Adaptive Re-Ranking as an Information-Seeking Agent

Sean MacAvaney
Nicola Tonellotto
Craig Macdonald

@macavaney

PASIR @ CIKM 2022

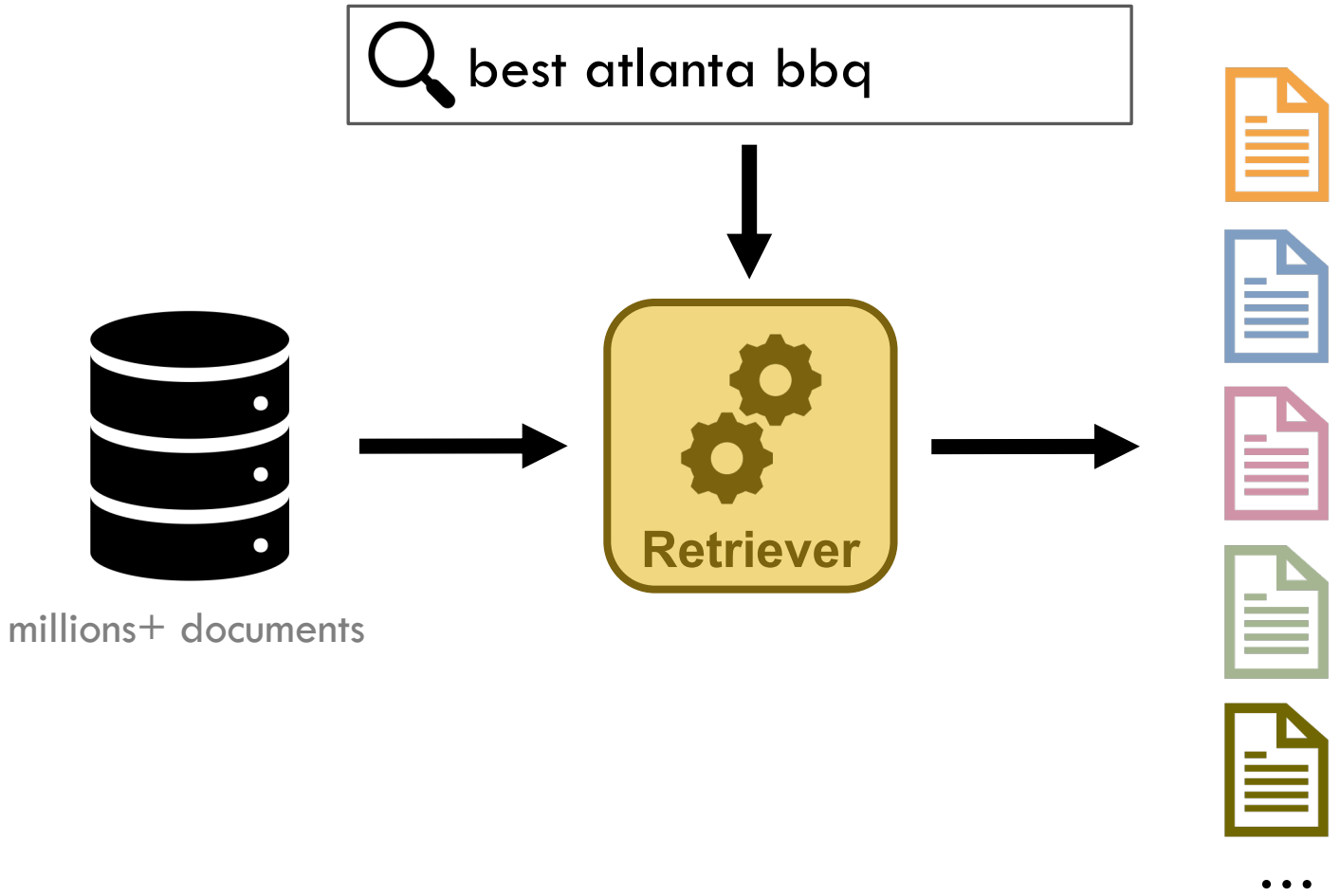
Terrier

University of Glasgow

Università di Pisa

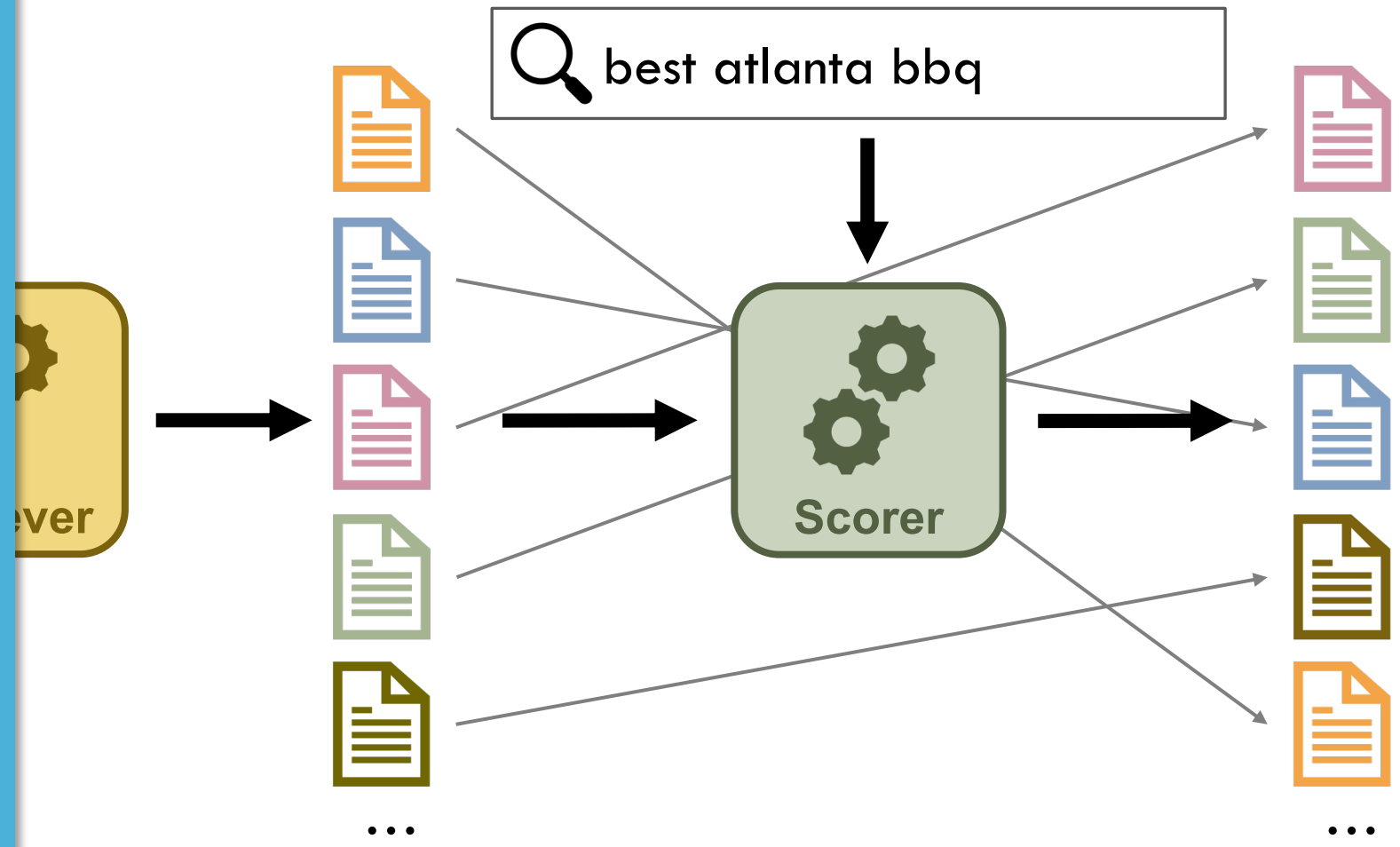
Re-Ranking

Stage 1: Retrieve candidate results from a corpus



Re-Ranking

Stage 2: Re-order the candidates

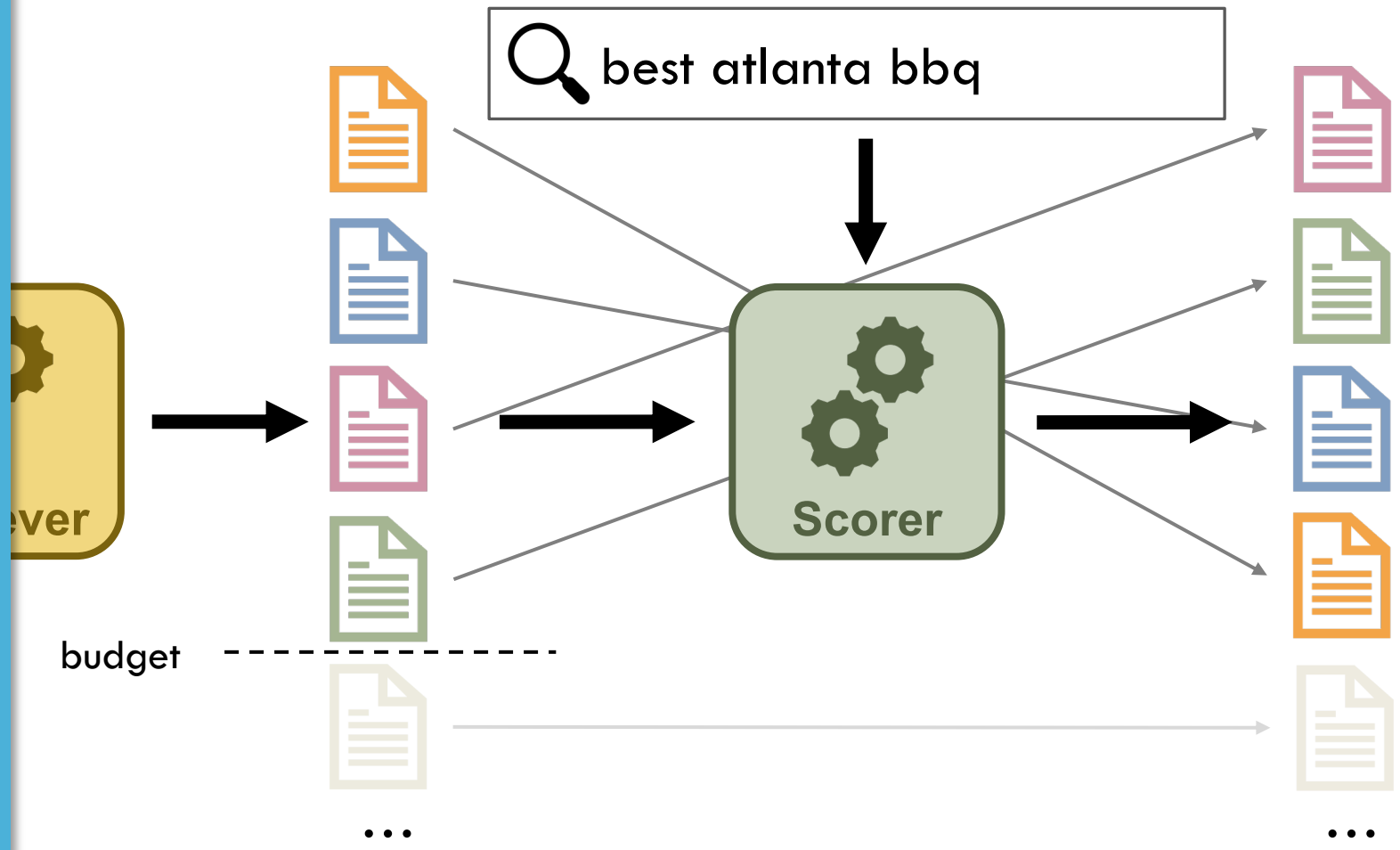


hundreds/thousands of documents

same documents, new order

Re-Ranking

Stage 2: Re-order the candidates



hundreds/thousands of documents

same documents, new order

Stage 2: Re-order the candidates

Re-Ranking

hundreds/thousands of documents

same documents, new order

What if we consider the re-ranking process as an agent?

