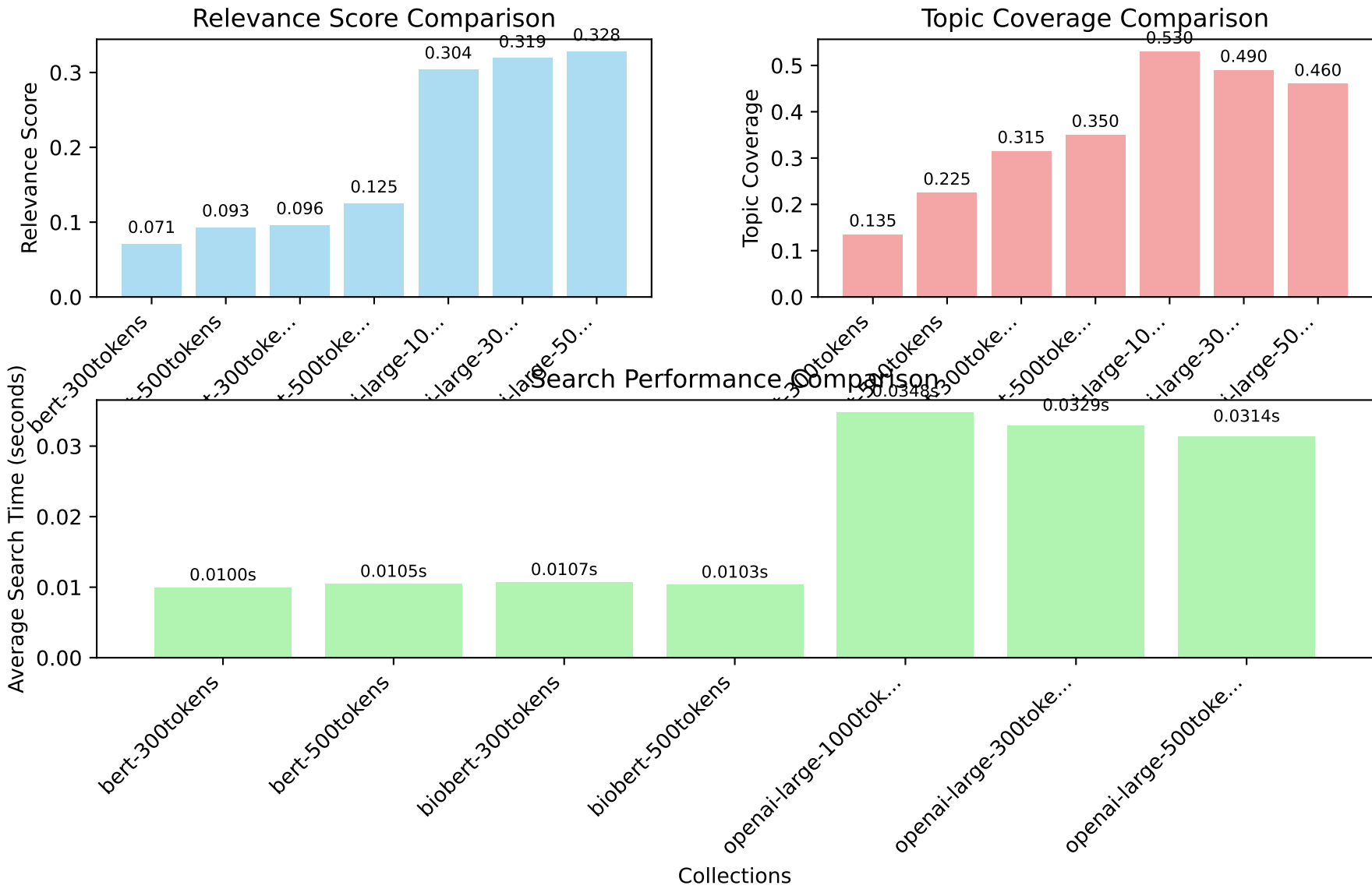# Embedding Method Comparison Report
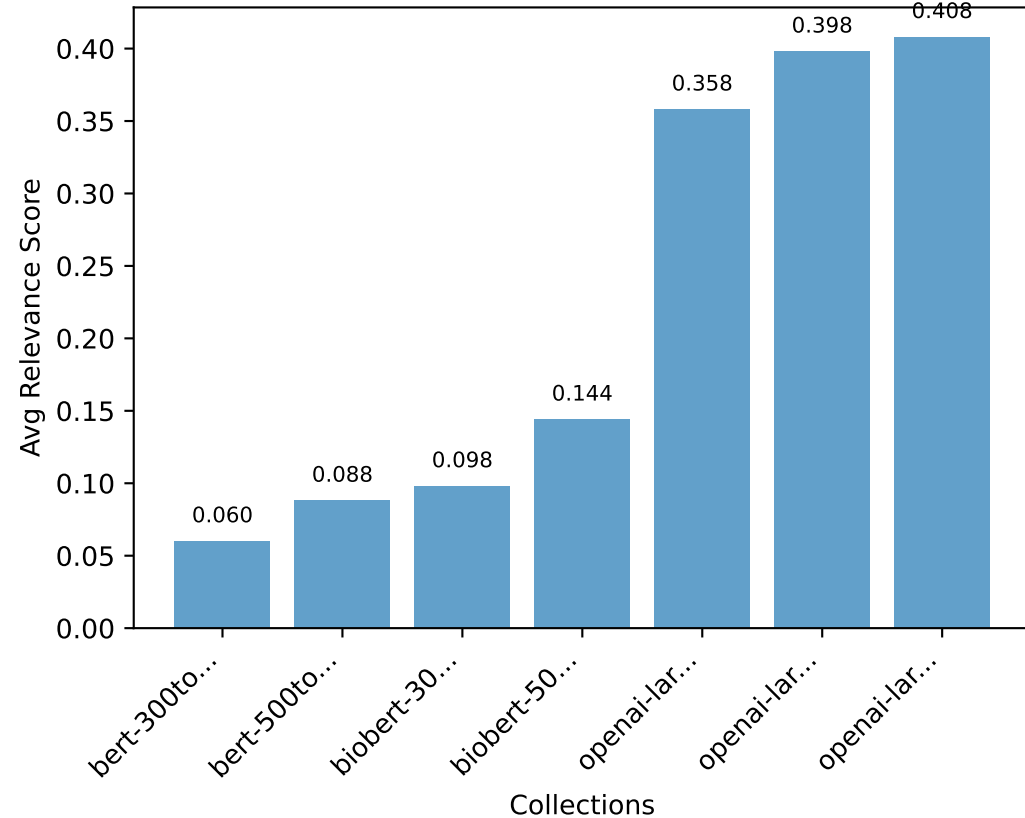
Performance Summary by Collection

