Spring25 CS598YP

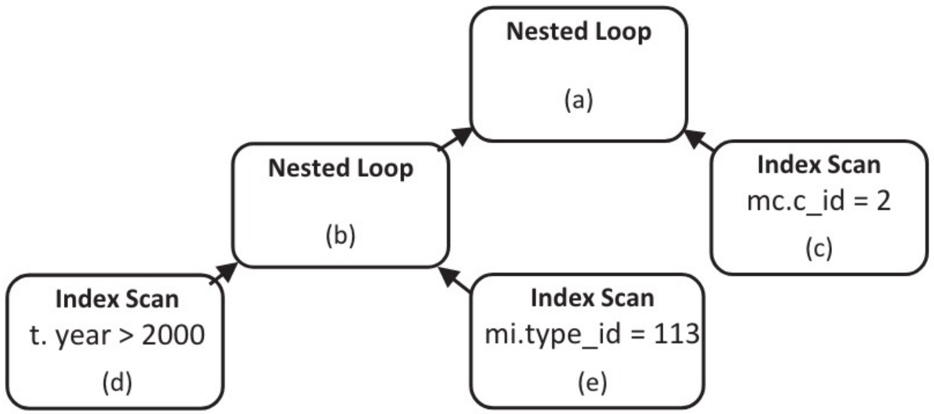
15.2: QueryFormer

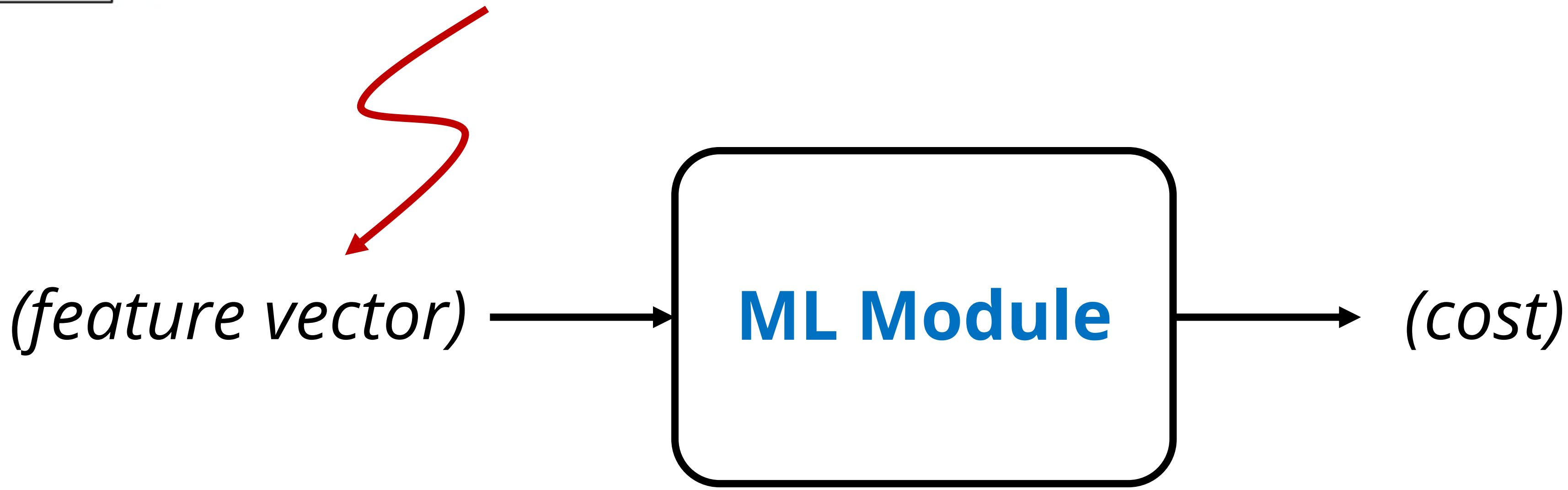
Yongjoo Park

University of Illinois Urbana-Champaign

Cost estimation as ML problem

```
SELECT * FROM
    title t,
    movie_info mi,
    movie_companies mc,
WHERE
    t.id = mi.movie_id AND
    t.id = mc.movie_id AND
    mi.type_id = 113 AND
    mc.c_id = 2 AND
    t.year > 2000
```





Previous approaches are less effective

Table 1: Summary of existing solutions to query plan representation.

