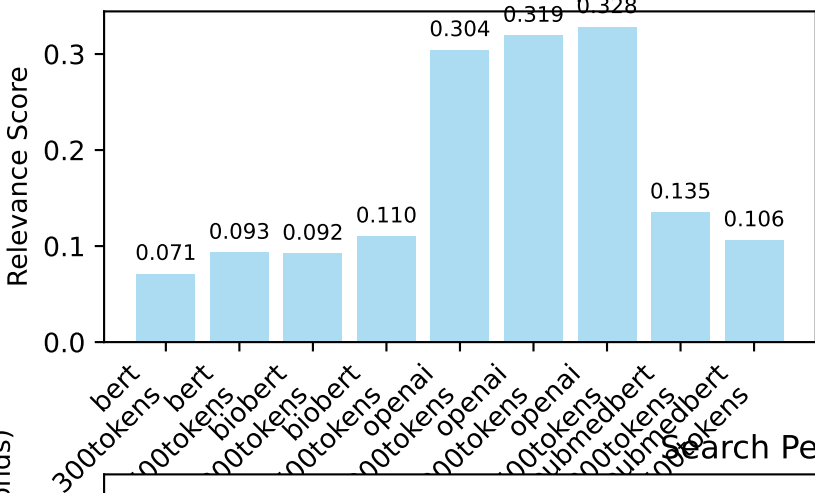# Embedding Method Comparison Report
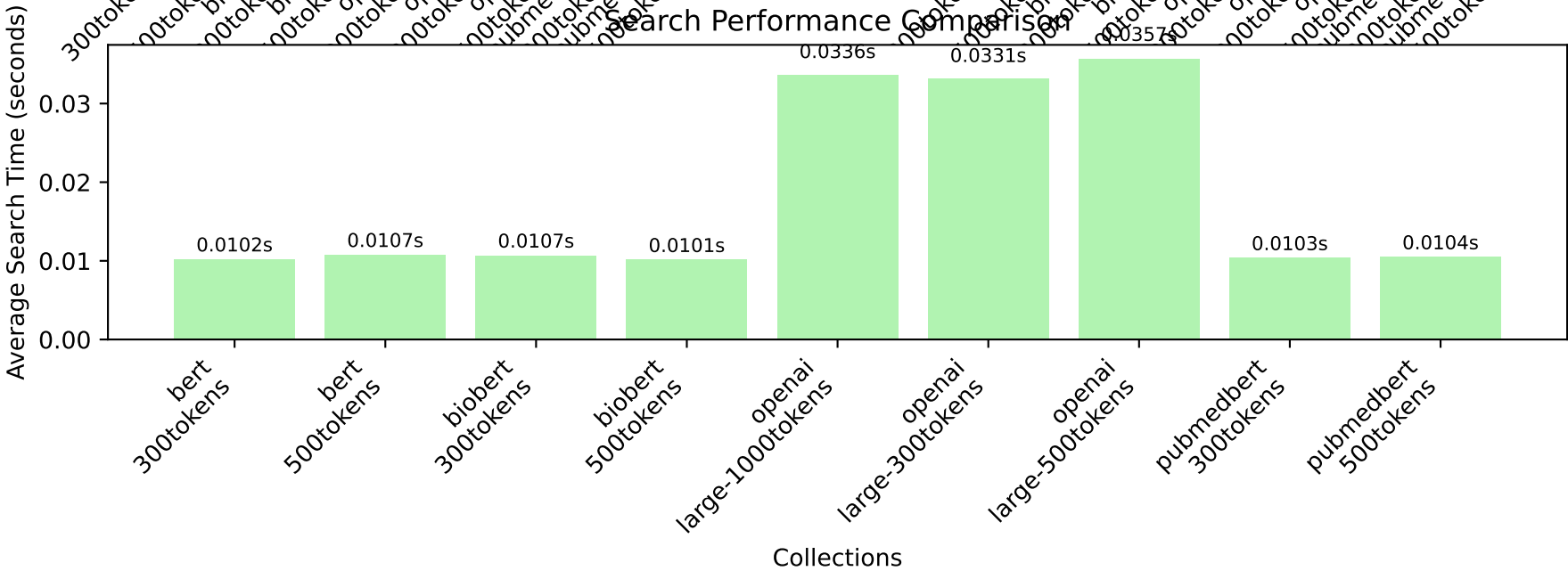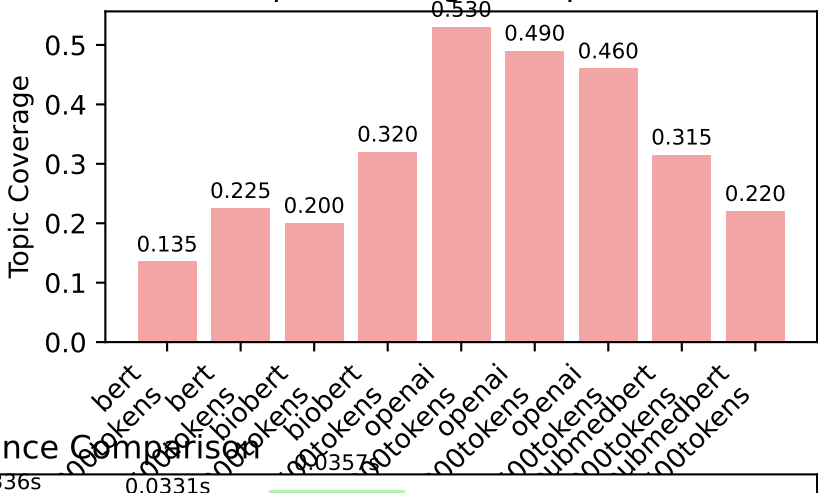
Performance Summary by Collection

