Knowledge Base Deduplication Analysis
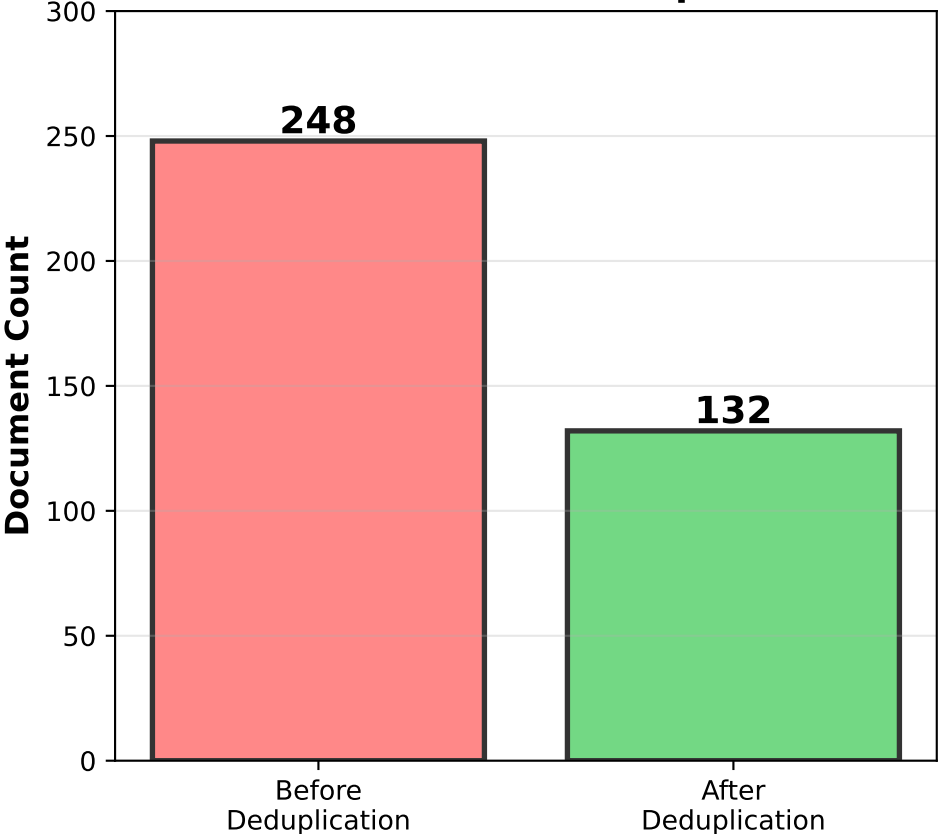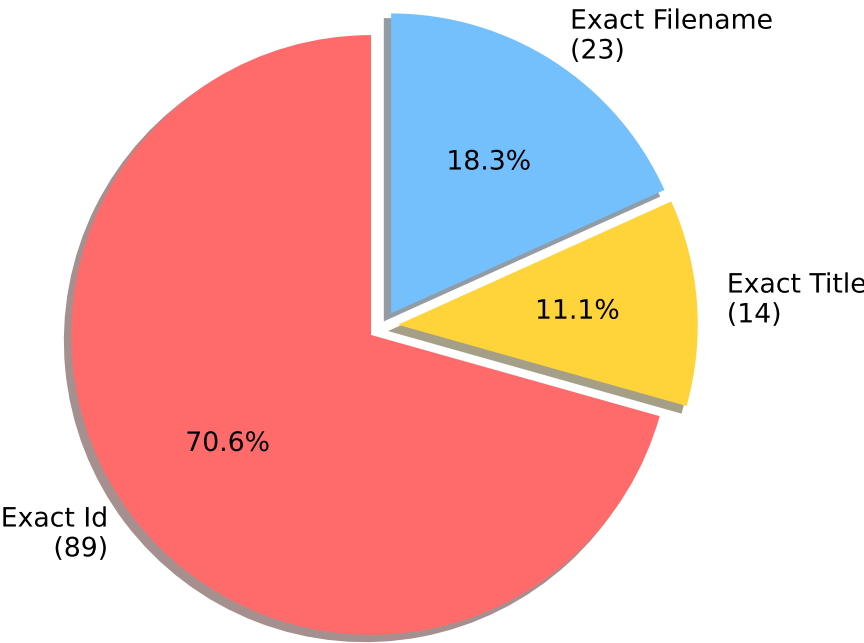

Document Count Comparison
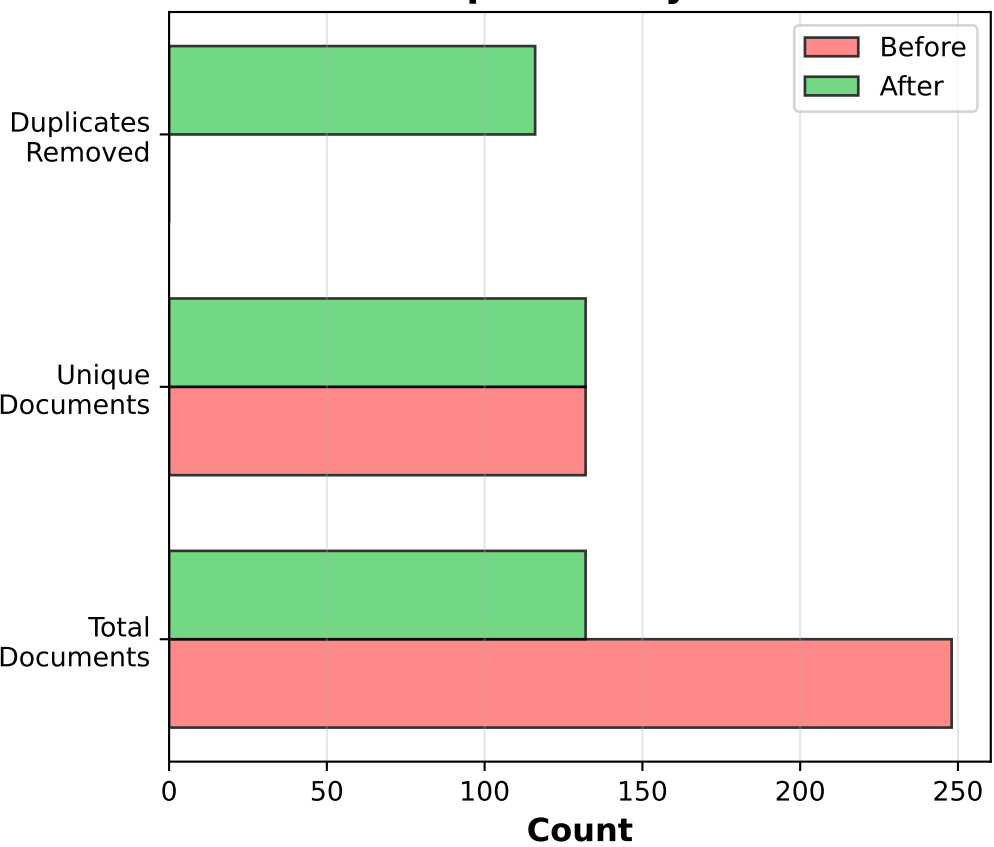- Before Deduplication: 248
- After Deduplication: 132

Documents by Source
- TypeScript KB: 16
- JSON KB: 116
- Embeddings: 116

Duplicates by Type
- Exact Id (89): 70.6%