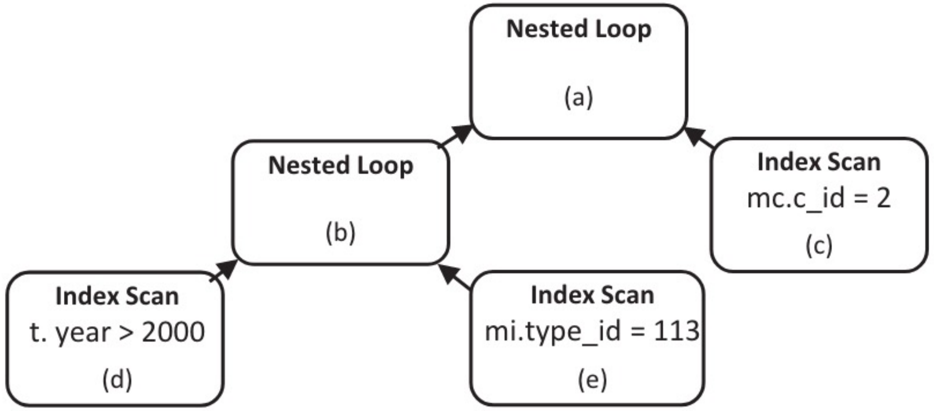
Category	Paper	Task	Parent-Children Dependency	Long Path Information Flow	Database Statistics	Training Difficulty
Flattened	AVGDL [38]	View Selection	No	Yes	NA	Hard
Tree-RNN	RTOS [36]	Join Order Selection	Yes	Yes	NA	Hard
	E2E-Cost [30]	Cost, Cardinality	Yes	Yes	Sample	Hard
	Plan-Cost [19]	Cost Estimation	Yes	Yes	Estimated card, cost	Hard
Tree-CNN	NEO [17]	Optimization	Yes	No	Estimated card	Easy
	BAO [16]	Optimization	Yes	No	Estimated card, cost	Easy
	Prestroid [39]	Cost Estimation	Yes	No	NA	Easy
Feature Vectors	ReJOIN [18]	Join Order Selection	No	No	NA	Easy
	AIMeetsAI [9]	Index Selection	No	No	Estimated card, cost	Easy
	LQPP [5]	Cost Estimation	No	No	Estimated card, cost	Easy
Transformer	QueryFormer (Ours)	All	Yes	Yes	Sample, Histogram	Easy

Previous approaches are less effective

Table 1: Summary of existing solutions to query plan representation.

Category	Paper	Task	Parent-Children Dependency	Long Path Information Flow	Database Statistics	Training Difficulty
Flattened	AVGDL [38]	View Selection	No	Yes	NA	Hard
	RTOS [36]	Join Order Selection	Yes	Yes	NA	Hard
Tree-RNN	E2E-Cost [30]	Cost, Cardinality	Yes	Yes	Sample	Hard
	Plan-Cost [19]	Cost Estimation	Yes	Yes	Estimated card, cost	Hard
					Estimated card	Easy
					Estimated card, cost	Easy
					NA	Easy
					NA	Easy
					Estimated card, cost	Easy
					Estimated card, cost	Easy
					Sample, Histogram	Easy

```
SELECT * FROM
    title t,
    movie_info mi,
    movie_companies mc,
WHERE
    t.id = mi.movie_id AND
    t.id = mc.movie_id AND
    mi.type_id = 113 AND
    mc.c_id = 2 AND
    t.year > 2000
```