	Search Time (s)	Similarity Score	Relevance Score	Topic Coverage
bert-300tokens	0.0102	0.7906	0.071	0.135
bert-500tokens	0.0107	0.7813	0.093	0.225
biobert-300tokens	0.0107	0.9113	0.092	0.2
biobert-500tokens	0.0101	0.9037	0.11	0.32
openai-large-1000tokens	0.0336	0.5986	0.304	0.53
openai-large-300tokens	0.0331	0.6509	0.319	0.49
openai-large-500tokens	0.0357	0.6328	0.328	0.46
pubmedbert-300tokens	0.0103	0.9079	0.135	0.315
pubmedbert-500tokens	0.0104	0.8993	0.106	0.22

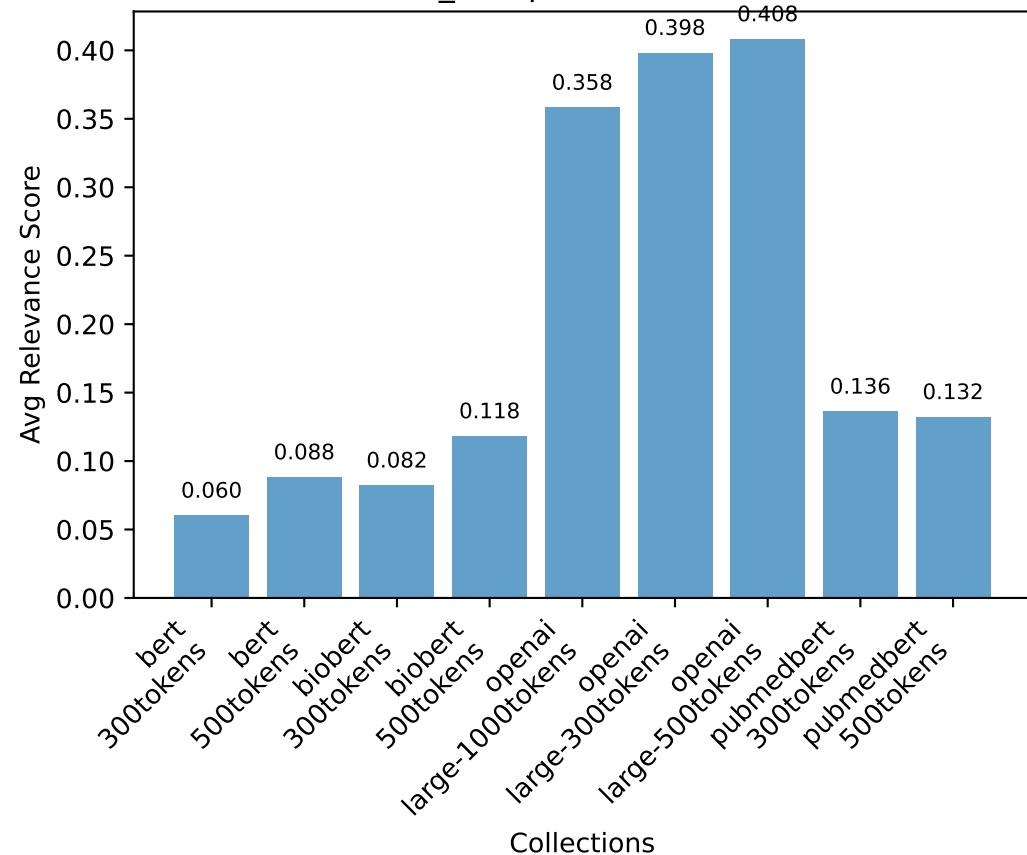Relevance Score Comparison

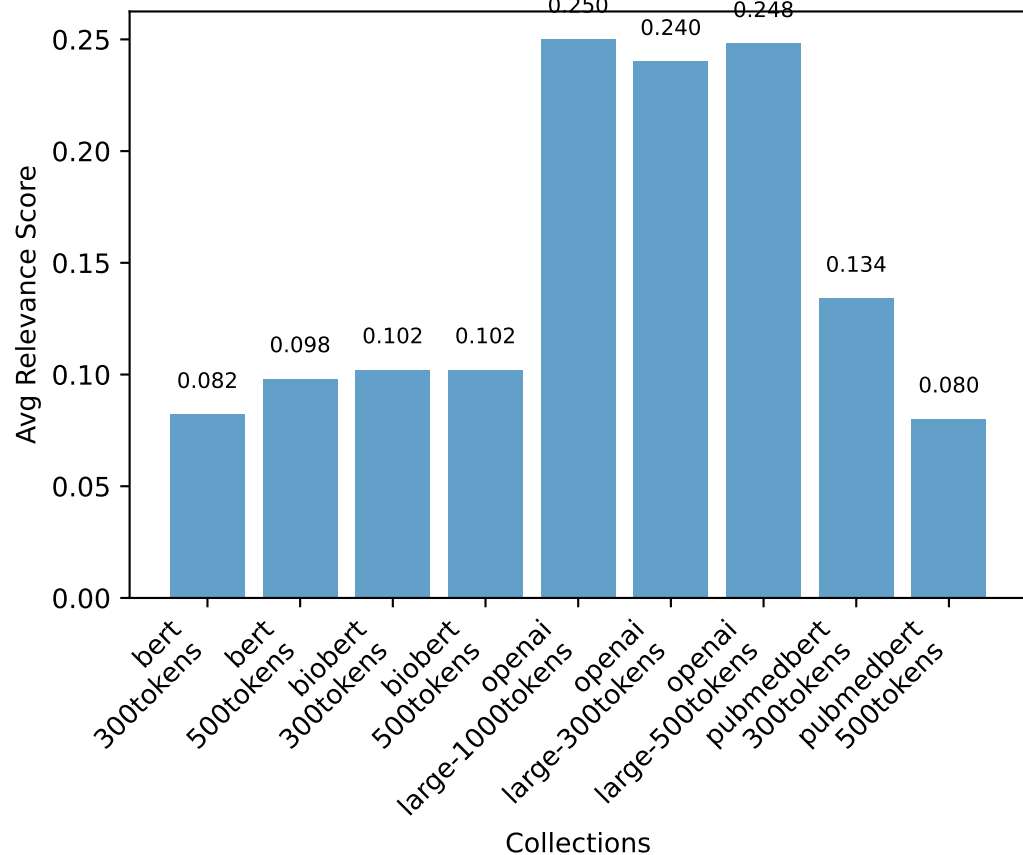Topic Coverage Comparison

Search Performance Comparison

Performance by Query Type

Factual_Comprehension Queries

Mechanistic Queries

Embedding Method Comparison Report
Generated: 2025-06-01 22:00:26

Collections Compared:
- bert-500tokens
- bert-300tokens
- pubmedbert-500tokens
- pubmedbert-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	0.0102	0.0009	0.7906	0.0179	0.071	0.1381	0.135	0.2334
bert-500tokens	0.0107	0.0010	0.7813	0.0184	0.093	0.1306	0.225	0.2348
biobert-300tokens	0.0107	0.0015	0.9113	0.0111	0.092	0.1227	0.200	0.2698
biobert-500tokens	0.0101	0.0006	0.9037	0.0113	0.110	0.1245	0.320	0.3450
openai-large-1000tokens	0.0336	0.0027	0.5986	0.1100	0.304	0.2718	0.530	0.3924
openai-large-300tokens	0.0331	0.0038	0.6509	0.1262	0.319	0.2960	0.490	0.3872
