Trie-NLG: Trie Context Augmentation to Improve Personalized Query Auto-Completion for Short and Unseen Prefixes

Accepted and Presented at ECML-PKDD 2023
Also to appear in Data Mining and Knowledge Discovery (DAMI)

Kaushal Kumar Maurya[1], Maunendra Sankar Desarkar[1],
Manish Gupta[2] and Puneet Agrawal[2]

[1]Department of CSE, IIT Hyderabad, India
[2]Bing AutoSuggest, Microsoft, Hyderabad, India

Outline

- Limitation of Existing QAC Models
- Goals and Problem Statement
- Proposed Approach: Trie-NLG
- Results and Analysis

PQAC: Personalized Query Auto-Completion

- Query Auto Completion (QAC): Recommend a list of relevant complete queries for partially typed search query (i.e. prefix)
- Helps in:
  - Saving keystrokes
  - System's understanding of search intent
  - Assisting users in efficiently expressing their intent

ecml InPrivate

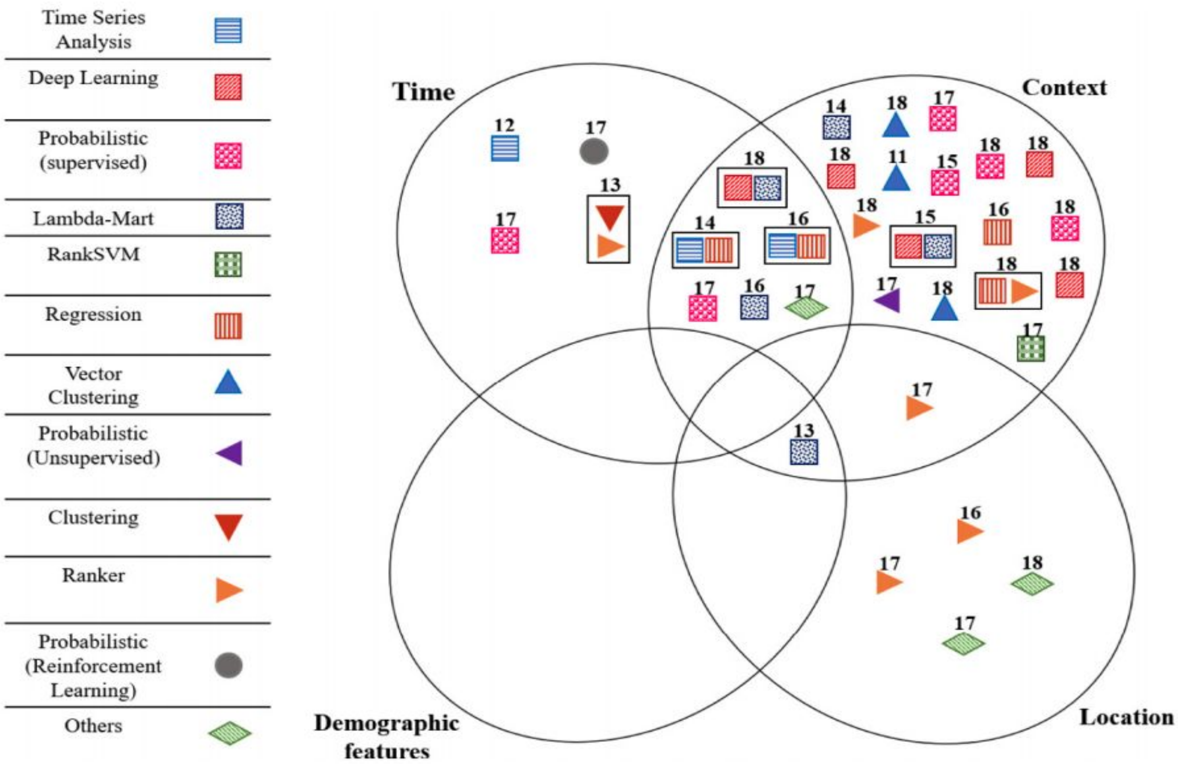
- ecml
- ecml **pkdd 2023**
- ecml **2023**
- ecml**ink**
- ecml **pkdd**
- ecml**-pkdd**

Session: mountains images||caves images||mountainside caves||mountain caves||timber wolves
Prefix: wolf p
Correct Query: wolf poetry

Generations:
1. wolf poetry
2. wolf pictures
3. wolf photos
4. wolf pics
5. wolf picture
6. wolf photo
7. wolf prints
8. wolf print

Sample from Bing Query log

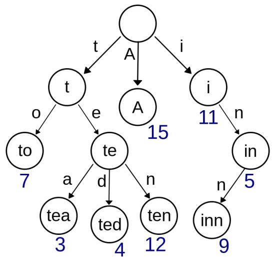
Existing QAC Models: A Taxonomy



Ref: Tahery, Saedeh, and Saeed Farzi. "Customized query auto-completion and suggestion—A review." *Information Systems* 87 (2020): 101415.

