Hierarchical Retrieval

Level 1: Documents
Search 10,000 documents → Top 100
~50ms

↓ **Coarse Filter**

Level 2: Sections
Search 500 sections → Top 50
~30ms

↓ **Fine Filter**

Level 3: Sentences
Search 200 sentences → Top 10
~20ms
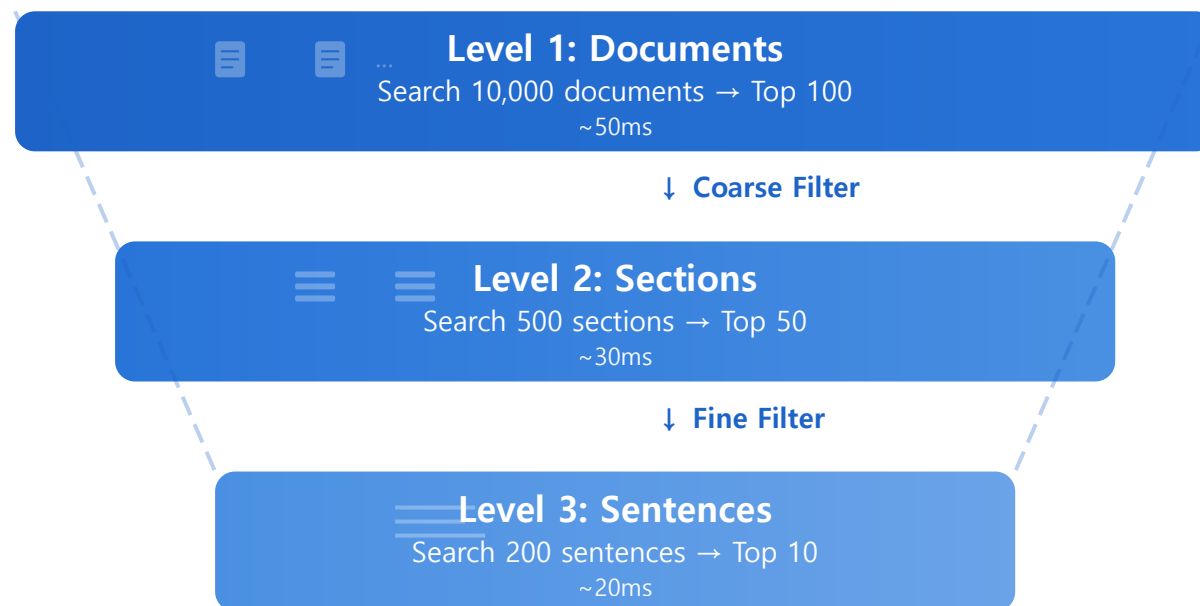
⚡ **Efficiency**

100ms total vs 500ms flat

5x faster

🎯 **Accuracy**

Precision@10: 94%

2% drop acceptable

💾 **Memory**

Incremental loading

Lower RAM usage