	3.0543	2.0527	3.4778	3.0717
bm25.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0297		0.0364	
avg mrr(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
#mrr(q^*) > mrr(q)	2,276,965	**4,503,321**	2,845,642	**4,891,602**
%	51%	**61%**	60%	**61%**
#oracle q^*	1,069,531	**1,132,741**	1,037,103	**1,347,612**
bm25.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0241		0.0308	
avg map(q^*)	0.4702	**0.5693**	0.4783	**0.5562**
#map(q^*) > map(q)	2,583,023	**4,491,856**	2,421,347	**4,884,799**
%map(q^*) > map(q)	58%	**61%**	52%	**61%**
#oracle q^*	649,764	**686,682**	591,001	**855,355**
colbert.mrr($\cdot : \mathcal{J}_{qu}$)				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
#mrr(q^*) > mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr(q^*) > mrr(q)	11%	**11.4%**	11.4%	**16%**
#oracle q^*	0	0	0	0
colbert.map($\cdot : \mathcal{J}_{qu}$)				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
#map(q^*) > map(q)	2,334	**3,355**	2,715	**3,321**
%map(q^*) > map(q)	11.6%	**16.7%**	13.57%	**16.60%**
#oracle q^*	0	0	0	0

BM25.MAP

	aol.title		aol.title.url	
	−user	+user	−user	+user
$\#q$	4,459,613	7,348,389	4,672,506	8,020,979
avg #words in q	3.5849	3.0245	3.5817	2.9766
avg #words in q^* vs. q_u^*	3.0543	2.0527	3.4778	3.0717
bm25.mrr$(\cdot : \mathcal{J}_{qu})$				
avg mrr(q)	0.0297		0.0364	
avg mrr(q^*)	0.6670	**0.7937**	0.6807	**0.8172**
$\#$mrr$(q^*) >$ mrr(q)	2,276,965	4,503,321	2,845,642	4,891,602

Computational need is very high for all queries per user

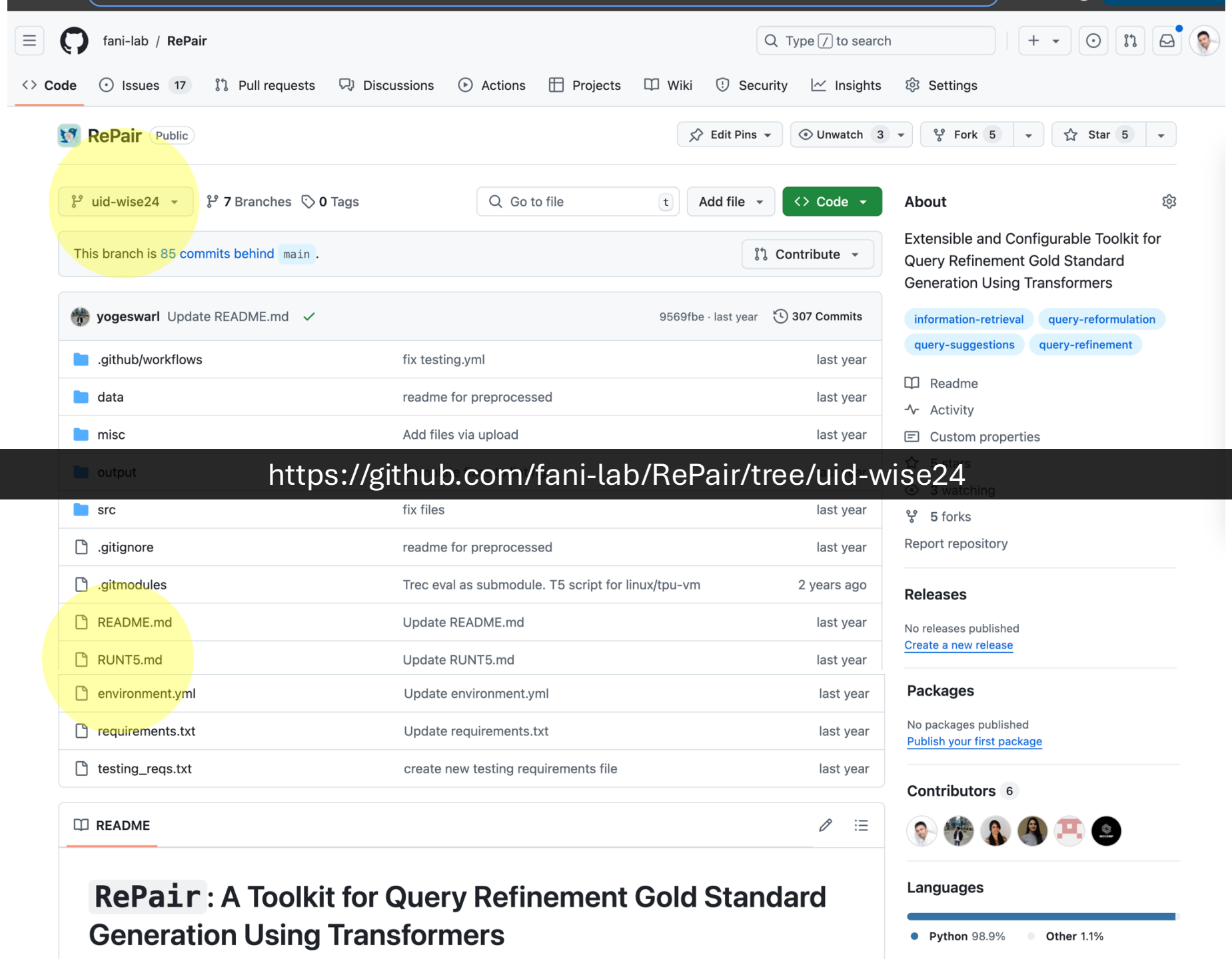
Let's evaluate it under pressure but for small number of queries