Trie/Ranking Models vs NLG Models

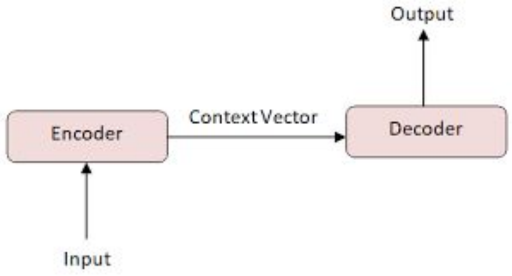
Trie/Ranking Models



+ Suggestions are more meaningful as they come from user log

- No personalization
- Provide limited number of suggestions
- No suggestions for unseen prefixes

NLG Models



+ Can model personalization
+ Generate suggestions for unseen prefixes

- For short-prefixes, suggestions are bad due to limited context
- Learn the generation biases from training dataset

Unseen and Short Prefixes in Bing Dataset

Char Length	Train			Validation			Test		
	Total	Seen	Unseen	Total	Seen	Unseen	Total	Seen	Unseen
Total	20.40M	17.86M	2.54M	100K	92.43K	7.57K	100K	92.80K	7.20K
[1-5]	9.10M	8.80M	0.30M	40.68K	40.39K	0.29K	40.19K	40.19K	0.27K
[6-10]	4.30M	4.10M	0.20M	21.40K	21.07K	0.33K	21.24K	21.24K	0.38K
10+	7.00M	4.96M	2.04M	37.92K	30.97K	6.95K	31.34K	31.37K	6.55K

- Unseen prefixes are those for which completions are not present in Trie created from Bing query log dataset of 1.5 years. Corresponding dataset is unseen dataset.
- Dataset is collected from Jul 2020 to Dec 2021
- Approximately 12% examples do not have any suggestions from Trie
- Approximately 65% examples has prefix length less than 11 and approximately 44% has less than 6

Microsoft

ECML PKDD 2023

Outline

- Limitation of Existing QAC Models
- Goals and Problem Statement
- Proposed Approach: Trie-NLG
- Results and Analysis

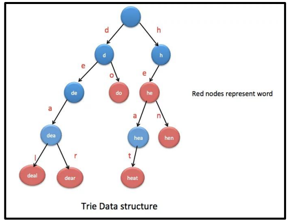
Factors to be Considered While Modeling

- Personalization
  - Leverage user’s session or query log
- Trie Context
  - Consider the suggested completions from Trie
- Deployable Latency

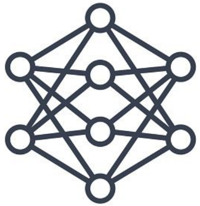
Problem Statement

Let previous ***n*** queries (earliest to latest order) in the current **session s** be $\{q_1, q_2, ..., q_n\}$. Current **query** is **q**, and **p** is the **query prefix** typed so far.

There are up to **m candidate query completions** (top-ranked to low-ranked order) $c_1, c_2, ..., c_m$ available as additional context **e** from a trie.



Trie Data structure



Generate top-**N** query completions conditioned on current **query prefix p**, additional trie context **e**, and session information **s** i.e.,

$$P_\theta(q \mid p; e; s)$$

Outline

- Limitation of Existing QAC Models
- Goals and Problem Statement
- Proposed Approach: Trie-NLG
- Results and Analysis

Proposed Trie-NLG Model

Three Major Components of Trie-NLG

- Trie Context Extraction (MPC_{Main}) for seen prefixes
- Synthetic Context Extraction (MPC_{Synth}) for unseen prefixes
- Context Augmentations in NLG (LLM)

