**벡터 데이터베이스 성능 평가 보고서**

**개요**

이 보고서는 **Pinecone**, **ChromaDB**, **Faiss** 세 가지 벡터 데이터베이스의 성능을 검색된 문서와 거리, 그리고 검색 속도를 기준으로 비교한 결과입니다. 각 데이터베이스는 5개의 문서에 대한 검색 결과를 제공하였으며, 성능 차이가 명확하게 나타났습니다.

**1. 검색 정확도 (거리)**

거리(metric)는 검색된 문서가 쿼리와 얼마나 가까운지를 나타내며, 거리가 낮을수록 문서가 더 유사하다는 것을 의미합니다.

* **Pinecone**은 가장 낮은 거리를 기록하여 가장 높은 검색 정확도를 보였습니다.
  + 가장 가까운 문서인 20240103\_1105\_instructor.txt의 거리는 **0.107**로, ChromaDB (**0.214**) 및 Faiss (**0.214**)보다 훨씬 낮습니다.
  + 나머지 문서들에서도 **Pinecone**은 **0.107**에서 **0.147** 사이의 거리로 우수한 성능을 보였으며, ChromaDB와 Faiss는 각각 **0.214**에서 **0.294** 사이의 거리를 기록했습니다.

**요약**: **Pinecone**은 **ChromaDB** 및 **Faiss**에 비해 훨씬 더 정확한 검색 결과를 제공했습니다.

**2. 검색 속도**

검색 속도는 각 데이터베이스가 검색 결과를 반환하는 데 걸리는 시간을 의미합니다.

* **Faiss**는 **0.001초**로 가장 빠른 검색 속도를 기록했으며, **ChromaDB**도 **0.004초**로 매우 빠릅니다.
* 반면, **Pinecone**은 **0.2278초**로 두 데이터베이스에 비해 상당히 느렸습니다.

**요약**: **Faiss**와 **ChromaDB**는 거의 즉각적인 검색 속도를 제공한 반면, **Pinecone**은 상대적으로 느렸습니다.

**3. 인덱스 통계**

* **Pinecone**: 총 20개의 벡터가 있으며, sample\_docs에는 16개의 벡터, ns1에는 4개의 벡터가 포함되어 있습니다.
* **ChromaDB**와 **Faiss**는 동일하게 16개의 벡터로 구성되어 있습니다.

**요약**: **Pinecone**의 벡터 수는 다른 두 데이터베이스보다 많아, 이는 검색 시간이 길어진 원인일 수 있습니다.

**4. 종합 성능**

* **정확도**: **Pinecone**은 가장 높은 정확도를 제공하며, 거리가 가장 짧았습니다.
* **속도**: **Faiss**는 가장 빠른 검색 속도를 기록했으며, **ChromaDB**도 빠른 속도를 제공했습니다. 반면, **Pinecone**은 느린 속도를 기록했습니다.

상황에 따라, 정확성이 중요한 경우 **Pinecone**이 적합하며, 반대로 실시간 검색 속도가 중요한 경우 **Faiss**나 **ChromaDB**가 더 나은 선택이 될 수 있습니다.

**결론**

정확성이 중요한 작업에서는 **Pinecone**이 최적의 선택이 될 수 있지만, 속도가 중요한 경우에는 **Faiss**와 **ChromaDB**가 더 적합한 옵션입니다.