	Search Time (s)	Relevance Score	Topic Coverage	Similarity Score
bert-300tokens	0.01	0.071	0.135	0.7906
bert-500tokens	0.0105	0.093	0.225	0.7813
biobert-300tokens	0.0107	0.096	0.315	0.8675
biobert-500tokens	0.0103	0.125	0.35	0.8548
openai-large-1000tokens	0.0348	0.304	0.53	0.5986
openai-large-300tokens	0.0329	0.319	0.49	0.6509
openai-large-500tokens	0.0314	0.328	0.46	0.6327

Performance by Query Type

Factual_Comprehension Queries

(bar chart; x-axis: Collections, y-axis: Avg Relevance Score)

Collection	Avg Relevance Score
bert-300to...	0.060
bert-500to...	0.088
biobert-30...	0.098
biobert-50...	0.144
openai-lar...	0.358
openai-lar...	0.398
openai-lar...	0.408

Mechanistic Queries

(bar chart; x-axis: Collections, y-axis: Avg Relevance Score)

Collection	Avg Relevance Score
bert-300to...	0.082
bert-500to...	0.098
biobert-30...	0.094
biobert-50...	0.106
openai-lar...	0.250
openai-lar...	0.240
openai-lar...	0.248
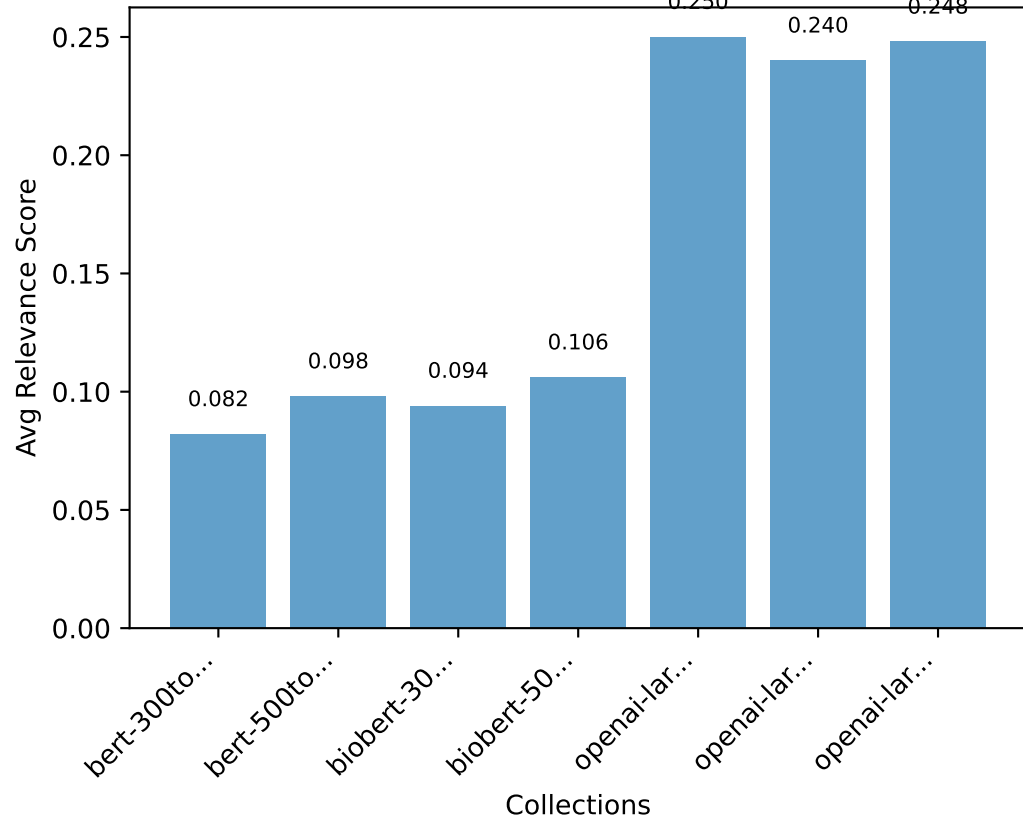
Embedding Method Comparison Report
Generated: 2025-06-01 12:10:35

Collections Compared:
- bert-500tokens
- bert-300tokens
- biobert-500tokens
- biobert-300tokens
- openai-large-1000tokens
- openai-large-500tokens
- openai-large-300tokens

Overall Performance Summary:

collection	search_time mean	std	avg_similarity_score mean	std	relevance_score mean	std	topic_coverage mean	std
bert-300tokens	0.0100	0.0006	0.7906	0.0179	0.071	0.1381	0.135	0.2334
bert-500tokens	0.0105	0.0006	0.7813	0.0184	0.093	0.1306	0.225	0.2348
biobert-300tokens	0.0107	0.0008	0.8675	0.0080	0.096	0.0971	0.315	0.3019