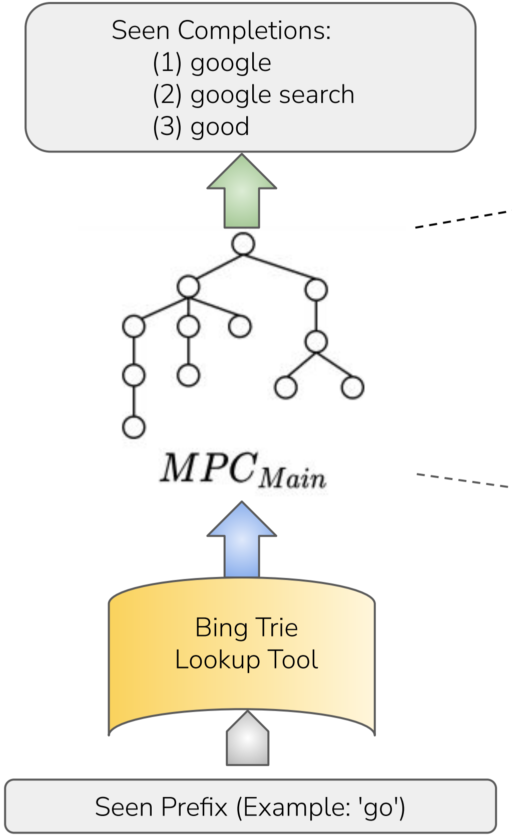
Example1: Short Prefix

Prefix: go
Clicked Query: google.com

Example2: Unseen Prefix

Prefix: kindle e-reader
Clicked Query: kindle e-reader questionnaire

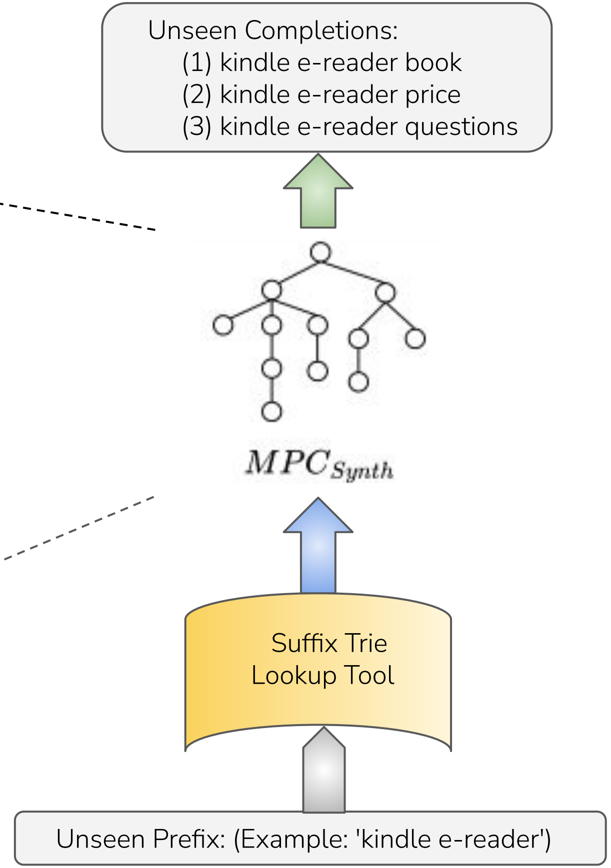
Trie Context Extraction (MPC_{Main})



- Created by Microsoft with 1.5 year's query log data
- Based on trie data structure
- Obtain m most popular completions
- For unseen prefix, no completions reported

Synthetic Context Extraction (MPC_{Synth})

- Provides synthetic completions for Unseen prefixes
Created from train datasets' queries
- Work at n-gram level
- Called as suffix trie (MPC_{synth})