Those queries that are *hard*: no q* found for them using BM25

20,000 *hard queries*

colbert.mrr$(\cdot : \mathcal{J}_{qu})$				
avg mrr(q)	0.0807		0.0802	
avg mrr(q^*)	0.2790	**0.2989**	0.2224	**0.310**
$\#$mrr$(q^*) >$ mrr(q)	2,200	**2,288**	2,281	**3,216**
%mrr$(q^*) >$ mrr(q)	11%	**11.4%**	11.4%	**16%**
$\#$oracle q^*	0	0	0	0
colbert.map$(\cdot : \mathcal{J}_{qu})$				
avg map(q)	0.0661		0.0603	
avg map(q^*)	0.1357	**0.215**	0.0990	**0.1967**
$\#$map$(q^*) >$ map(q)	2,334	**3,355**	2,715	**3,321**
%map$(q^*) >$ map(q)	11.6%	**16.7%**	13.57%	**16.60%**
$\#$oracle q^*	0	0	0	0

Colbert.*



fani-lab / RePair

Type / to search

Code | Issues 17 | Pull requests | Discussions | Actions | Projects | Wiki | Security | Insights | Settings

RePair Public

Edit Pins | Unwatch 3 | Fork 5 | Star 5

uid-wise24 | 7 Branches | 0 Tags | Go to file t | Add file | Code

This branch is 85 commits behind main. | Contribute

yogeswarl Update README.md ✓ 9569fbe · last year | 307 Commits

.github/workflows	fix testing.yml	last year
data	readme for preprocessed	last year
misc	Add files via upload	last year
output		
src	fix files	last year
.gitignore	readme for preprocessed	last year
.gitmodules	Trec eval as submodule. T5 script for linux/tpu-vm	2 years ago
README.md	Update README.md	last year
RUNT5.md	Update RUNT5.md	last year
environment.yml	Update environment.yml	last year
requirements.txt	Update requirements.txt	last year
testing_reqs.txt	create new testing requirements file	last year

README

RePair: A Toolkit for Query Refinement Gold Standard Generation Using Transformers

About

Extensible and Configurable Toolkit for Query Refinement Gold Standard Generation Using Transformers

information-retrieval | query-reformulation | query-suggestions | query-refinement