Environment:

Initial Ranking

...

Final Ranking

[empty]

Agent:

Actions:

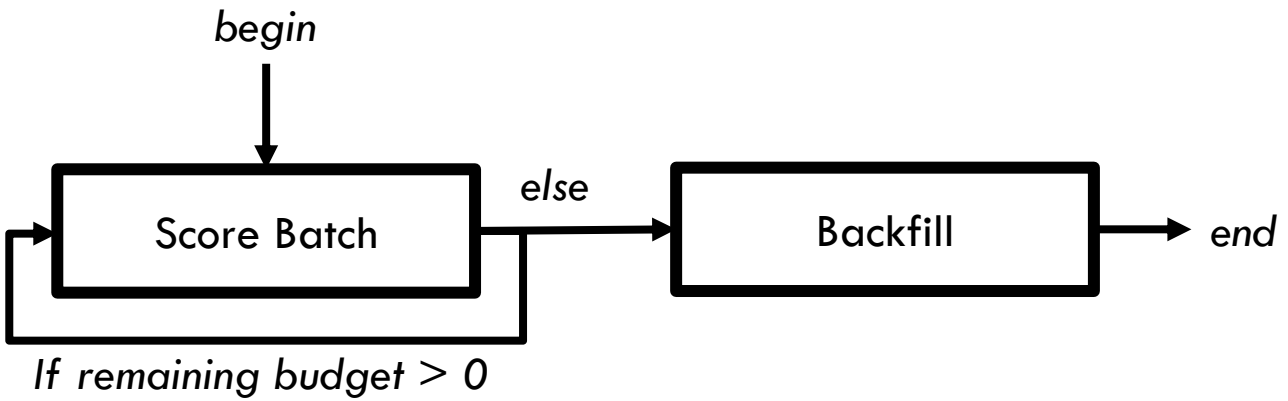
Score Batch

Backfill

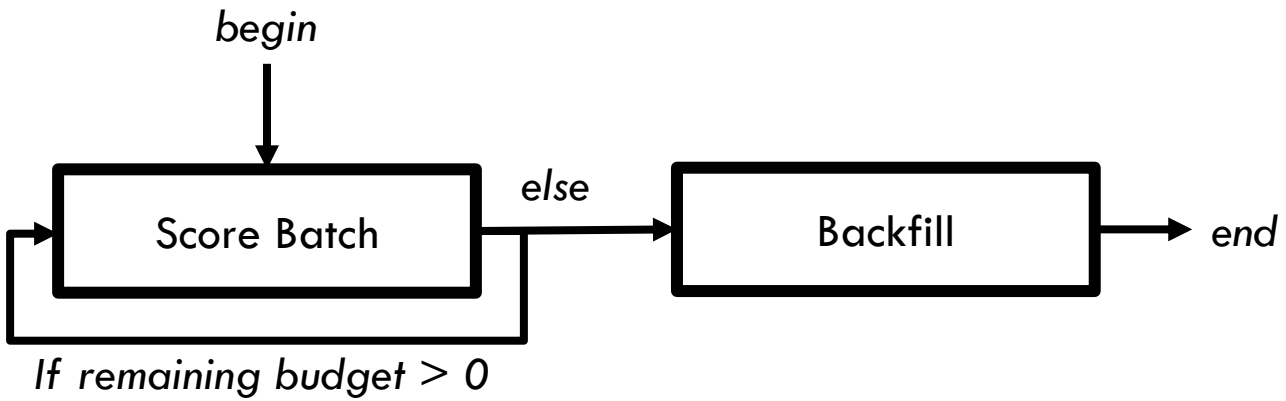
State:

Remaining budget

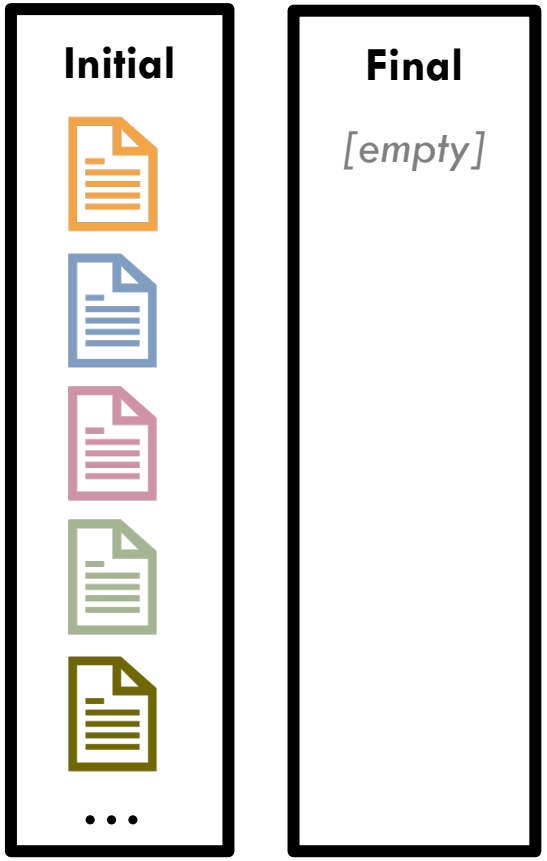
Traditional Re-Ranking



Traditional Re-Ranking



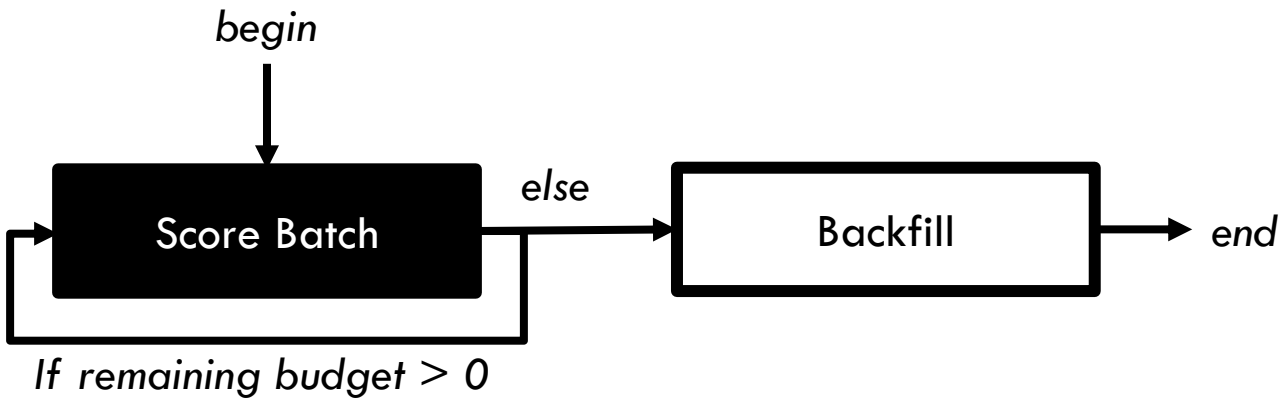
Environment:



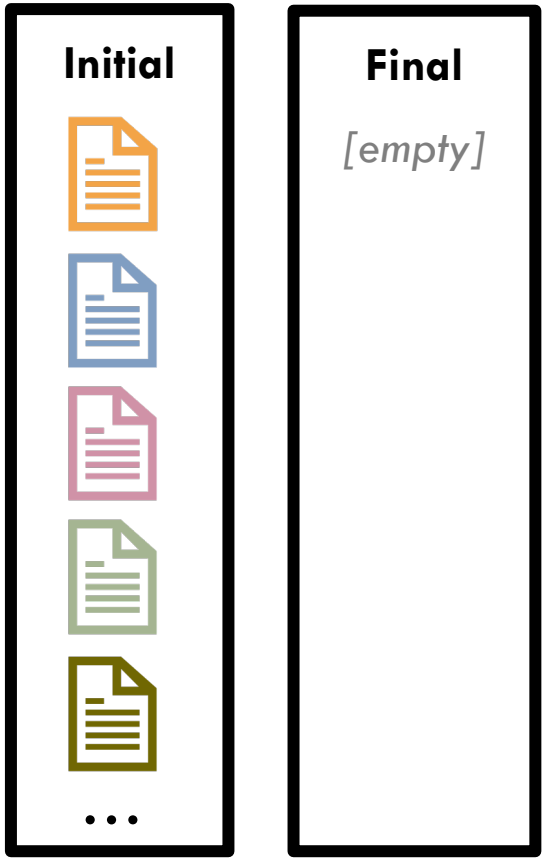
State:

Remaining budget = 3

Traditional Re-Ranking



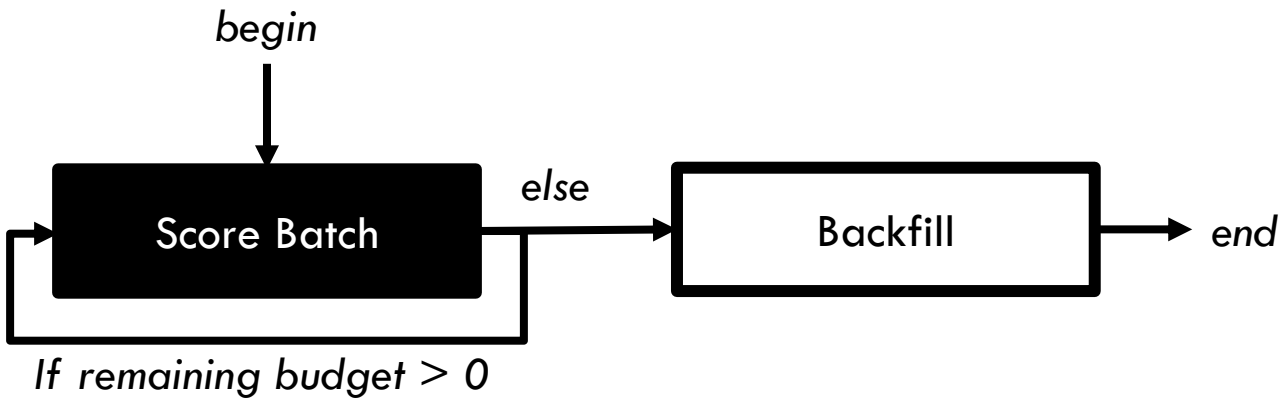
Environment:



State:

Remaining budget = 3

Traditional Re-Ranking

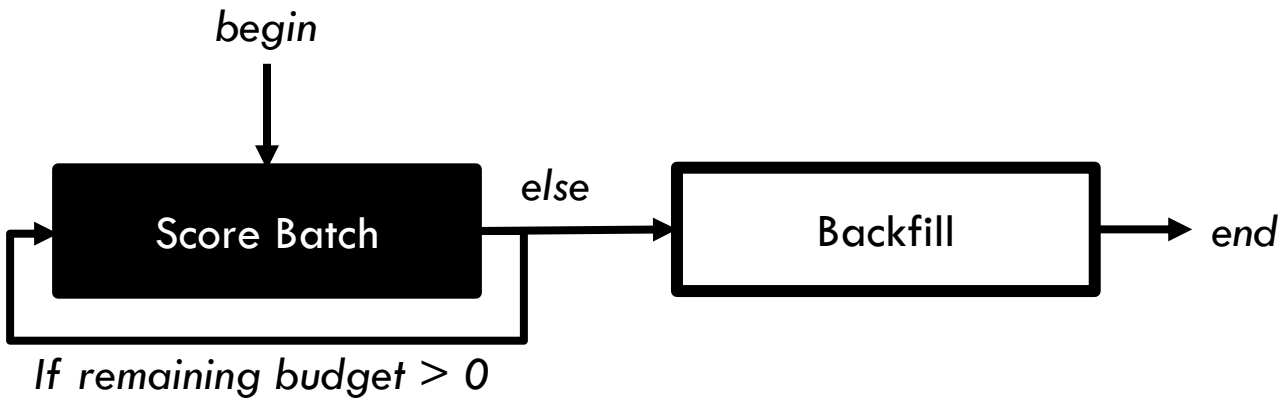


Environment:

State:

Remaining budget = 2

Traditional Re-Ranking

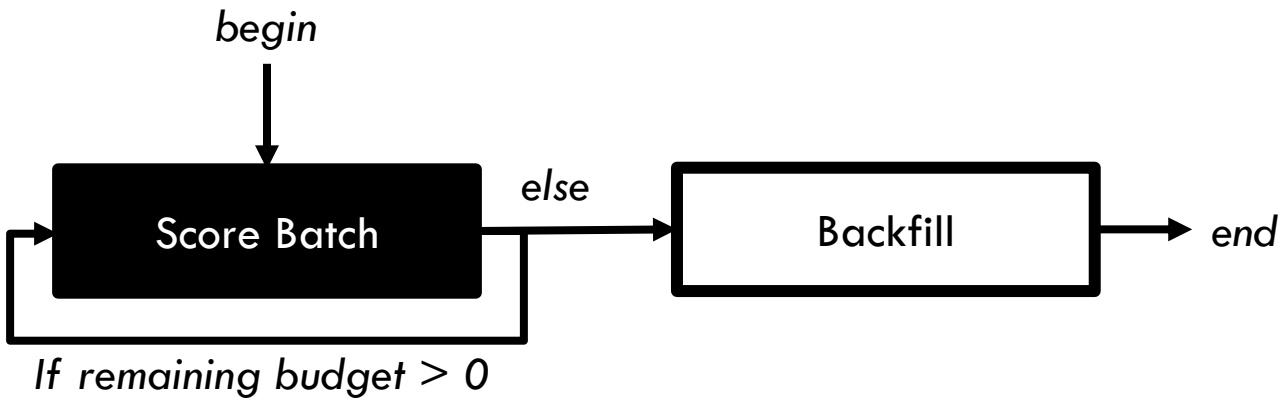


Environment:

State:

Remaining budget = 1

Traditional Re-Ranking

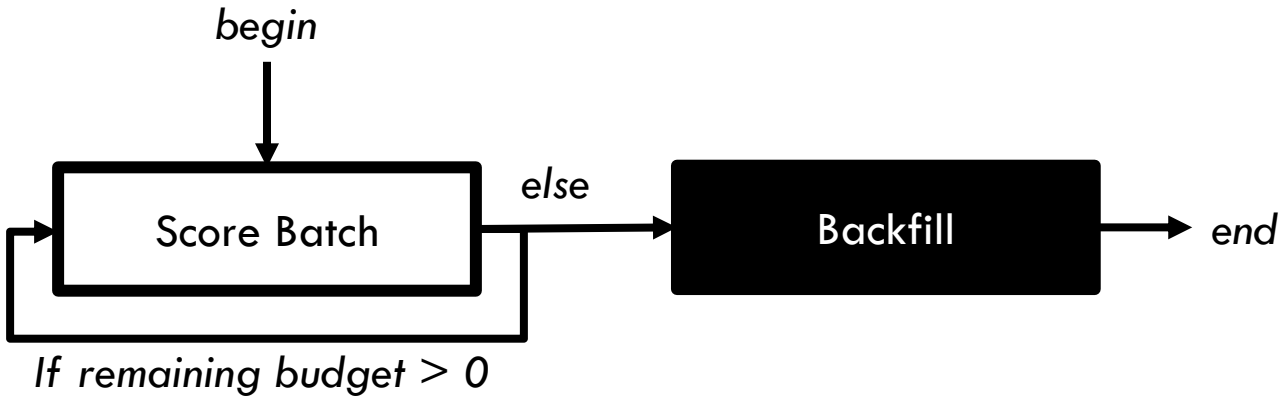


Environment:

State:

Remaining budget = 0

Traditional Re-Ranking

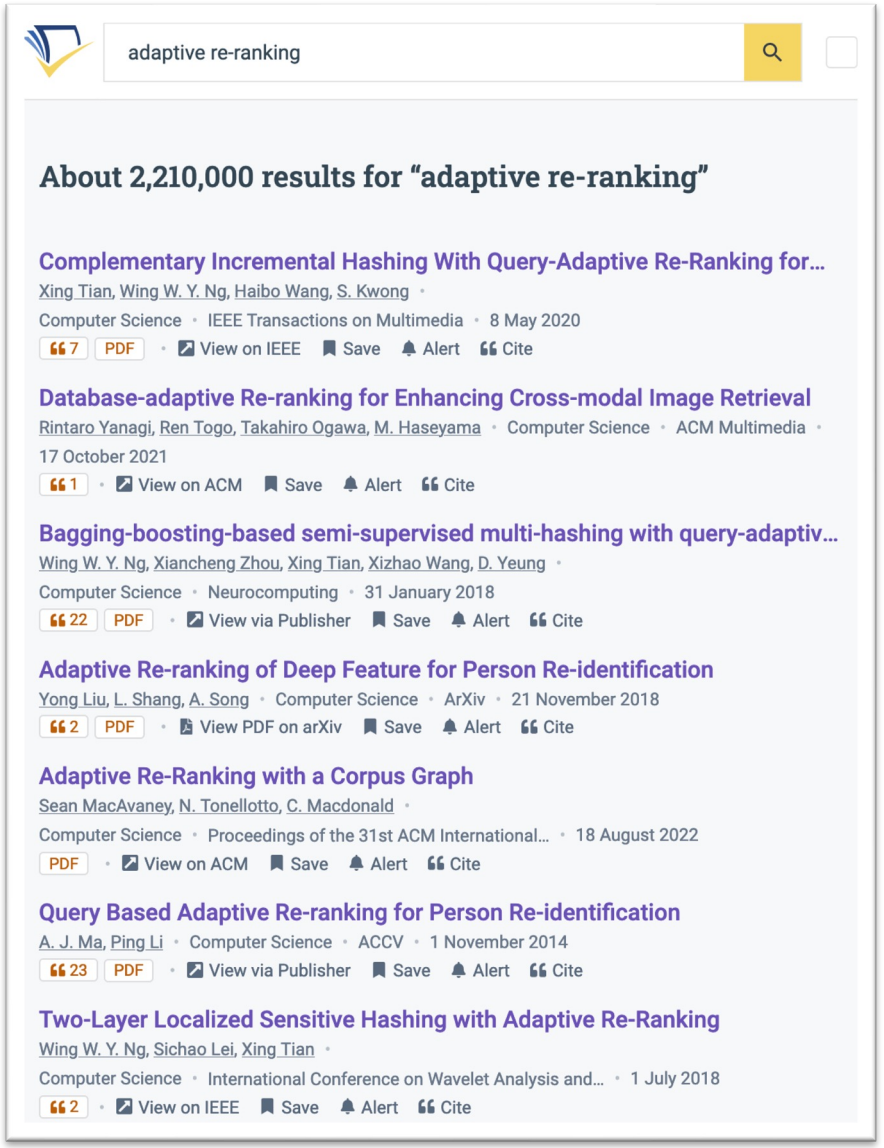


Environment:

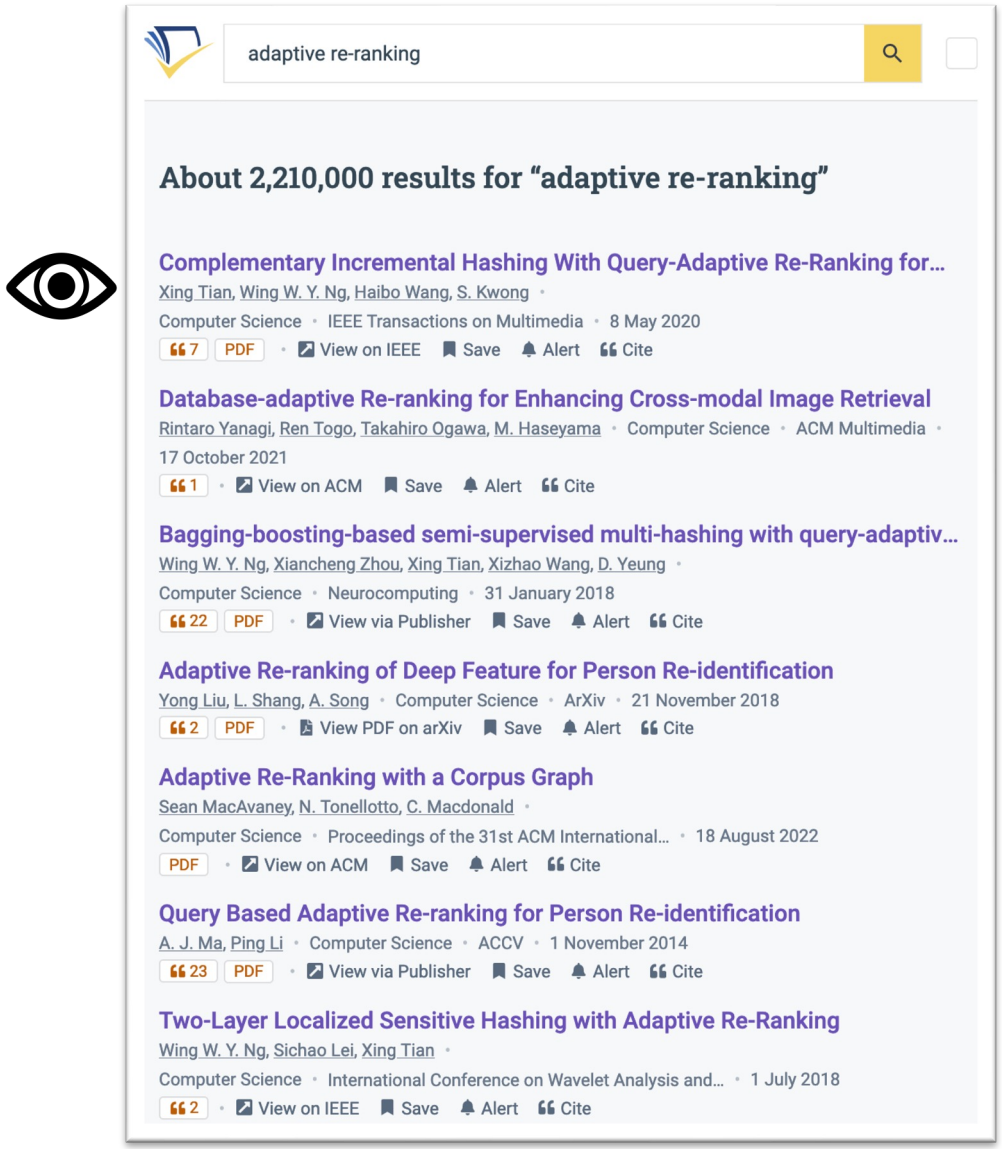
State:

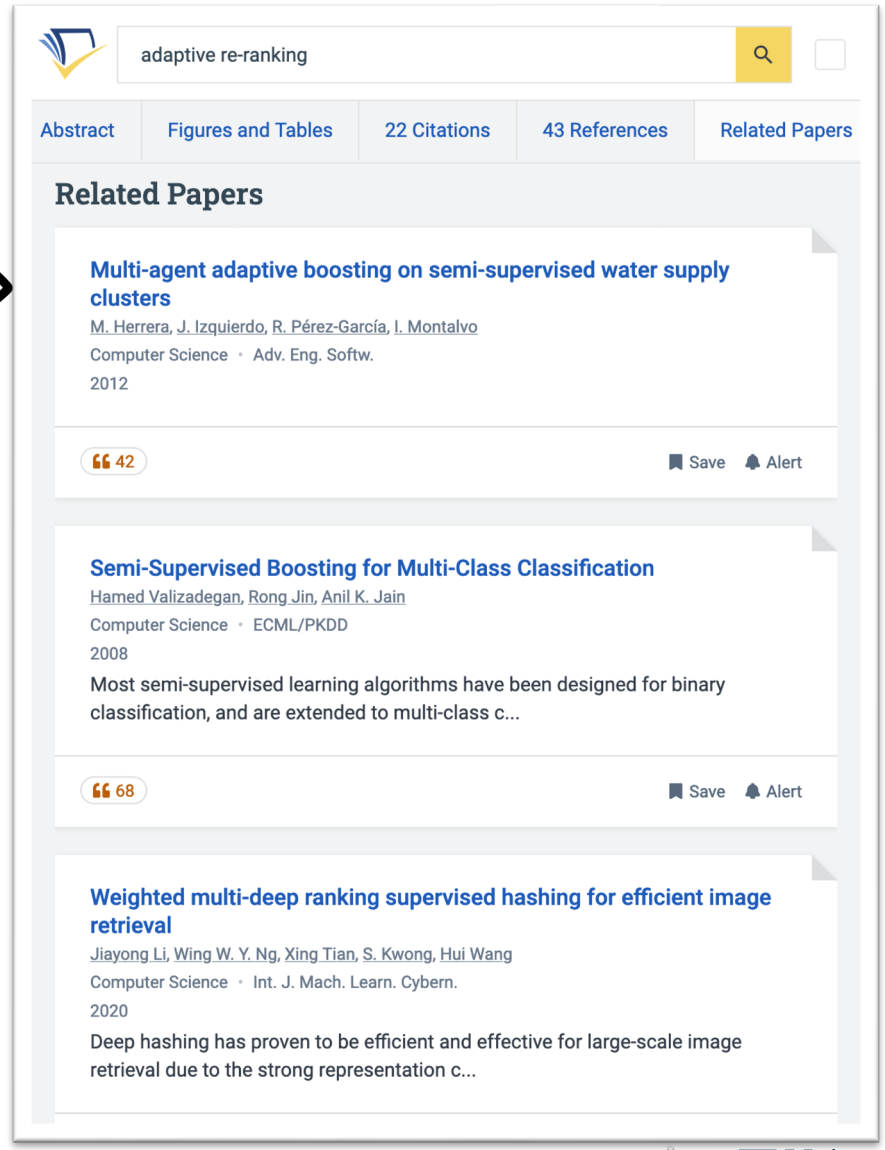
Remaining budget = 0

This is similar to how a user might traverse search results.



But users may apply other strategies, too.





Introducing the find neighbours action

Based on the Clustering Hypothesis: Relevant documents are often "close" to one another.

Environment:

Initial Ranking

...

Final Ranking

[empty]

Agent:

Actions:

- Score Batch (initial)
- Backfill
- Find Neighbours
- Score Batch (frontier)

State:

Remaining budget

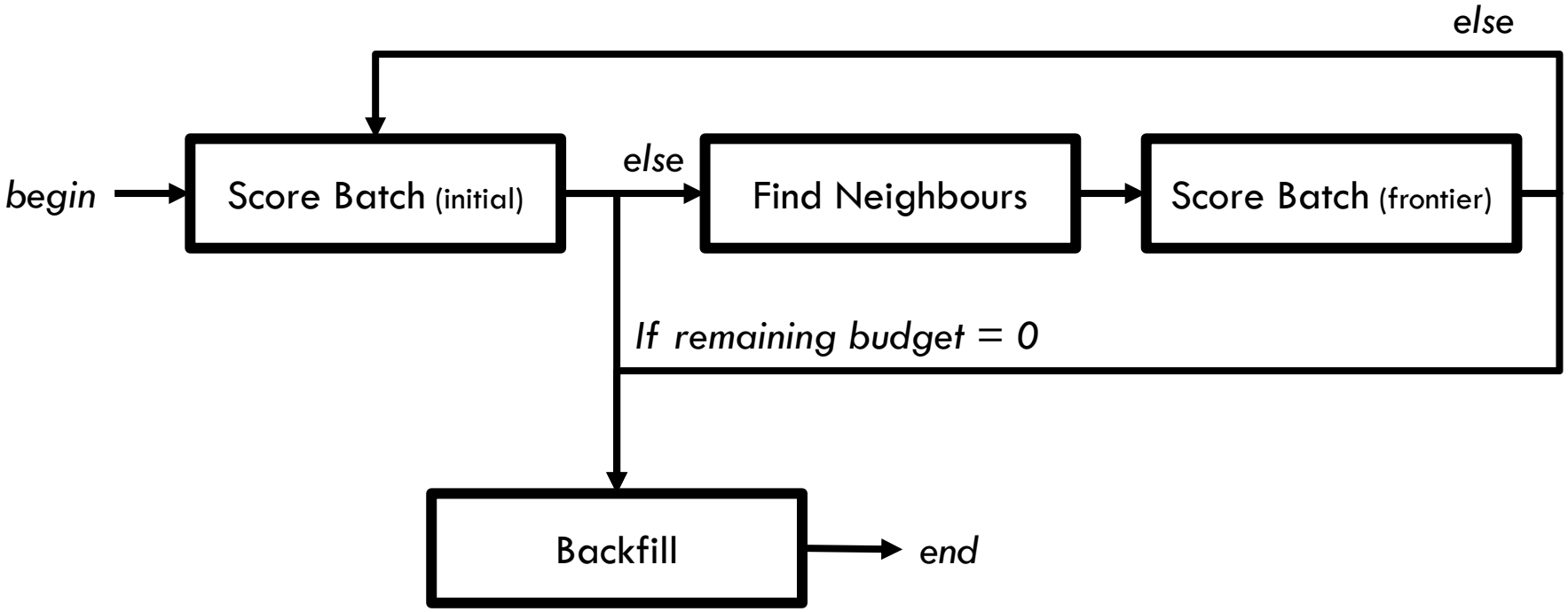
Frontier

[empty]

Adaptive Re-Ranking

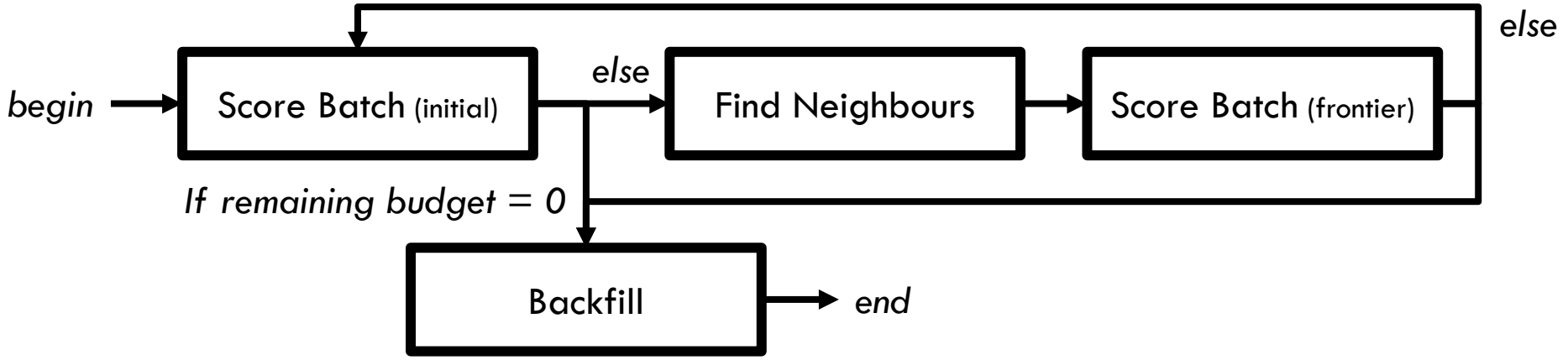
ALTERNATE

"Alternate" Strategy: Alternate between scoring from initial pool and neighbours. [cikm2022]

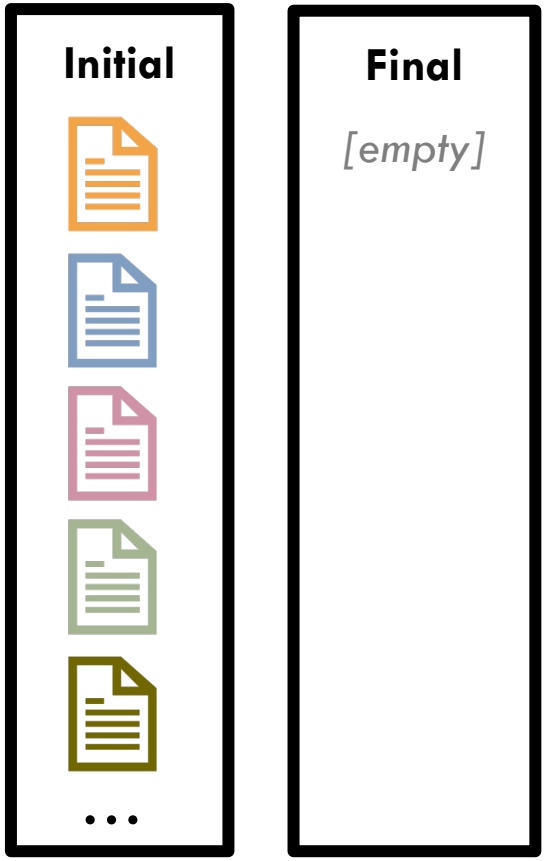


Adaptive Re-Ranking

ALTERNATE

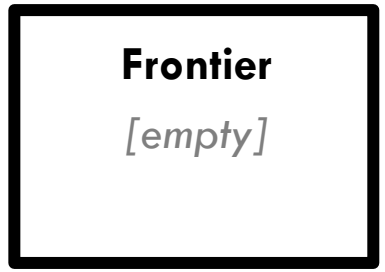


Environment:



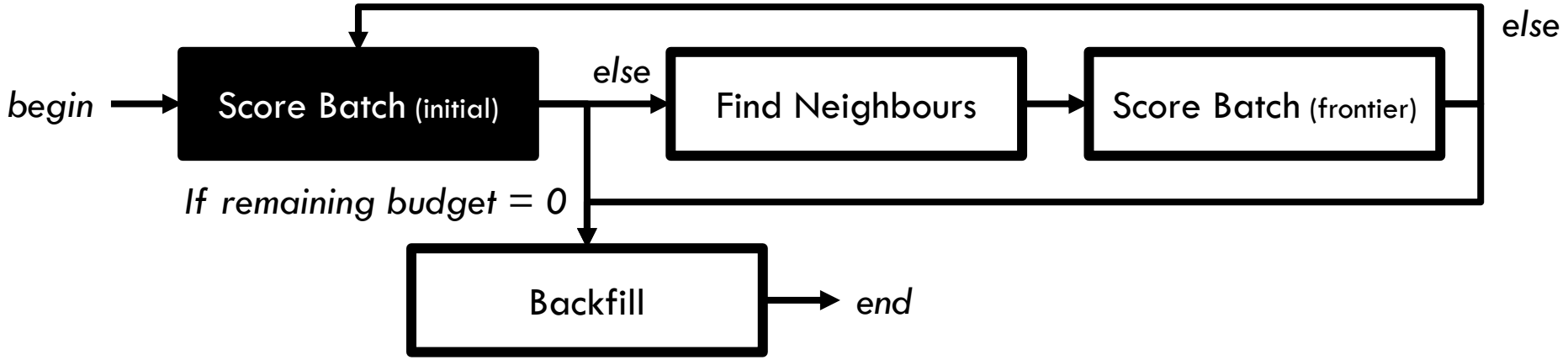
State:

Remaining budget = 3

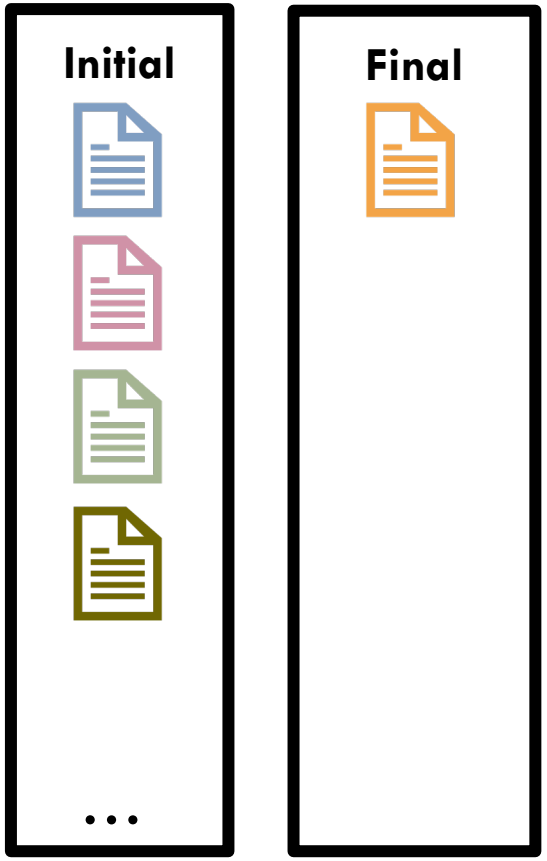


Adaptive Re-Ranking

ALTERNATE



Environment:



State:

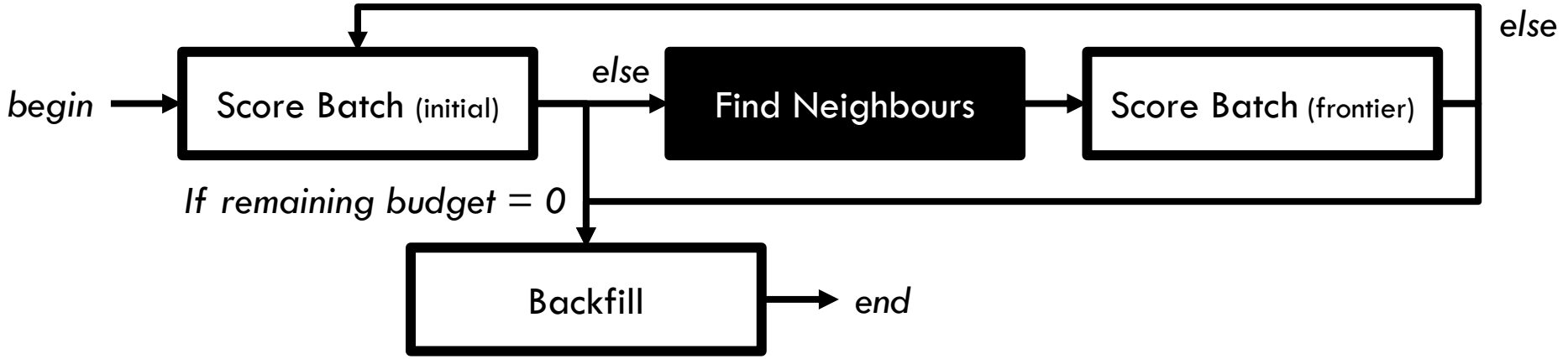
Remaining budget = 2

Frontier

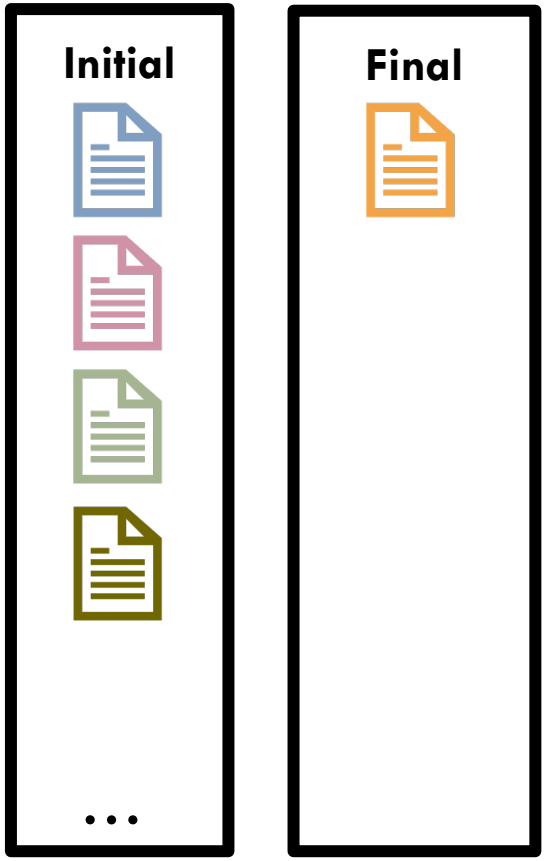
[empty]

Adaptive Re-Ranking

ALTERNATE



Environment:

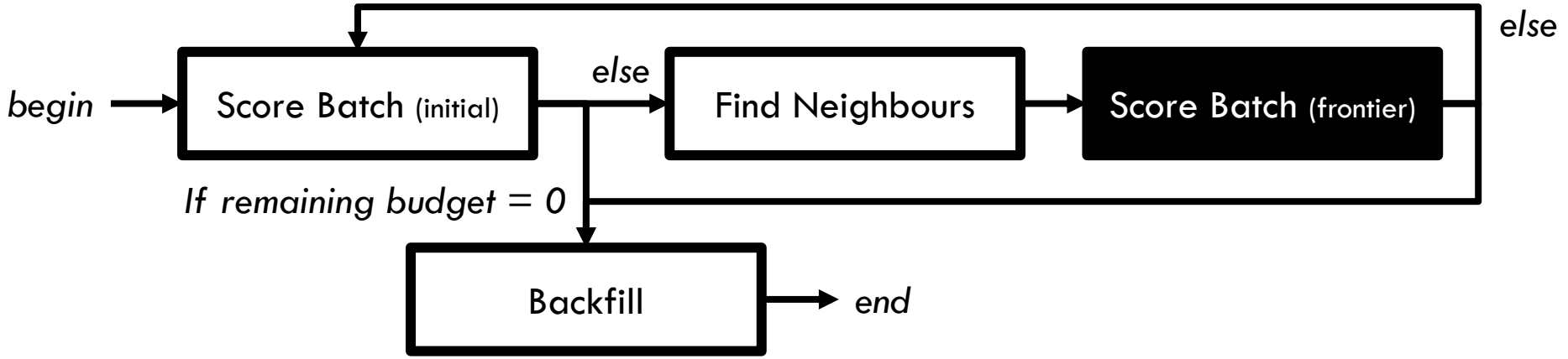


State:

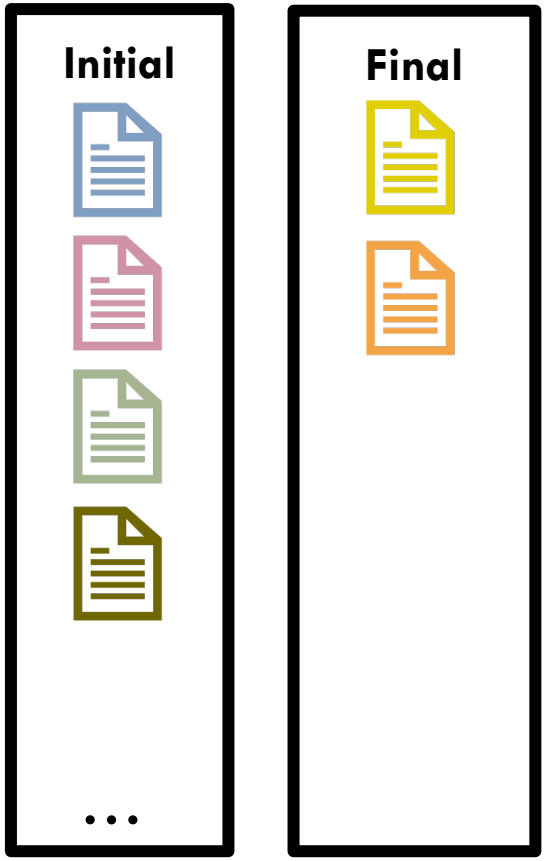
Remaining budget = 2

Adaptive Re-Ranking

ALTERNATE



Environment:



State:

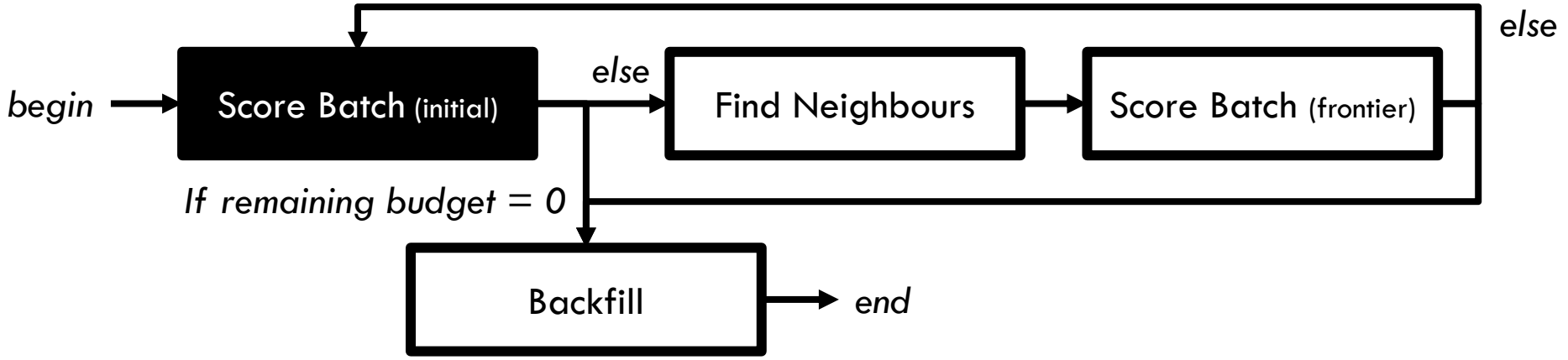
Remaining budget = 1

Frontier

[empty]

Adaptive Re-Ranking

ALTERNATE



Environment:

State:

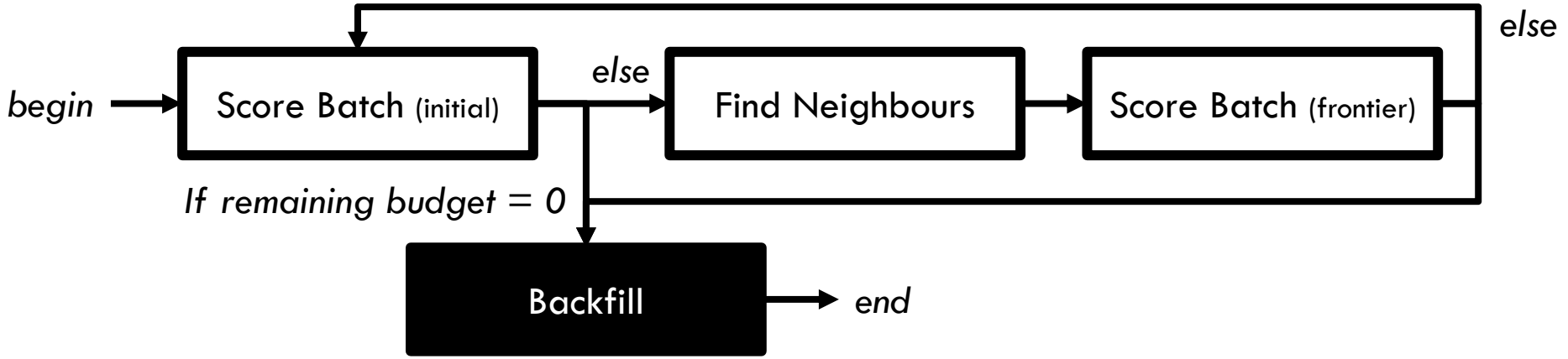
Remaining budget = 0

Frontier

[empty]

Adaptive Re-Ranking

ALTERNATE



Environment:

State:

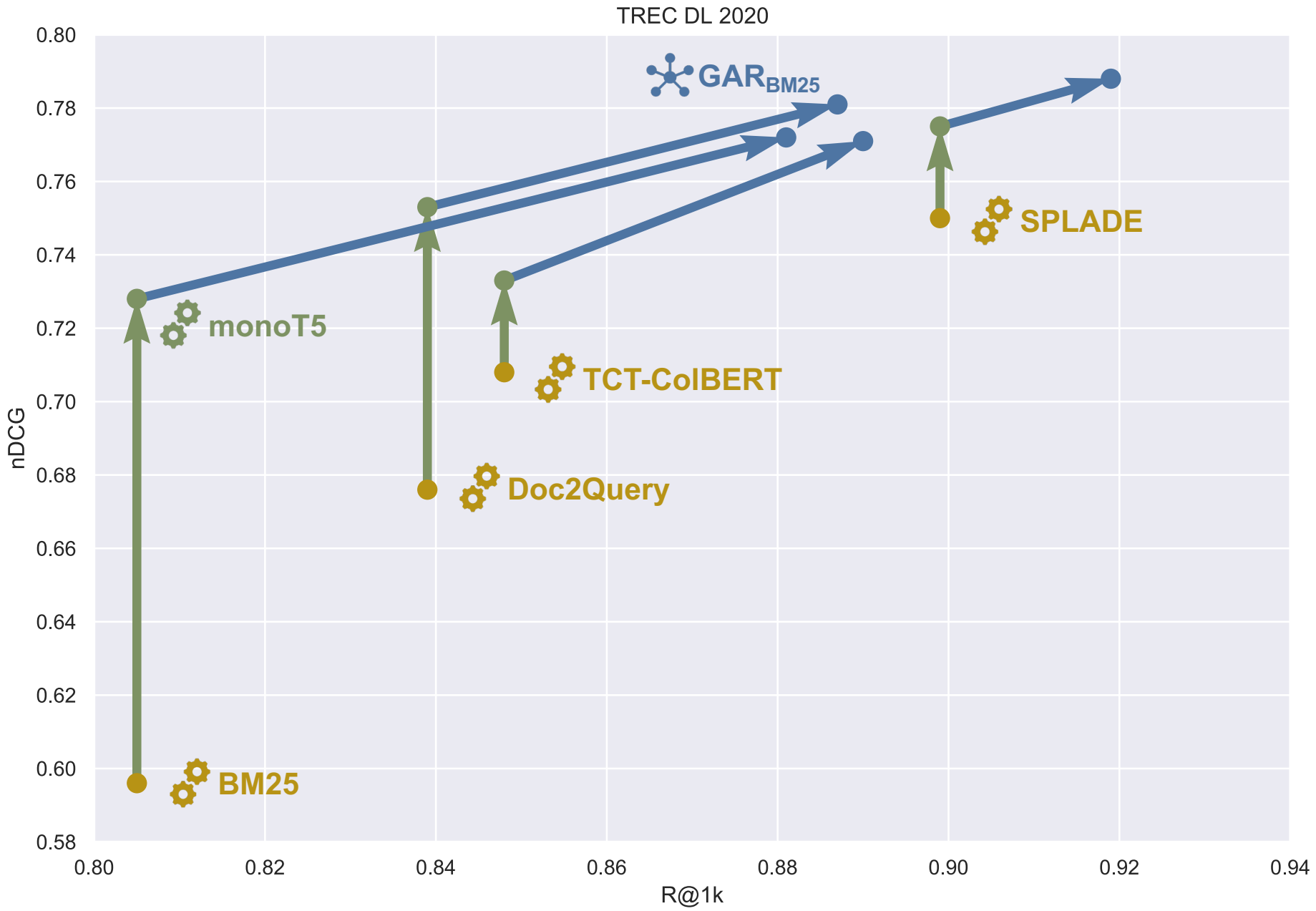
Remaining budget = 0

Frontier

[empty]

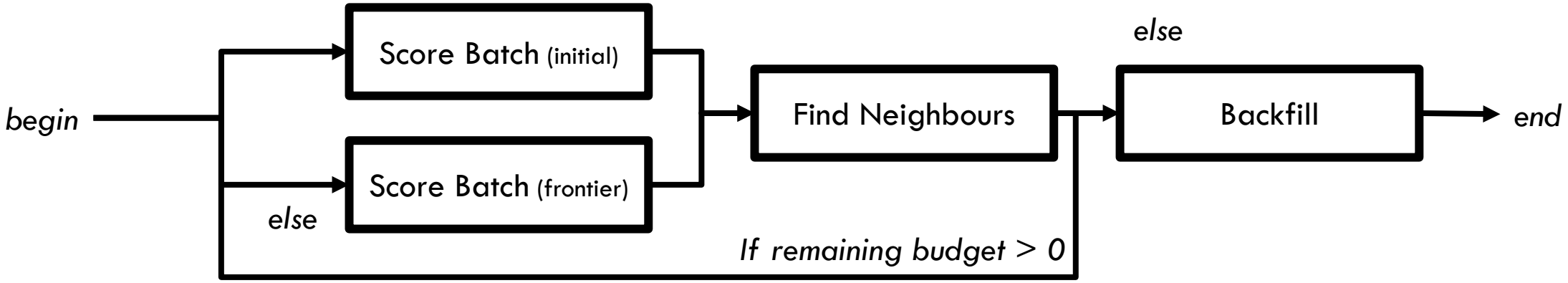
Adaptive Re-Ranking

ALTERNATE



“Alternate” works, but there must be a smarter strategy, right?

If batch from initial will give more relevant documents than frontier



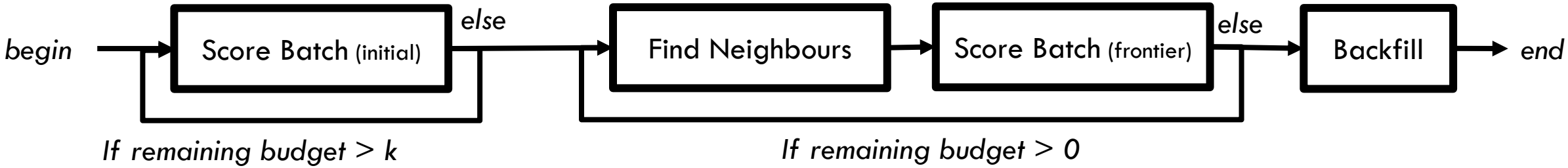
There's considerable room to improve upon Alternate.

Pipeline	Agent	TREC DL 2019 (dev)		TREC DL 2020 (test)	
		GAR_{bm25}	GAR_{tct}	GAR_{bm25}	GAR_{tct}
BM25»MonoT5	Non-Adaptive	0.699	0.699	0.711	0.711
	Alternate	0.726	0.743	0.743	0.749
	Oracle	**0.747**	**0.786**	**0.748**	**0.768**
TCT»MonoT5	Non-Adaptive	0.704	0.704	0.693	0.693
	Alternate	0.733	0.724	0.719	0.710
	Oracle	**0.793**	**0.766**	**0.762**	**0.754**
D2Q»MonoT5	Non-Adaptive	0.747	0.747	0.731	0.731
	Alternate	0.757	0.766	0.748	0.748
	Oracle	**0.797**	**0.798**	**0.791**	**0.793**
SPLADE»MonoT5	Non-Adaptive	0.737	0.737	0.731	0.731
	Alternate	0.745	0.737	0.737	0.734
	Oracle	**0.807**	**0.783**	**0.777**	**0.781**

nDCG

Two-Phase:

Spend *k* of your budget on the initial re-ranking upfront, then move to neighbours.

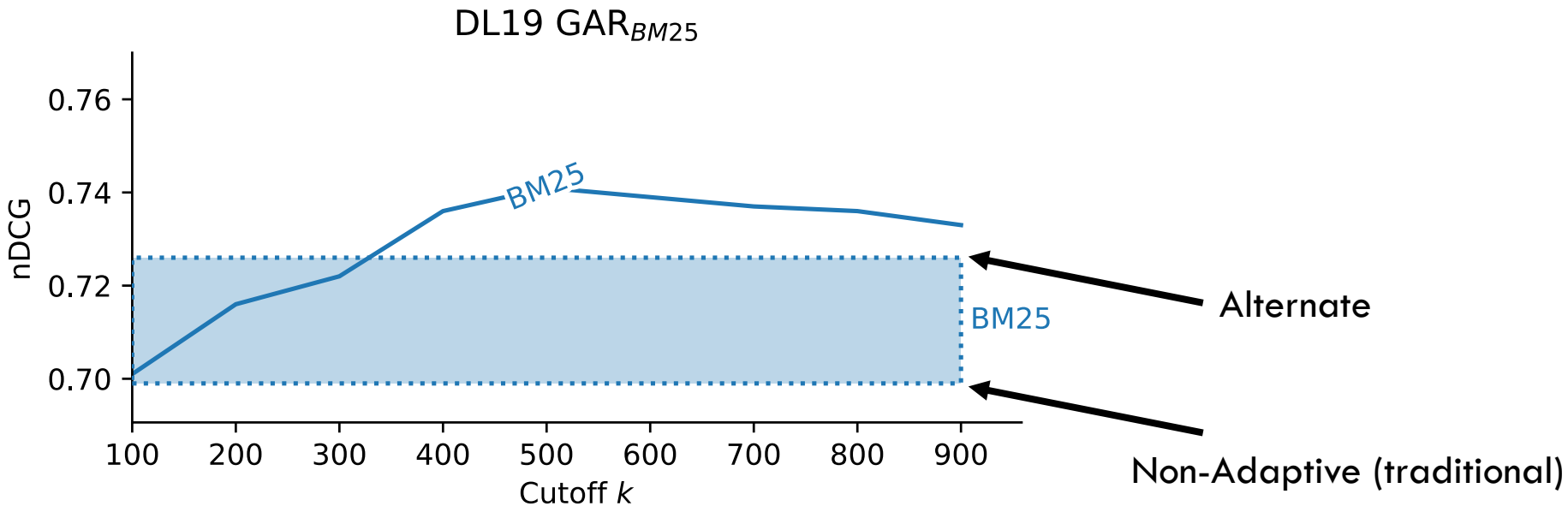


Adaptive

Re-Ranking

TWO-PHASE

Can outperform Alternate… But not always.