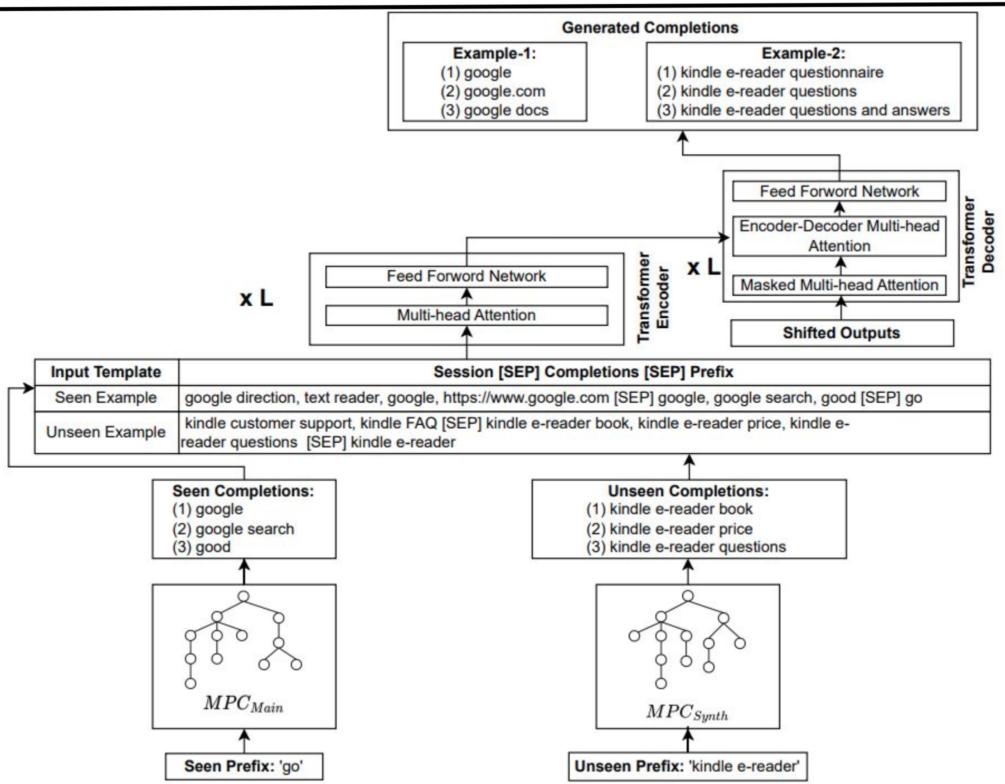


Example: Suffix Creation
Input Query: University of west florida Candidate Suffixes: 1. florida 2. west florida 3. of west florida

MPC_{Synth} Elements
{ University of west : florida, University of : west florida, University : of west florida}

Ref: Mitra, Bhaskar, and Nick Craswell. "Query auto-completion for rare prefixes." *Proceedings of the 24th ACM international on CIKM* 2015.

Context Augmentations in NLG (LLM)



Prefix: go
Clicked Query: google.com

Example1: Short Prefix

Prefix: kindle e-reader
Clicked Query: kindle e-reader questionnaire

Example2: Unseen Prefix

Outline

- Limitation of Existing QAC Models
- Goals and Problem Statement
- Proposed Approach: Trie-NLG
- Results and Analysis

Experimental Setup: Evaluation Metrics

- Mean Reciprocal Rank (MRR) [From IR Literature]

$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{r_i}$$

- Bilingual Evaluation Understudy (BLEU) [From NLP Literature]

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}.$$

Then,

$$\textsc{Bleu} = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right).$$

- BLEU Reciprocal Rank ($\mathrm{BLEU}_{\mathrm{RR}}$) [From IR+NLP]

$$\textsc{Bleu}_{\mathbf{RR}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{k}\frac{1}{j}\textsc{Bleu}(q_i^{\star}, \hat{q}_{i,j})}{\sum_{j=1}^{k}\frac{1}{j}}$$

1. MRR & RR_BLEU: Yadav, Nishant, et al. "Session-aware query auto-completion using extreme multi-label ranking." KDD 2021.
2. BLEU: Papineni, Kishore, et al. "Bleu: a method for automatic evaluation of machine translation." ACL 2002.

Microsoft

ECML PKDD 2023

Experimental Setup: Datasets and Baselines

Char Length	Bing Train			AOL Train Validation		
	Total	Seen	Unseen	Total	Seen	Unseen
Total	20.40M	17.86M	2.54M	3.91M	3.47	0.44M
[1-5]	9.10M	8.80M	0.30M	1.42M	1.42M	0.00M
[6-10]	4.30M	4.10M	0.20M	1.15M	1.11M	0.04M
10+	7.00M	4.96M	2.04M	1.34M	0.94M	0.40M

- Multi-level pre-processing for AOL
- AOL unseen splits are obtained with Bing Trie
- For both datasets, there is each 100K val and test split

Baselines

1. $\text{MPC}_{\text{Train}}$
2. MPC_{Main}
3. $\text{MPC}_{\text{Main}} + \text{MPC}_{\text{Synth}}$
4. GRM
5. Seq2Seq LSTM
6. Seq2Seq Transformer
7. T5
8. BART
9. BART + ITC
10. BART + MPC_{Main}