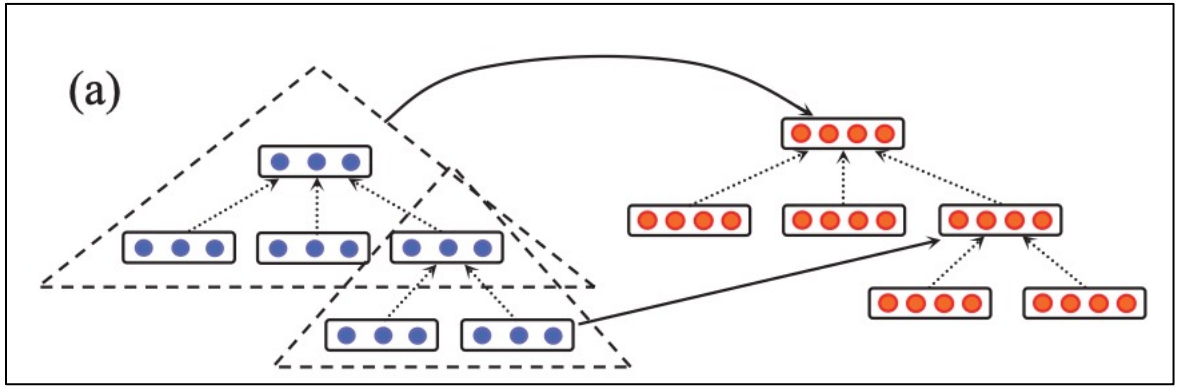
flattened: (e) (d) (b) (c) (a) *cannot capture dependency*

Previous approaches are less effective

Table 1: Summary of existing solutions to query plan representation.

Category	Paper	Task	Parent-Children Dependency	Long Path Information Flow	Database Statistics	Training Difficulty
Flattened	AVGDL [38]	View Selection	No	Yes	NA	Hard
Tree-RNN	RTOS [36]	Join Order Selection	Yes	Yes	NA	Hard
	E2E-Cost [30]	Cost, Cardinality	Yes	Yes	Sample	Hard
	Plan-Cost [19]	Cost Estimation	Yes	Yes	Estimated card, cost	Hard
Tree-CNN	NEO [17]	Optimization	Yes	No	Estimated card	Easy
	BAO [16]	Optimization	Yes	No	Estimated card, cost	Easy
	Prestroid [39]	Cost Estimation	Yes	No	NA	Easy
Feature Vectors	ReJOIN [18]	Join Order Selection	No	No	NA	Easy
	AIMeetsAI [9]	Index Selection	No	No	Estimated card, cost	Easy
	LQPP [5]	Cost Estimation	No	No	Estimated card, cost	Easy
Transformer	QueryFormer (Ours)	All	Yes	Yes	Sample, Histogram	Easy