Readme | Activity | Custom properties | 3 watching | 5 forks | Report repository

Releases

No releases published
Create a new release

Packages

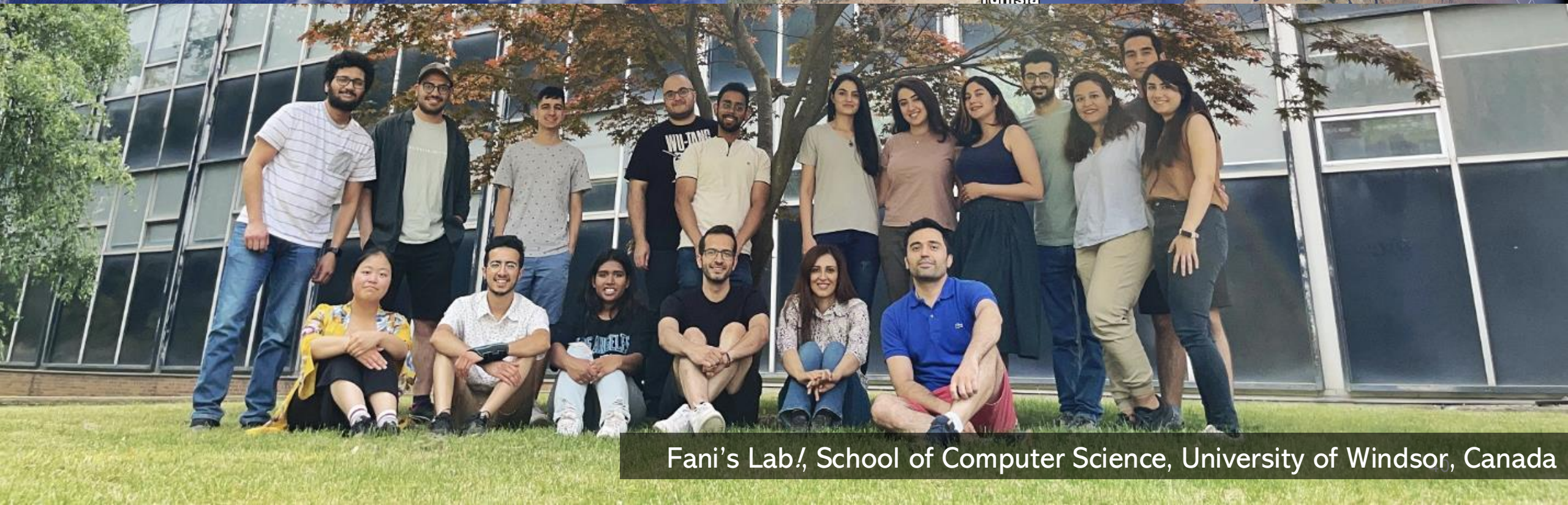
No packages published
Publish your first package