Results: AOL Datasets

Models	AOL ALL Dataset				AOL Seen Dataset				AOL Unseen Dataset		
	MRR	$BLEU_{RR}$	BLEU		MRR	$BLEU_{RR}$	BLEU		MRR	$BLEU_{RR}$	BLEU
TrainMPC	20.6	5.4	15.25		23.2	6.1	19.49		-	-	-
TrainMPC + SynthMPC	31.3	12.0	40.10		23.2	6.1	19.49		**95.2**	**58.7**	**95.66**
TrieMPC	19.7	9.9	24.44		22.2	11.2	29.75		-	-	-
TrieMPC + SynthMPC	30.4	16.6	47.76		22.2	11.2	29.75		**95.2**	**58.7**	**95.6**
GRM	21.8	7.3	20.69		24.6	8.3	23.33		-	-	-
Seq2Seq LSTM	43.9	14.7	51.43		43.6	12.9	49.16		47.1	39.3	69.18
Seq2Seq Transformer	45.4	16.8	57.50		44.5	14.8	51.79		51.9	32.3	73.62
T5	48.1	17.4	59.63		46.6	15.2	53.42		59.5	34.6	77.18
BART	51.9	18.3	61.89		50.3	16.1	55.64		64.7	35.8	79.55
BART + ITC	50.7	18.3	61.55		49.1	16.0	55.29		63.5	35.9	79.24
BART + MPC_{Main}	53.2	18.6	62.48		51.8	16.4	**56.58**		64.1	35.6	79.21
Trie-NLG	**56.5**	**19.3**	**66.63**		**52.0**	**16.5**	56.56		92.1	41.2	94.62

(1) Unseen in only ~11%
(2) The trie suggestions for Unseen prefixed is limited for NLG models
(3) Synth MPC results in better context for MPC models

Results: Bing Datasets

Percentage (%) Improvement over TrainMPC + SynthMPC Baseline model

Models	Bing ALL Dataset				Bing Seen Dataset				Bing Unseen Dataset		
	ΔMRR	$\Delta BLEU_{RR}$	ΔBLEU		ΔMRR	$\Delta BLEU_{RR}$	ΔBLEU		ΔMRR	$\Delta BLEU_{RR}$	ΔBLEU
TrainMPC	-7.90	-19.87	-24.64		0.00	0.00	0.00		-	-	-
TrieMPC	-0.11	36.61	26.38		8.49	70.38	63.68		-	-	-
TrieMPC + SynthMPC	7.78	56.42	47.01		8.49	70.38	63.68		0.00	0.00	0.00
GRM	-1.43	-5.32	-5.51		4.30	19.02	16.58		-	-	-
Seq2Seq LSTM	9.61	39.30	56.78		10.24	44.02	72.20		5.34	16.61	91.45
Seq2Seq Transformer	15.74	54.22	76.16		18.24	55.07	82.46		10.53	24.20	99.29
T5	20.78	61.81	78.70		20.76	70.83	85.13		21.60	27.33	103.9
BART	36.73	73.47	91.09		36.95	84.51	100.4		34.53	29.03	110.3
BART + ITC	31.66	71.43	88.72		31.58	81.97	96.84		33.05	29.12	111.0
BART + MPC_{Main}	54.12	86.77	110.7		56.14	101.3	129.0		34.10	28.04	110.8
Trie-NLG	**56.78**	**88.26**	**114.5**		**56.56**	**101.9**	**130.0**		**59.7**	**33.0**	**123.0**

Microsoft ECML PKDD 2023

Results: Short Prefixes

	Prefix Length in [1-5]			Prefix Length in [6-10]		
Bing Dataset	**ΔMRR**	**$\Delta\text{BLEU}_{\text{RR}}$**	**ΔBLEU**	**ΔMRR**	**$\Delta\text{BLEU}_{\text{RR}}$**	**ΔBLEU**
BART	21.6	-17.6	2.9	38.5	38.4	56.8
BART + MPC_{Main}	52.1	-16.8	10.1	55.2	49.3	72.1
TRIE-NLG	**53.2**	**-15.9**	**11.4**	**56.4**	**50.1**	**73.6**
AOL Dataset	**MRR**	**BLEU_{RR}**	**BLEU**	**MRR**	**BLEU_{RR}**	**BLEU**
BART	41.3	8.0	35.39	52.2	14.3	53.39
BART + MPC_{Main}	41.5	8.0	35.25	54.1	14.7	54.08
TRIE-NLG	**42.1**	**8.1**	**35.58**	**55.2**	**14.9**	**55.56**