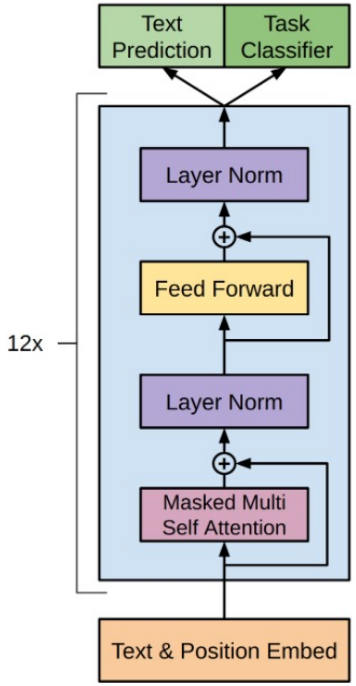
CNN

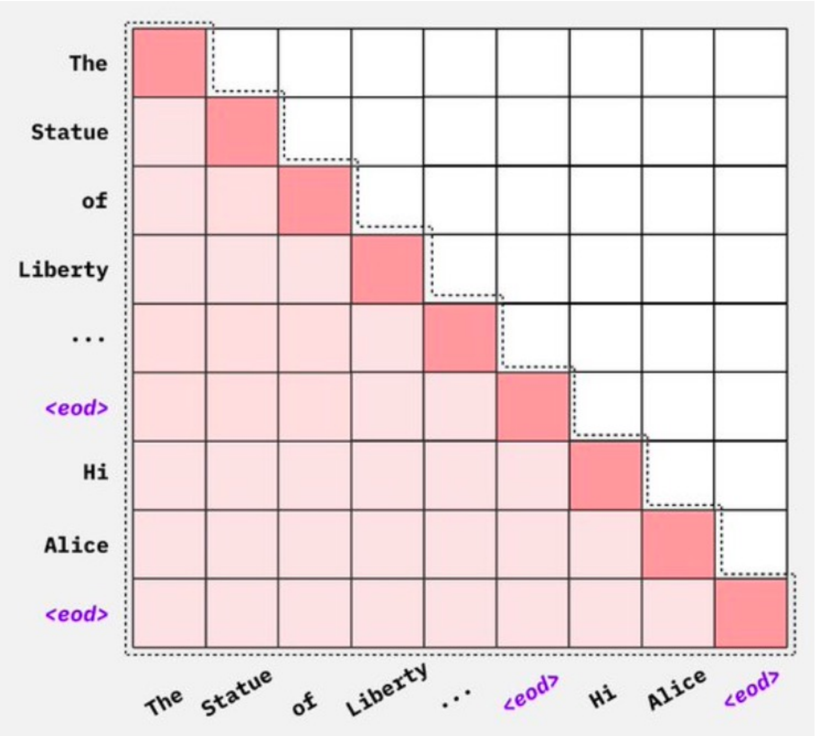


cannot capture long distance

How can we adapt Transformer / Attention?

GPT's Attention models P(new token | all previous tokens)





We can manipulate this attention via masking

QueryFormer Architecture

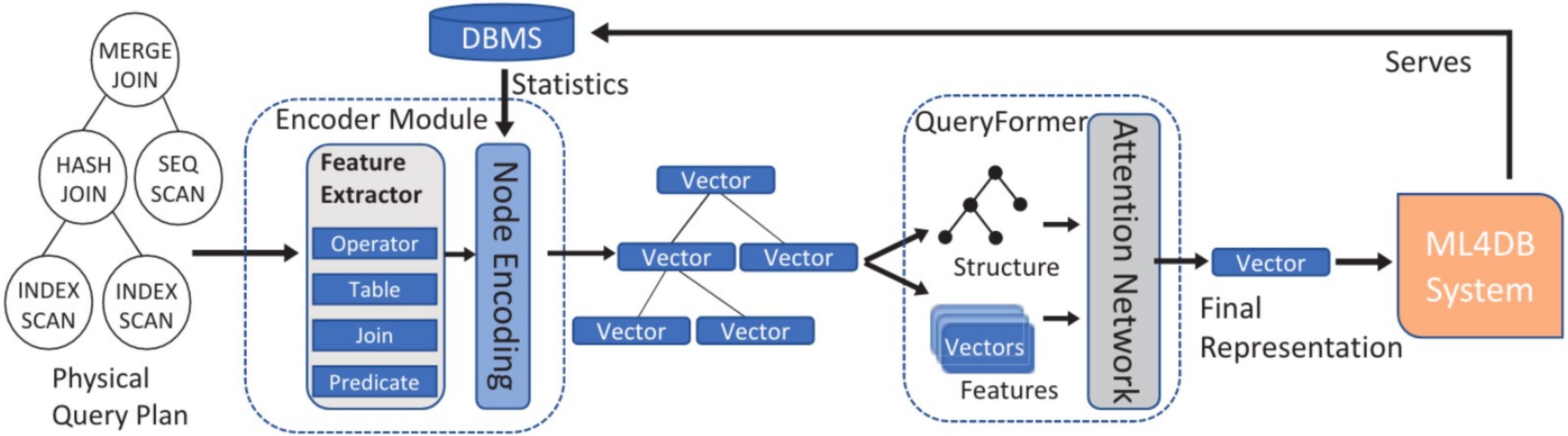
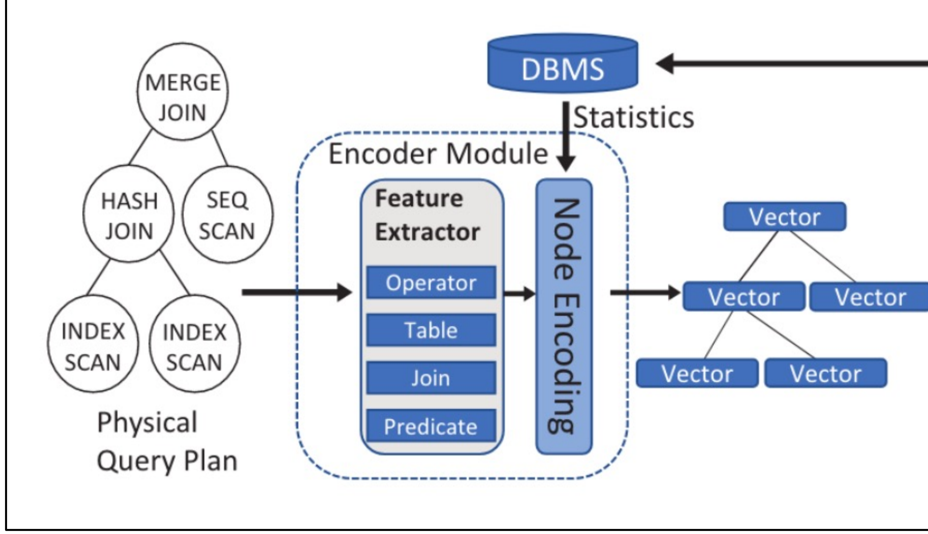


