

## Metadata-Driven Hybrid RAG

### מטרה כללית

לניהול מסמכים הכוללים מספר מסמכים הכוללים טקסט, טבלאות ותרשימים. המערכת Agent RAG פיתוח עם מטא-דאטה וריראנקר, כאשר גודל כל צ'אנק עד ~5% מגודל (Dense + Sparse) מבוססת אחזור היברידי. המסמך.

---

### הנחיות לתרגיל

#### 1) Parsing & Chunking

(חלוקה לפי מבנה טבעי (סעיפים, כותרות, טבלאות, נספחים, תרשימים).

.לתקציב צ'אנקים (עד ~5% מהמסמך) — לשמור רק את ליבת המידע Distillation

.מיקום בשורה/עמודה, SectionType, TableId/FigureId, PageNumber: בכל צ'אנק Anchors שמירת

#### 2)

חמישה שדות לפחות של metadata לכל צ'אנק

FileName

PageNumber

ChunkSummary

Keywords

CriticalEntities

IncidentType (אם רלוונטי)

IncidentDate (אם קיים)

SectionType (Summary / Timeline / Table / Figure / Analysis / Conclusion)

AmountRange (אם רלוונטי)

TableId / FigureId (אם קיים)

ClientId / CaseId

### 3) Indexing

אינדוקס מידע טבלאי: המרת טבלאות ל-CSV/Markdown:  
(TableId, PageNumber) דחוס, שמירת עוגנים  
+  
תיאור טקסטואלי קצר של תוכן הטבלה לשיפור אחזור סמנטי.  
שמירת מטא-דאטה צמודה לווקטורים לצורך פילטרים בזמן ה-retrieval.

### 4) Hybrid Retrieval

ניתוח שאילתה → חילוך מילות מפתח, ישויות, תאריכים → בניית Metadata Filters.

Dense Retrieval ( $K \approx 10$ )  $\cup$  Sparse Retrieval ( $K \approx 10$ )

סינון לפי מטא-דאטה

Reranker (למשל  $K \approx 20$ ) → צמצום ל-6–8 תוצאות לפני שליחה ל-LLM.

### 5) Multi-Document Support (תיקי עבודה)

תמיכה באחזור ממסמכים מרובים בתיק לקוח/מקרה.

שימוש ב-ClientId/CaseId במטא-דאטה לחיבור בין מסמכים.

אחזור אפשרי על כלל התיק או מסמך בודד לפי הצורך.

מבנה סוכנים (3 + Router תתי-סוכנים)

סוכן נתב (Router Agent)

מנתב את השאילתה לתת-סוכן הרלוונטי לפי סוג הבקשה:

1. Summary Agent – תמצות חלקים נבחרים.

2. Needle Agent – ספציפי Anchor/איתור פסקה/סעיף –

3. Table-QA Agent – מענה על שאלות כמותיות/טבלאיות –

### קריטריונים לדאטה

1. היקף מסמך – מינ' 10 עמודים למסמך

2. גודל צ'אנק – ממוצע 250–500 tokens; מקסימום 800 לטבלה/תרשים.

3. עקרונות חלוקה – כל צ'אנק הוא יחידה לוגית; Anchors חובה: PageNumber, SectionType, ואם יש – TableId/FigureId.

4. מבנה תוכן מומלץ – Textbook (כותרות ותתי-כותרות), Timeline (חותמות זמן), Data (מדידות עם יחידות), Figures (גרפים עם תיאור טקסטואלי).

5. מטא-דאטה חובה – PageNumber, SectionType, IncidentDate/Type אם קיימים, Keywords/CriticalEntities, TableId/FigureId אם רלוונטי, ClientId/Caseld במסמכים מרובי תיקים.

### הערכת ביצועים

#### Evaluation

Answer Correctness – Ground Truth מלא

Context Precision  $\geq 0.75$

Context Recall  $\geq 0.70$

Faithfulness  $\geq 0.85$

Table-QA Accuracy  $\geq 0.90$