Contributors 6

Languages

Python 98.9% | Other 1.1%

https://github.com/fani-lab/RePair/tree/uid-wise24

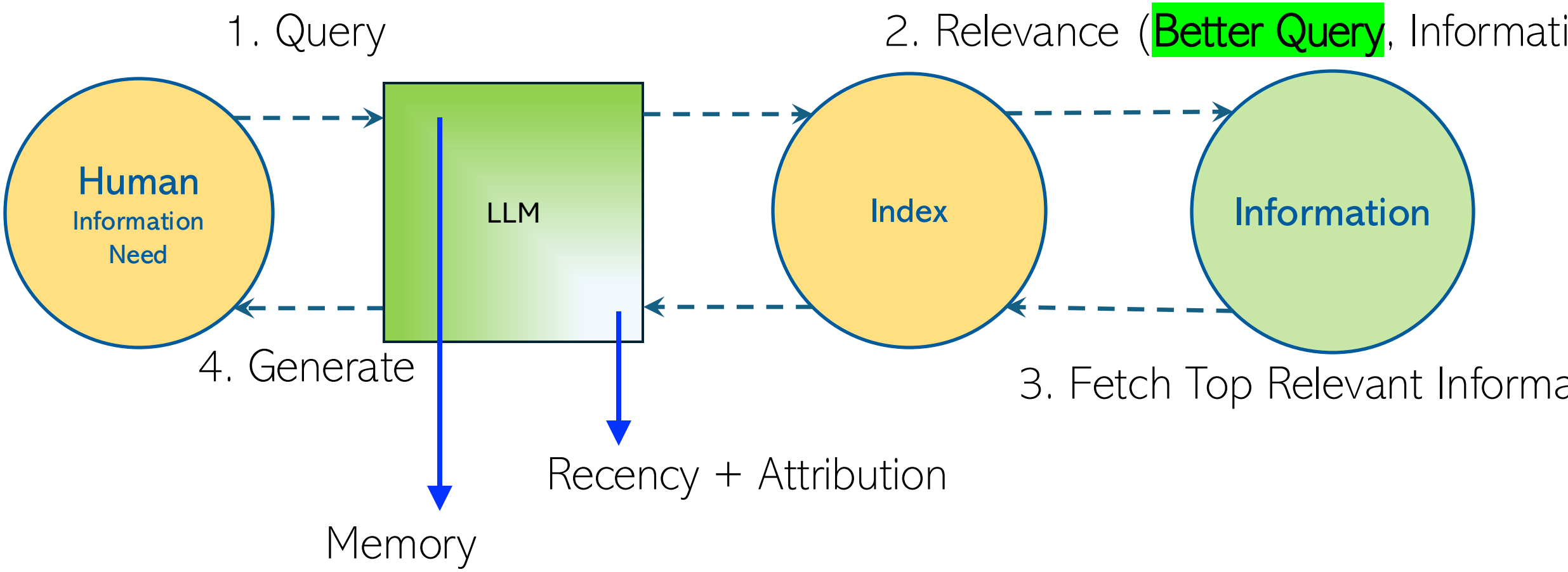


Fani's Lab!, School of Computer Science, University of Windsor, Canada


Yoges
Hossein

Future Direction: User Vector Representation vs. uid

RAGs: Nowadays



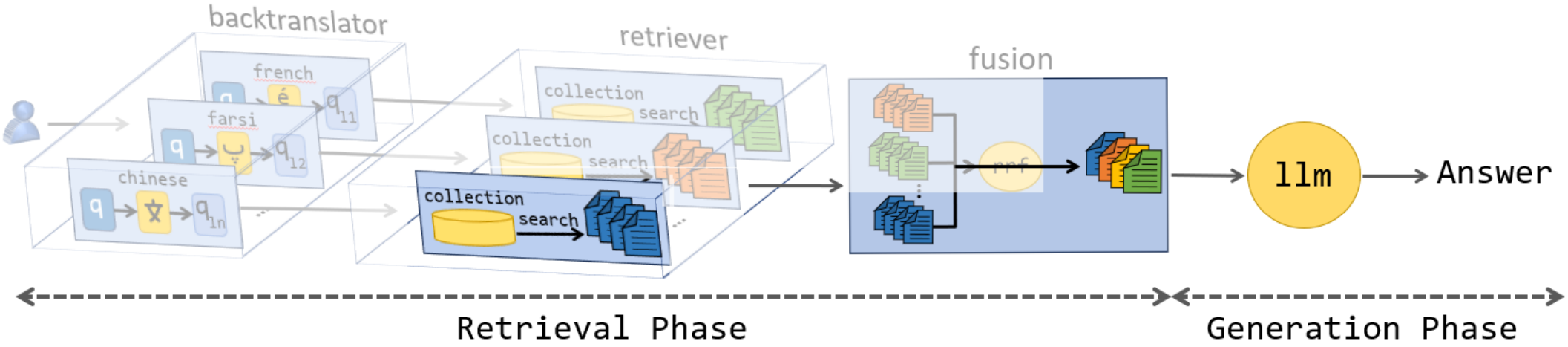
Retrieval-augmented generation (rag)

Session 27 – Query Processing & Information Extraction
Same Room (B116) @ 13:30


Delaram Rajaei (MSc)


Zahra Taheri (PhD)


Hossein Fani



A slide for people affected by the disaster of wars ...