Figure 2: System overview.

Encoder: Node -> Feature Vector

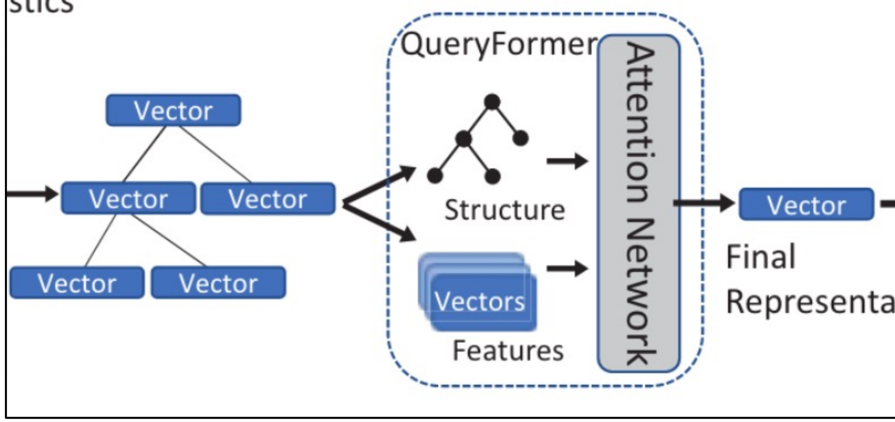


Learned embedding for

- operator: merge join, index scan
- predicate: *t.year > 2000*
- table
- join condition
- per-table statistics: histogram and samples