	Prefix Length 10+		
Bing Dataset	**ΔMRR**	**$\Delta\text{BLEU}_{\text{RR}}$**	**ΔBLEU**
BART	49.1	190.0	120.1
BART + MPC_{Main}	55.6	218.5	145.1
TRIE-NLG	**60.7**	**221.1**	**149.5**
AOL Dataset	**MRR**	**BLEU_{RR}**	**BLEU**
BART	63.2	32.8	75.40
BART + MPC_{Main}	65.1	33.4	76.20
TRIE-NLG	**73.4**	**35.1**	**82.79**

Analysis: Case Study

<table>
<tr><th colspan="2">Example-1: Seen Short Prefix</th><th colspan="2">Example-2: Unseen Prefix</th></tr>
<tr><td colspan="2">Session: kysportsradio || kysportsradio || cincinnati reds || cincinnati reds || espn sports || espn sports || ebth || ebth.com || ebth.com || cnn news || cnn news || politico news
Prefix: p
Correct Query: politico</td><td colspan="2">Session: hurricane resistant || hurricane lines || houston crap || houston crap plan || hurricane climate
Prefix: houston climate actio
Correct Query: houston climate action plan</td></tr>
<tr><td>Completions (MPC_{Main}):
1. pinterest
2. paypal
3. pittsburgh penguins
4. pandora
5. prime video
6. paypal login account
7. pennlive
8. pogo official site</td><td>Completions (Trie-NLG):
1. politico
2. profootballtalk
3. politico news
4. pittsburgh pirates
5. pogo official site
6. page tour
7. philadelphia inquirer
8. pennlive</td><td>Completions(MPC_{Main}):
None

Completions(MPC_{Synth}):
1. houston climate action policy</td><td>Suggestions (Trie-NLG):
1. houston climate action plan
2. houston climate action policy
3. houston climate action play
4. houston climate plan action
5. houston climate action plan tx
6. houston climate action program
7. houston climate action plan plan
8. houstonclimate action plan</td></tr>
</table>

Other Analyses

- Ablation Study
- Trie Completion Retention Analysis
  - Number of examples where t trie completions were retained in the Trie-NLG
  - Trie retention wrt to position
- Runtime Analysis

Contributions

- We motivated the need for incorporating both popularity signals from tries and personalization signals from session to develop effective PQAC model, especially for short and unseen prefixes.
- We proposed a novel architecture, Trie-NLG, this is the first attempt of Trie knowledge augmentation in NLG models for personalized QAC.
- Achieved SOTA performance on two real prefix-to-query click behaviour QAC datasets from Bing and AOL.

Contributions

- ➢ Anonymous reviewers for Constructive feedback
- ➢ Microsoft for MAPG 2022 grant for Research and Travel Support

{Problem definition, Related Work, Methodology, Results, Contributions}

Thank You

Contact us:

cs18resch11003@iith.ac.in

nlip@cse.iith.ac.in

https://nlip-lab.github.io/

https://kaushal0494.github.io/

References

1. Bar-Yossef, Z., & Kraus, N. (2011, March). Context-sensitive query autocompletion. In Proceedings of the 20th international conference on World wide web (pp. 107-116).
2. Shokouhi, M., & Radinsky, K. (2012, August). Time-sensitive query auto-completion. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 601-610).
3. Shokouhi, M. (2013, July). Learning to personalize query autocompletion. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 103-112).
4. Kannadasan, M. R., & Aslanyan, G. (2019). Personalized Query AutoCompletion Through a Lightweight Representation of the User Context. arXiv preprint arXiv:1905.01386.
5. Bhaskar Mitra and Nick Craswell. 2015. Query auto-completion for rare prefixes. In CIKM.
6. Bhaskar Mitra. 2015. Exploring session context using distributed representations of queries and reformulations. In Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. ACM, 3–12.
7. Taihua Shao, Honghui Chen, and Wanyu Chen. 2018. Query Auto-Completion Based on Word2vec Semantic Similarity. In Journal of Physics: Conference Series, Vol. 1004. IOP Publishing, 012018
8. Yin, Di, et al. "Learning to Generate Personalized Query Auto-Completions via a Multi-View Multi-Task Attentive Approach." *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020.
9. Kim, G. (2019). Subword Language Model for Query Auto-Completion. arXiv preprint arXiv:1909.00599.
10. Park, D. H., & Chiba, R. (2017, August). A neural language model for query auto-completion. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1189-1192).
11. Qureshi, Abdur Rehman Anwar, and M. Ali Akcayol. "Long Short-Term Memory Based Query Auto-Completion." *2021 8th International Conference on Electrical and Electronics Engineering (ICEEE)*. IEEE, 2021.