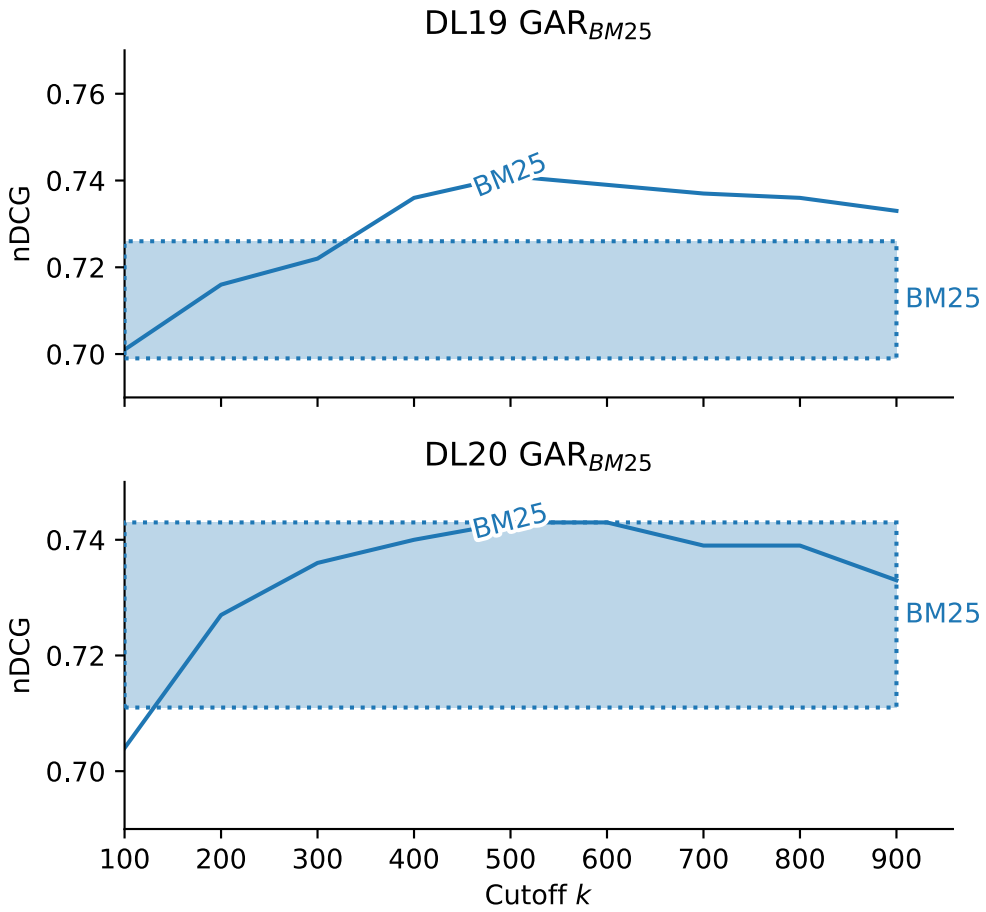


Adaptive

Re-Ranking

TWO-PHASE

Can outperform Alternate… But not always.