Similar to *learned embedding* inside the Transformer architecture

QueryFormer: Tree -> Vector

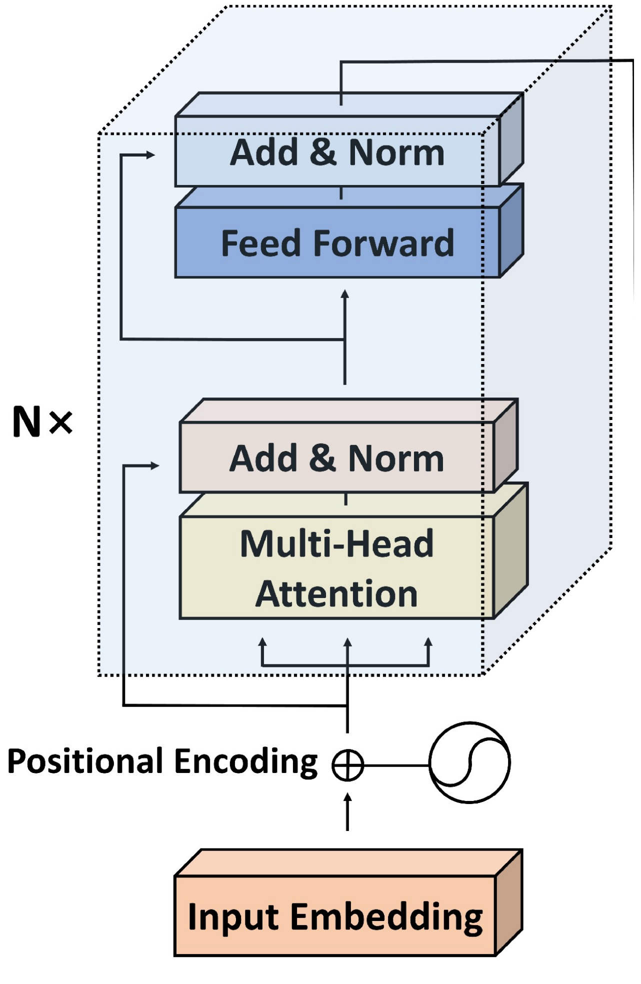


Tree-structured Transformer

- Height Encoding