- Exact Filename (23): 18.3%
- Exact Title (14): 11.1%

Impact Analysis
- Duplicates Removed
- Unique Documents
- Total Documents
- Legend: Before, After
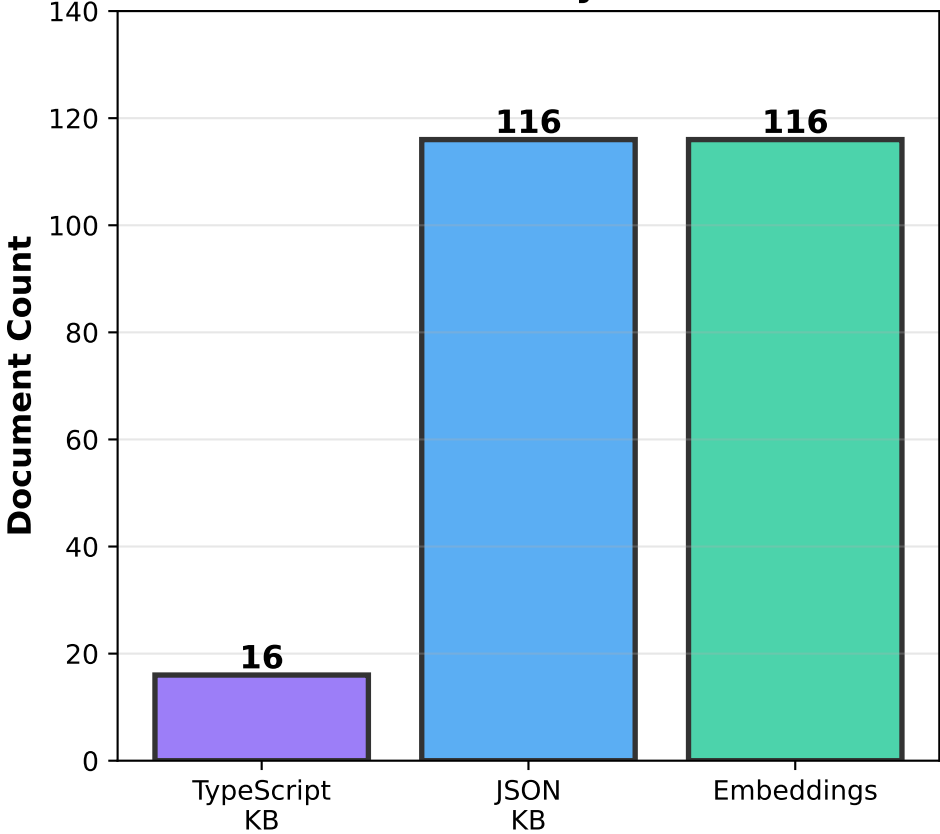
Reduction Rate: 46.8%
- 46.8% Reduced
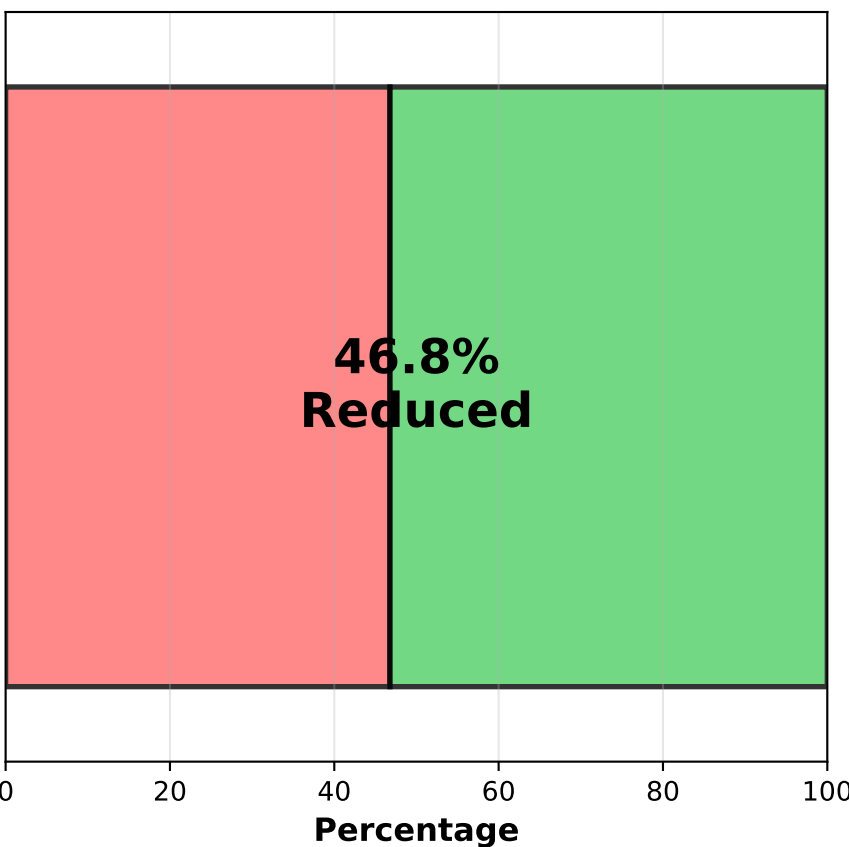
DEDUPLICATION SUMMARY

```
Original Count:     248 documents
Final Count:        132 documents
Duplicates Removed: 116 documents

Reduction Rate:     46.8%

DUPLICATE BREAKDOWN:
• Exact ID:          89 duplicates
• Exact Filename:    23 duplicates
• Exact Title:       14 duplicates

DATA SOURCES:
• TypeScript KB:     16 documents
• JSON KB:          116 documents
• Embeddings:       116 documents

DUPLICATE GROUPS:   118 groups

Status: ✓ COMPLETED SUCCESSFULLY
```