清理entityTable中的重复项：

DELETE FROM entityTable

WHERE ROWID NOT IN (

SELECT MIN(ROWID)

FROM entityTable

GROUP BY document\_id, entity\_id

);

合并摘要嵌入表与实体聚合嵌入表：

UPDATE summaries\_embeddings

SET aggregated\_entity\_embeddings = (

SELECT aggregated\_embedding

FROM aggregated\_entity\_embeddings

WHERE summaries\_embeddings.文件名 = aggregated\_entity\_embeddings.document\_title

)

WHERE 文件名 IN (

SELECT document\_title

FROM aggregated\_entity\_embeddings

);

合并数据表：

CREATE TABLE similarity\_result AS

SELECT

a.DocumentTitle1,

a.DocumentTitle2,

a.摘要\_embedding余弦相似度 AS cosine\_similarity\_summary,

a.aggregated\_embedding余弦相似度 AS cosine\_similarity\_aggregated,

a.聚合后的余弦相似度 AS cosine\_similarity\_combined,

b.CosineSimilarity AS description\_cosine\_similarity,

c.CoOccurrenceCount AS document\_co\_occurrence,

d.JaccardSimilarity AS jaccard\_similarity

FROM

cosine\_similarity\_results a

JOIN

description\_cosine\_similarity\_results b

ON

a.DocumentTitle1 = b.DocumentTitle1 AND a.DocumentTitle2 = b.DocumentTitle2

JOIN

document\_title\_co\_occurrence c

ON

a.DocumentTitle1 = c.DocumentTitle1 AND a.DocumentTitle2 = c.DocumentTitle2

JOIN

jaccard\_similarity\_results d

ON

a.DocumentTitle1 = d.DocumentTitle1 AND a.DocumentTitle2 = d.DocumentTitle2;