- Tree-biased Attention

Aggregate nodes into a vector

Bert vs QueryFormer



BERT

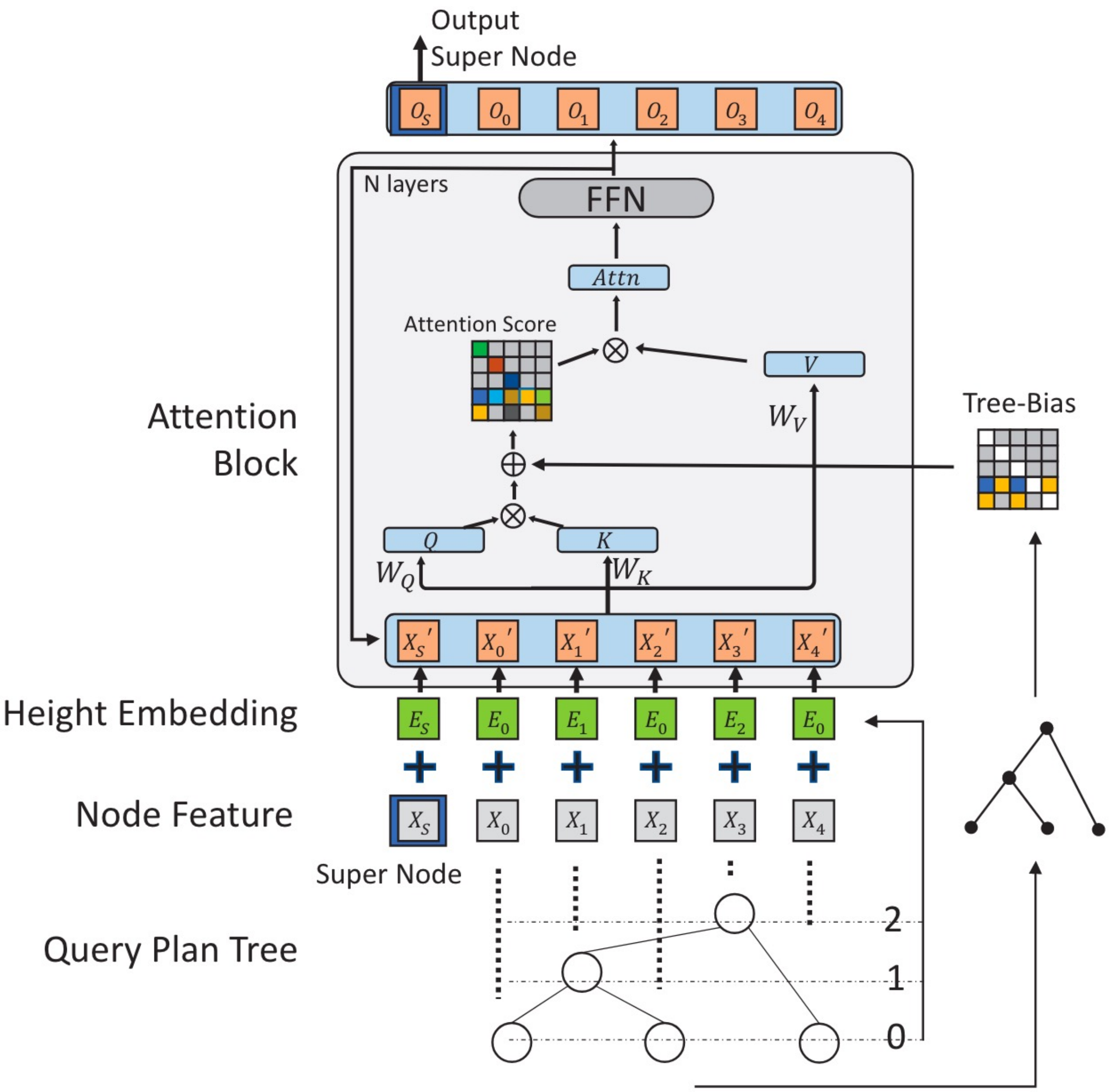
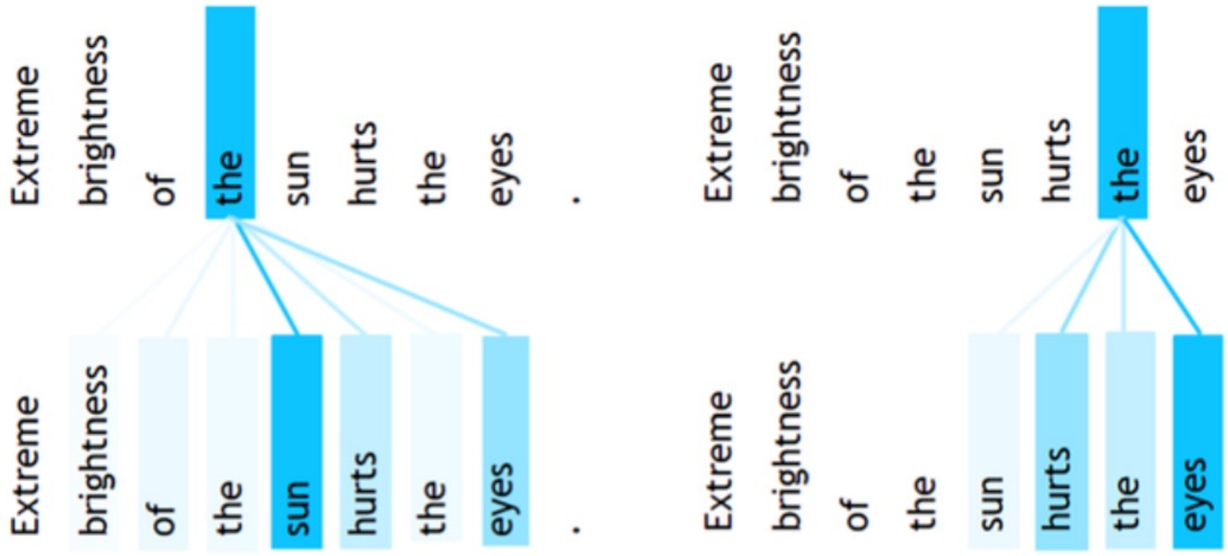
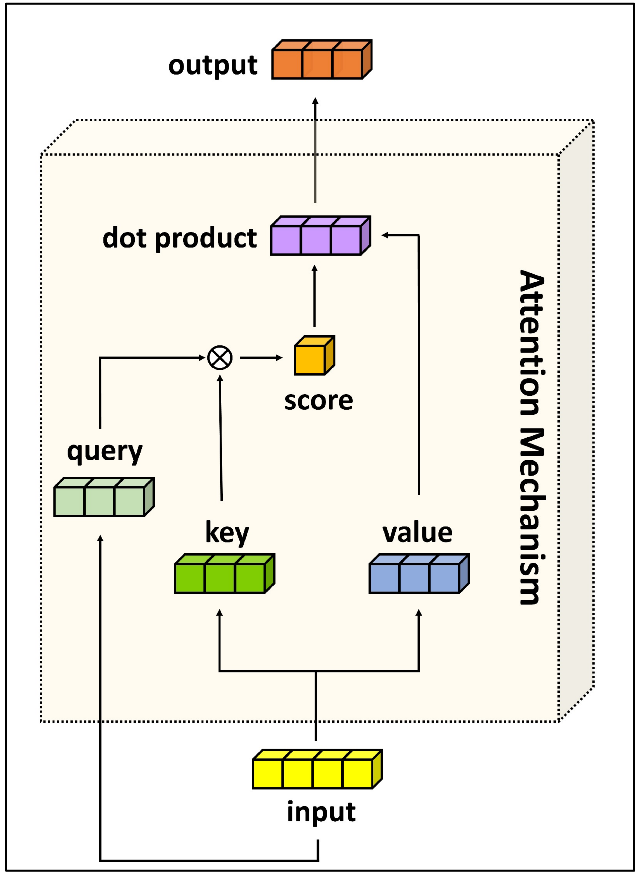