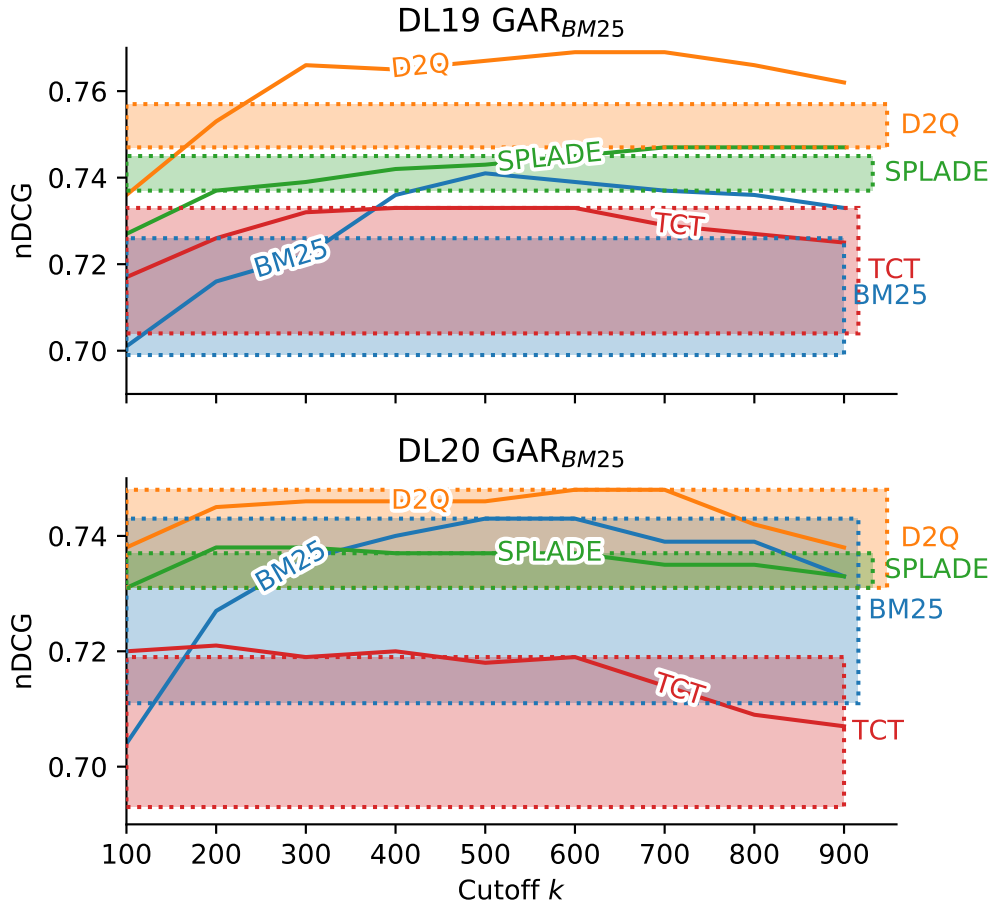


Adaptive Re-Ranking

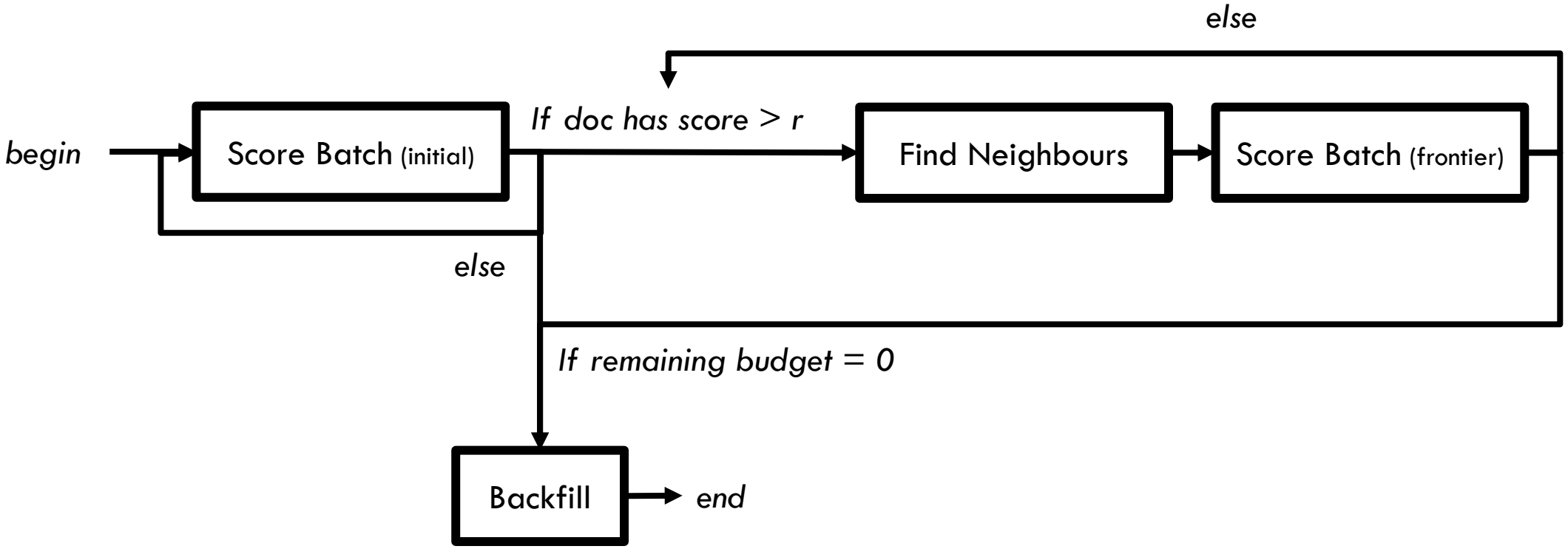
TWO-PHASE

Can outperform Alternate… But not always…

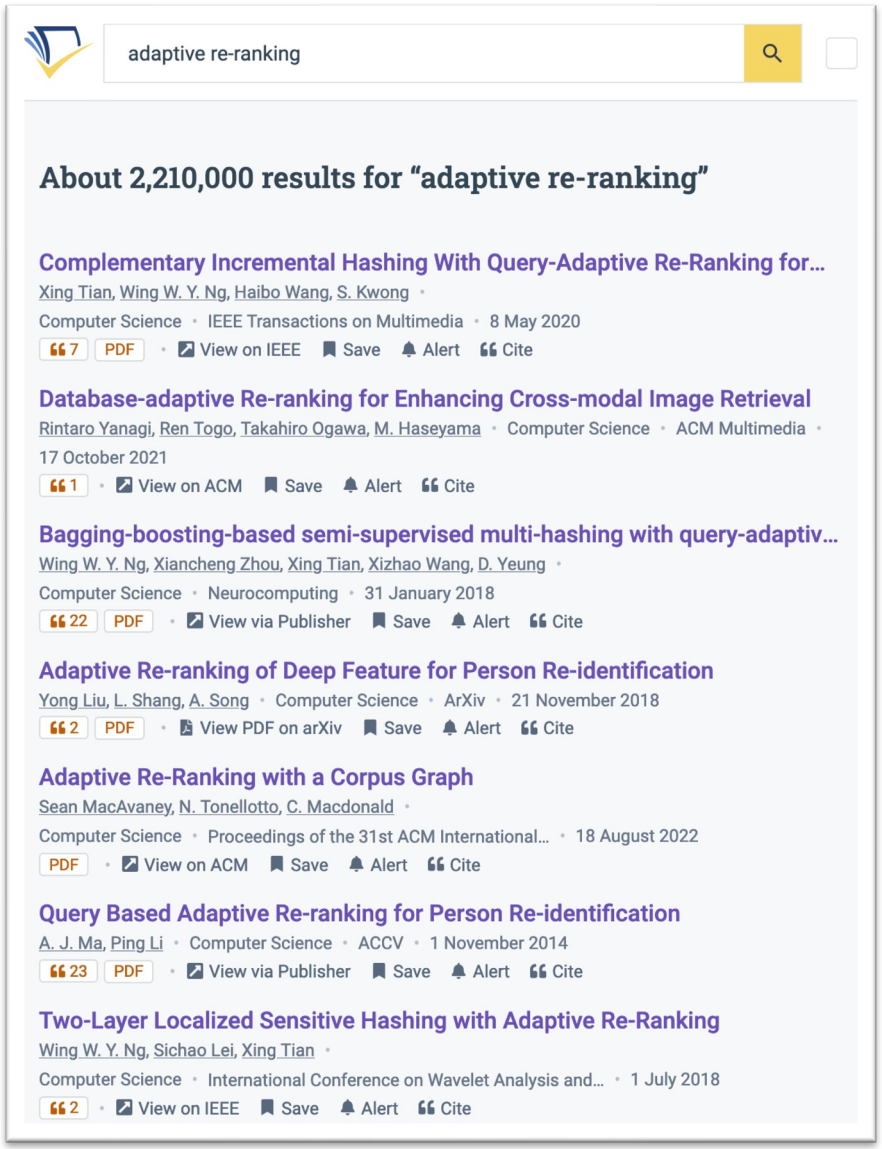
And is sensitive to *k*

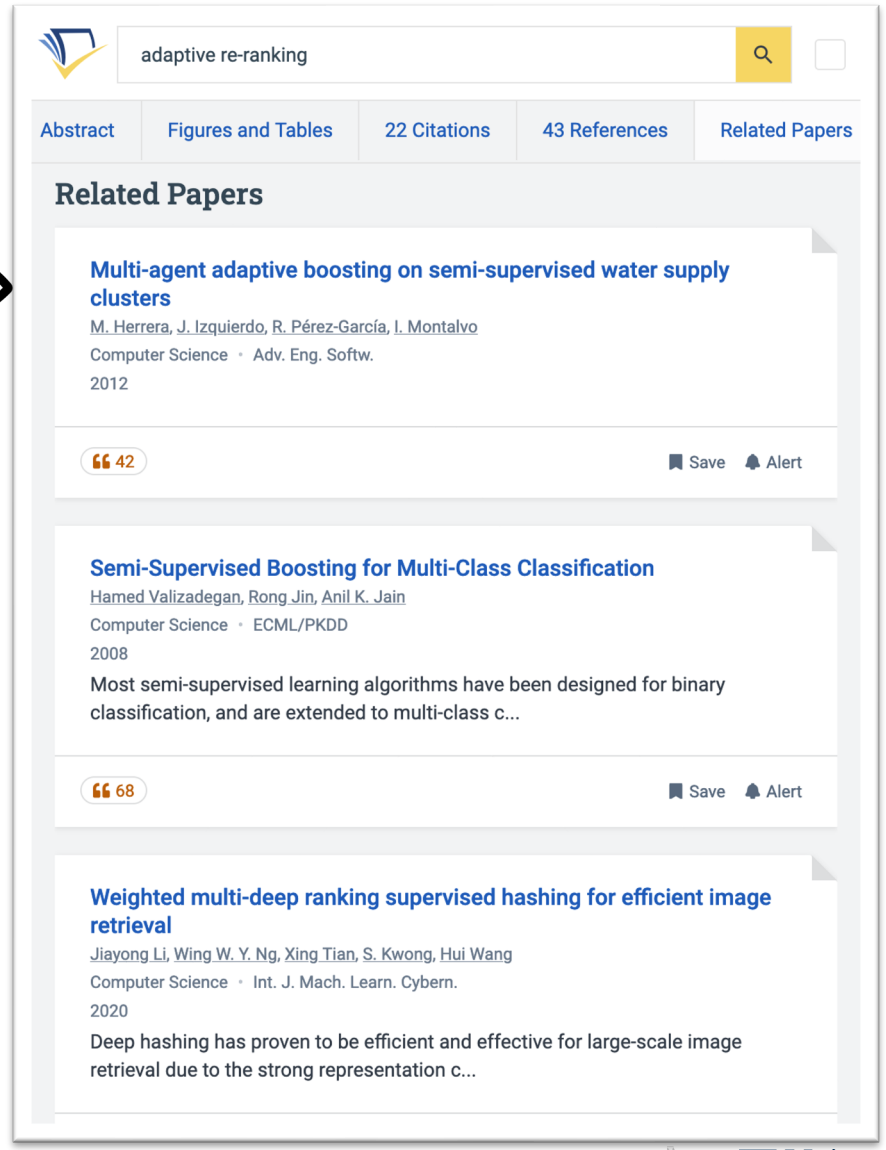


Threshold:
Only explore neighbours of documents that meet relevance score threshold *r*.



But users may apply other strategies, too.



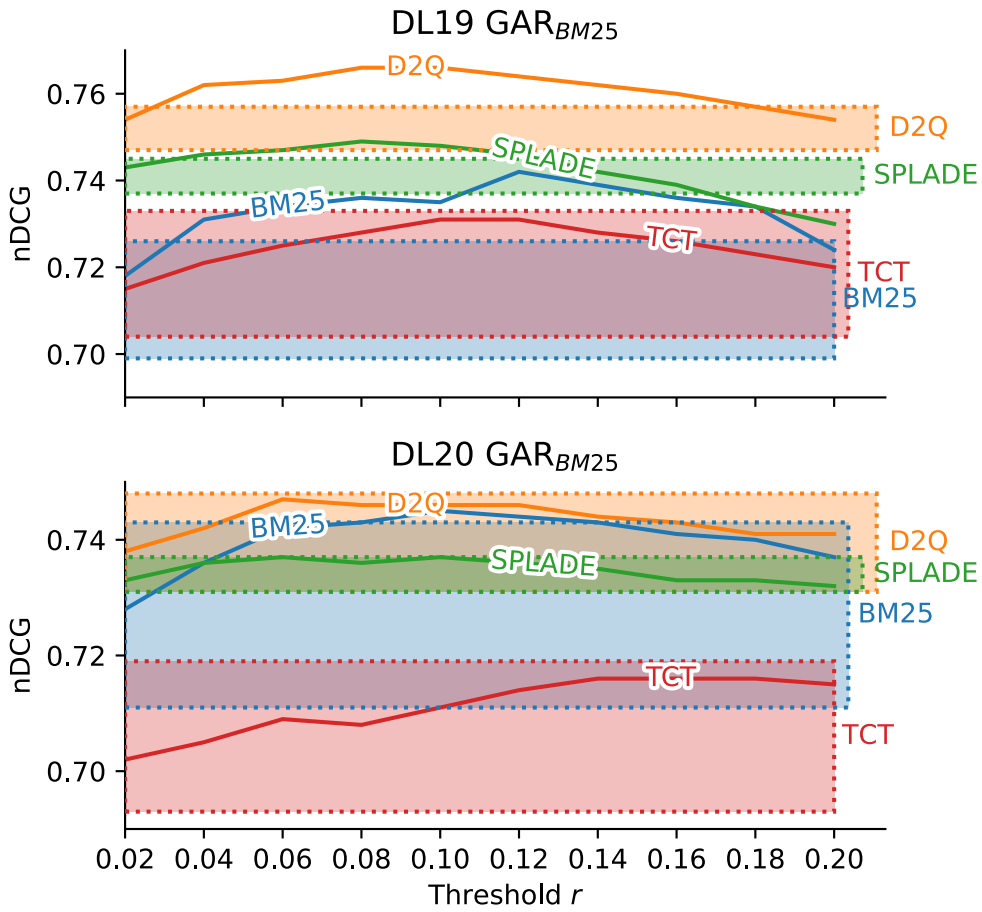


Adaptive Re-Ranking

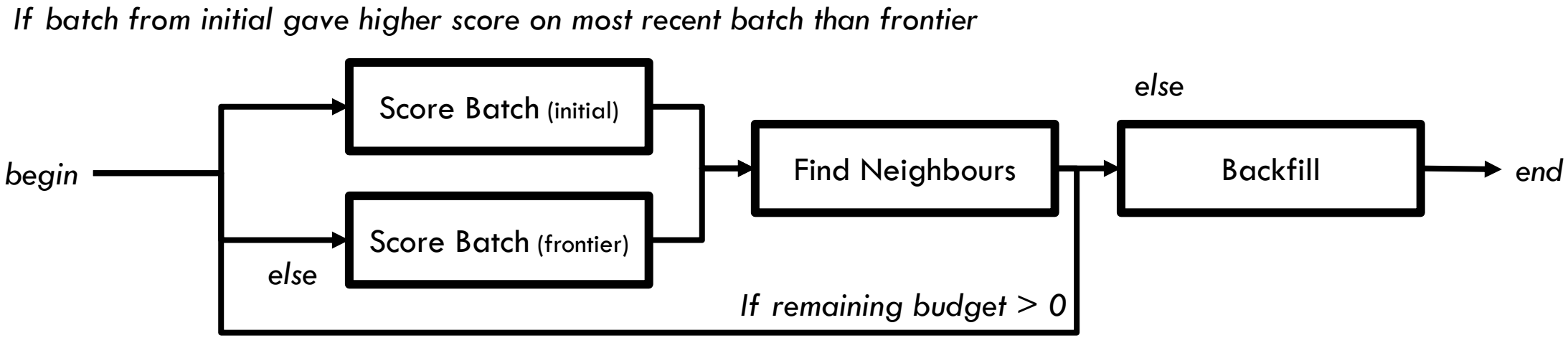
THRESHOLD

Can outperform Alternate… But not always…

And is sensitive to *r*



Greedy:
Score from initial ranking or frontier based on what gave the highest score recently.



Adaptive Re-Ranking

GREEDY

Also no better than Alternate

Pipeline	Agent	TREC DL 2019 (dev) GAR_{bm25}	TREC DL 2019 (dev) GAR_{tct}	TREC DL 2020 (test) GAR_{bm25}	TREC DL 2020 (test) GAR_{tct}
BM25»MonoT5	Non-Adaptive	0.699	0.699	0.711	0.711
	Oracle	0.747	0.786	0.748	0.768
	Alternate	0.726	0.743	0.743	**0.749**
	TwoPhase	0.741	0.743	0.743	0.744
	Threshold	**0.742**	**0.751**	**0.744**	0.744
	Greedy	0.723	0.737	0.743	0.744
TCT»MonoT5	Non-Adaptive	0.704	0.704	0.693	0.693
	Oracle	0.793	0.766	0.762	0.754
	Alternate	**0.733**	0.724	**0.719**	**0.710**
	TwoPhase	**0.733**	0.722	**0.719**	0.707
	Threshold	0.731	0.720	0.711	0.705
	Greedy	0.731	**0.725**	0.713	0.708
D2Q»MonoT5	Non-Adaptive	0.747	0.747	0.731	0.731
	Oracle	0.797	0.798	0.791	0.793
	Alternate	0.757	0.766	**0.748**	**0.748**
	TwoPhase	**0.769**	**0.767**	**0.748**	0.747
	Threshold	0.766	**0.767**	0.746	0.745
	Greedy	0.754	0.757	0.744	**0.748**
SPLADE»MonoT5	Non-Adaptive	0.737	0.737	0.731	0.731
	Oracle	0.807	0.783	0.777	0.781
	Alternate	0.745	0.737	0.737	0.734
	TwoPhase	**0.769**	**0.764**	**0.748**	0.736
	Threshold	0.766	0.759	0.746	**0.744**
	Greedy	0.747	0.740	0.734	0.734

Outlook

Let's make re-rankers smarter!

Moving beyond simply re-scoring an initial set of documents

Allowing re-rankers to pull in new relevant documents.

Simple heuristics are strong, but there's room to improve

Oracle results show considerable headroom.

Fertile ground: more actions, multiple initial rankings, etc.

Learning-to-re-rank? Learning to stop? Multi-armed bandits? Etc.

Code: https://github.com/terrierteam/pyterrier_adaptive

Adaptive Re-Ranking as an Information-Seeking Agent

- Building off CIKM 2022 proposal for "adaptive re-ranking"
- There is considerable room for improvement in pipelines by employing smarter agents
- The "Alternate" adaptive strategy is a remarkably strong baseline
- Sets the stage for work that learns a good strategy for re-ranking

SEAN MACAVANEY
NICOLA TONELLOTTO
CRAIG MACDONALD

@MACAVANEY

PASIR @ CIKM 2022

University of Glasgow

UNIVERSITÀ DI PISA

EPSRC grant EP/R018634/1: Closed-Loop Data Science for Complex, Computationally- & Data-Intensive Analytics.