biobert-500tokens	0.0103	0.0010	0.8548	0.0093	0.125	0.1011	0.350	0.3055
openai-large-1000tokens	0.0348	0.0047	0.5986	0.1100	0.304	0.2718	0.530	0.3924
openai-large-300tokens	0.0329	0.0033	0.6509	0.1262	0.319	0.2960	0.490	0.3872
openai-large-500tokens	0.0314	0.0048	0.6327	0.1168	0.328	0.3307	0.460	0.3978

Key Findings:
- **Fastest Search**: bert-300tokens
- **Highest Relevance**: openai-large-500tokens
- **Best Topic Coverage**: openai-large-1000tokens
- **Highest Similarity Scores**: biobert-300tokens

Performance by Query Type:

query_type	collection	relevance_score	topic_coverage	search_time
factual_comprehension	bert-300tokens	0.060	0.10	0.0097
	bert-500tokens	0.088	0.24	0.0108
	biobert-300tokens	0.098	0.36	0.0106
	biobert-500tokens	0.144	0.39	0.0107
	openai-large-1000tokens	0.358	0.61	0.0373
	openai-large-300tokens	0.398	0.66	0.0330
	openai-large-500tokens	0.408	0.56	0.0352
mechanistic	bert-300tokens	0.082	0.17	0.0103
	bert-500tokens	0.098	0.21	0.0101
	biobert-300tokens	0.094	0.27	0.0108
	biobert-500tokens	0.106	0.31	0.0100
	openai-large-1000tokens	0.250	0.45	0.0323
	openai-large-300tokens	0.240	0.32	0.0328
	openai-large-500tokens	0.248	0.36	0.0276

Recommendations:
Based on the analysis, consider the following:
1. For speed-critical applications: Use the fastest performing model
2. For accuracy-critical applications: Use the highest relevance scoring model
3. For domain-specific queries: Consider how well each model handles medical terminology

Files Generated:
- PDF Report: experiments/results/embedding_comparison_report_20250601_121034.pdf

Total Results: 70 comparisons across 10 queries