openai-large-500tokens	0.0357	0.0041	0.6328	0.1168	0.328	0.3307	0.460	0.3978
pubmedbert-300tokens	0.0103	0.0016	0.9079	0.0092	0.135	0.1445	0.315	0.3019
pubmedbert-500tokens	0.0104	0.0008	0.8993	0.0091	0.106	0.1484	0.220	0.2751

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

Performance by Query Type:

query_type	collection	search_time	avg_similarity_score	relevance_score	topic_coverage
factual_comprehension	bert-300tokens	0.0098	0.7992	0.060	0.10
	bert-500tokens	0.0108	0.7893	0.088	0.24
	biobert-300tokens	0.0109	0.9190	0.082	0.19
	biobert-500tokens	0.0104	0.9123	0.118	0.39
	openai-large-1000tokens	0.0325	0.5841	0.358	0.61
	openai-large-300tokens	0.0328	0.6189	0.398	0.66
	openai-large-500tokens	0.0340	0.6051	0.408	0.56
	pubmedbert-300tokens	0.0107	0.9046	0.136	0.31
	pubmedbert-500tokens	0.0105	0.8956	0.132	0.24
mechanistic	bert-300tokens	0.0106	0.7819	0.082	0.17
	bert-500tokens	0.0107	0.7733	0.098	0.21
	biobert-300tokens	0.0104	0.9036	0.102	0.21
	biobert-500tokens	0.0098	0.8951	0.102	0.25
	openai-large-1000tokens	0.0347	0.6130	0.250	0.45
	openai-large-300tokens	0.0334	0.6829	0.240	0.32
	openai-large-500tokens	0.0374	0.6606	0.248	0.36
	pubmedbert-300tokens	0.0100	0.9112	0.134	0.32
	pubmedbert-500tokens	0.0104	0.9031	0.080	0.20

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_220025.pdf

Total Results: 90 comparisons across 10 queries