Figure 3: QueryFormer architecture.

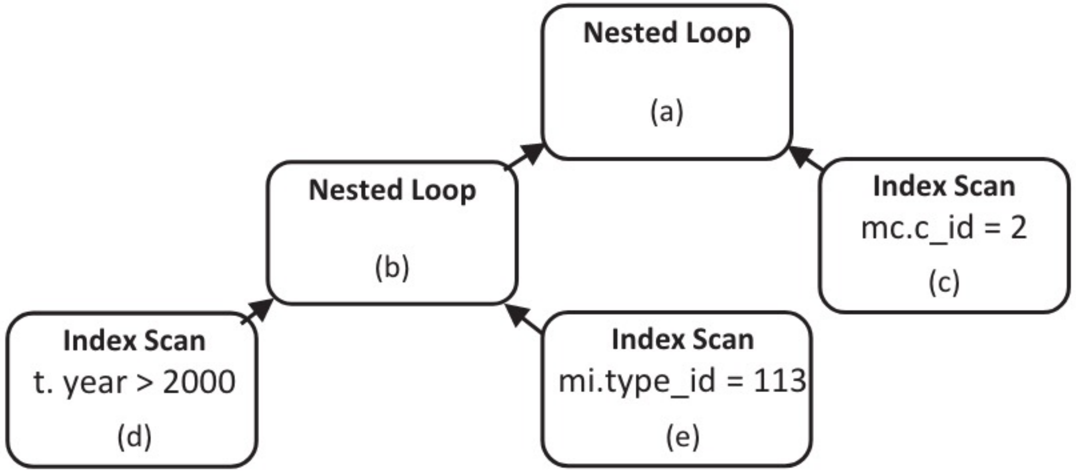
Self-Attention in Bert / Transformer



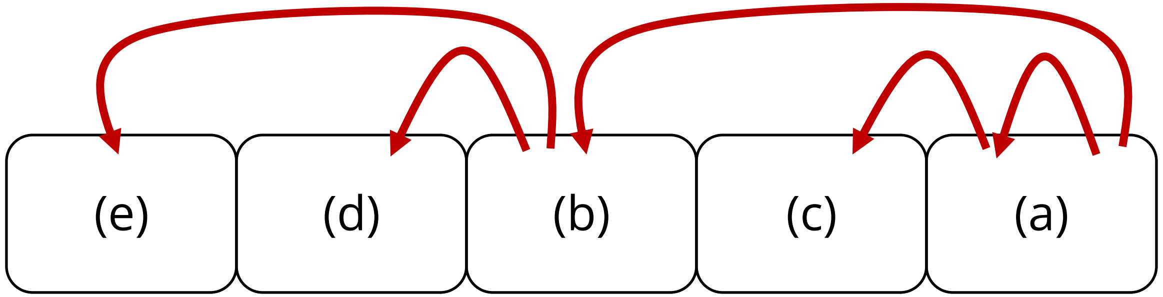


QueryFormer: Tree-biased Attention

```
SELECT * FROM
    title t,
    movie_info mi,
    movie_companies mc,
WHERE
    t.id = mi.movie_id AND
    t.id = mc.movie_id AND
    mi.type_id = 113 AND
    mc.c_id = 2 AND
    t.year > 2000
```



QueryFormer attention:



Summary

- **The QueryFormer paper** adapts Transformer to cost estimation
- ***Encoder***: An individual node -> a vector
- ***QueryFormer***: A tree of vector -> final vector   (-> cost estimation)
- *Tree-biased attention* controls information flow

Questions?