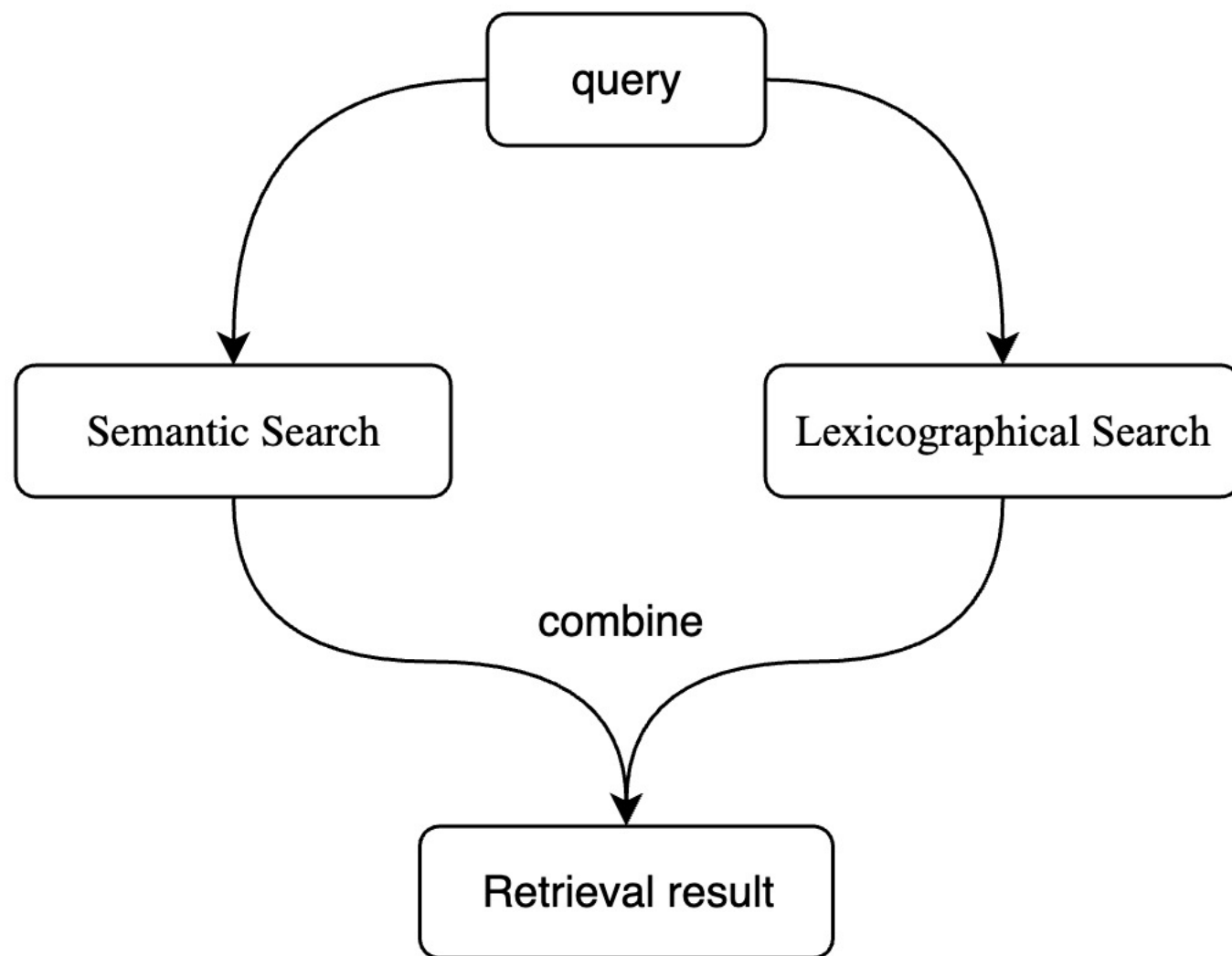


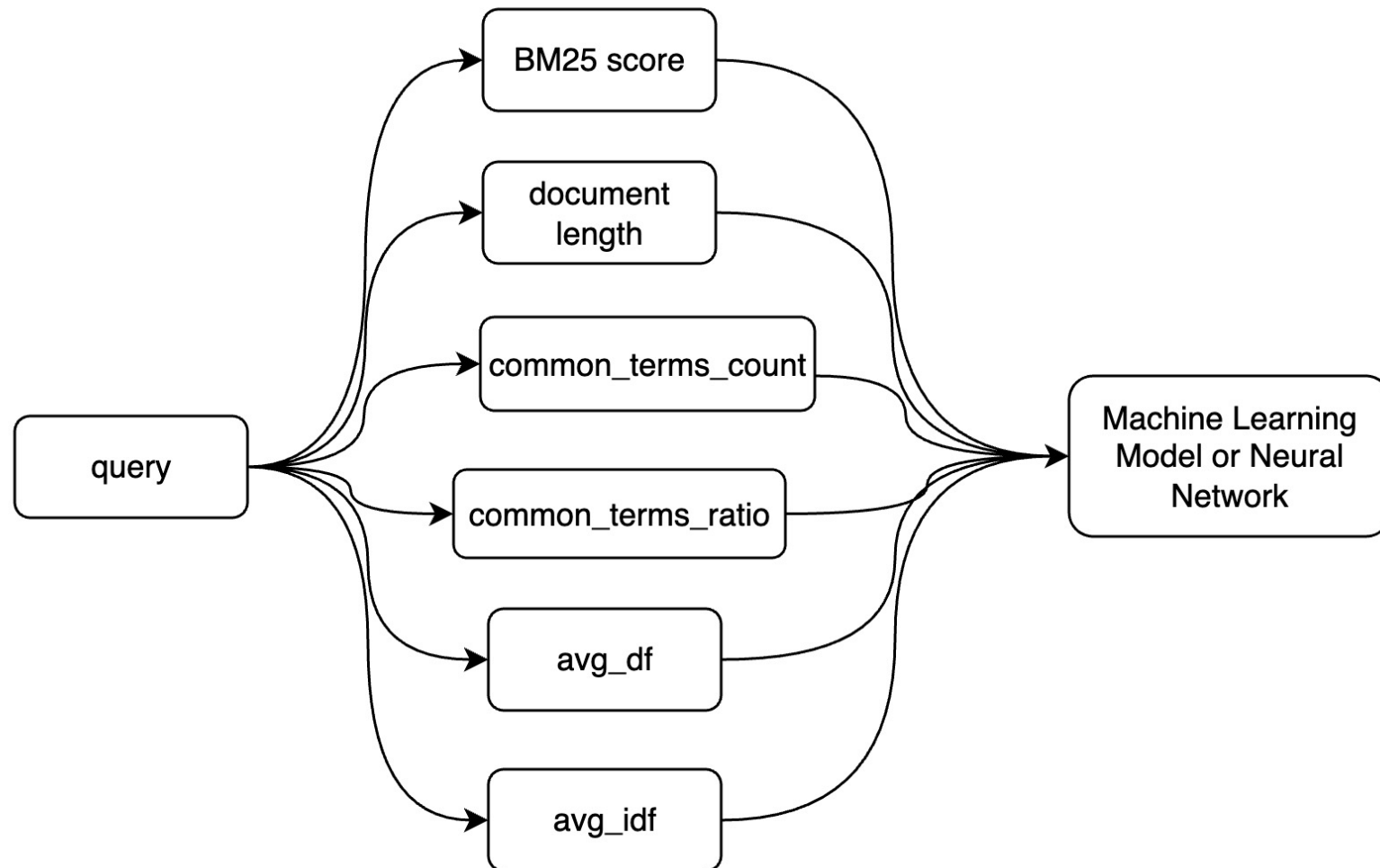
Text Retrieval



Lexicographical Search

Baseline Algorithm: BM25

BM25 combined with Machine Learning:



Lexicographical Search

Baseline Algorithm: BM25

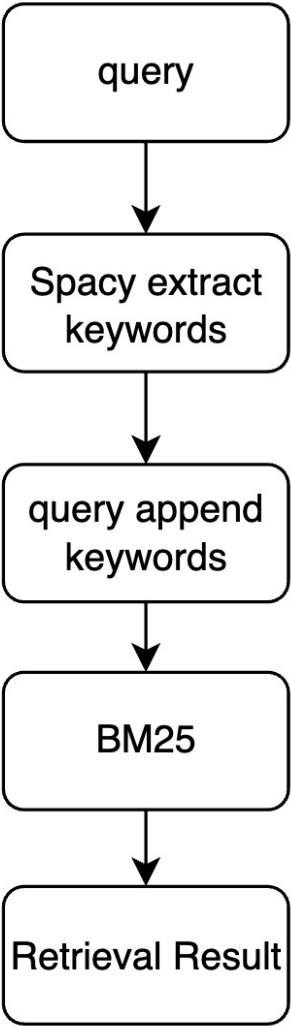
BM25 combined with Machine Learning:

| baseline | | | | |
|----------|-----------|------------|-----------|-------------------|
| BM25 | BM25 + LR | BM25 + SVM | BM25 + NN | BM25 + LambdaMART |
| 73% | 76% | 32% | 77% | 73% |
| 64% | 64% | null | 62% | 55.33% |
| 69% | 71% | null | 69% | 70% |

Top5 Recall

Lexicographical Search

Weighted BM25:



$$\text{Score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \left(\frac{\text{TF}(q_i, D) \cdot (k_1 + 1)}{\text{TF}(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{\text{avgdl}})} \right) \cdot w_i$$

From abstracts published in **2022**, in the context of **cancer classification**, what did the proposed hybrid approach consisting of **ANFIS**, **FCM**, and **SA** algorithm demonstrate?

['did', 'hybrid', 'consisting', 'fcm', 'approach', 'anfis', 'classification', 'cancer', 'context', 'proposed', 'demonstrate', 'abstracts', 'sa', '2022', 'algorithm', '2022', '2022', '2022', '2022', '2022', '2022', 'did', 'hybrid', 'consisting', 'fcm', 'approach', 'anfis', 'classification', 'cancer', 'context', 'proposed', 'demonstrate', 'abstracts', 'sa', '2022', 'algorithm', '2022', '2022', '2022', '2022', '2022', '2022', 'did', 'hybrid', 'consisting', 'fcm', 'approach', 'anfis', 'classification', 'cancer', 'context', 'proposed', 'demonstrate', 'abstracts', 'sa', '2022', 'algorithm']

| opensearch | Weighted BM25 |
|------------|---------------|
| 73% | 75% |
| 75% | 78% |
| 70% | 73% |

Semantic Search

Candidate improvement methods:

Negative sampling strategy

Cross-tower interaction Attention!

