



THE INSTITUTE
Empowering Israel's Future Through AI

AI Developers

Practice

TF-IDF RAG

Biblical RAG

עליכם להכין RAG תנ"כי שמשתמש בTF-IDF עבור החיפוש
(כלומר במקום חיפוש סמנטי).

לרשותכם מאגר בפורמט jsonl המכיל מקורות שונים כמו
התנ"ך, תלמוד וכו'.

נקודות למחשבה:

- ברשותנו הטקסט בעברית ואנגלית, במה עדיף להשתמש?
- לחלקים של התנ"ך לעומת שאר המסמכים יש פורמט שונה. איך תתמודדו עם ההבדלים בפורמט?

```
"chapter": 1, "verse": 1, "text_he": "ת ברא אלהים את השמים ואת הארץ"  
"chapter": 1, "verse": 2, "text_he": "ורוח אלהים מרחפת על-פני המים"  
"chapter": 1, "verse": 3, "text_he": "ויאמר אלהים יהי אור ויהי-אור"  
"chapter": 1, "verse": 4, "text_he": "יבדל אלהים בין האור ובין החושך"  
"chapter": 1, "verse": 5, "text_he": "ילה ויהי-ערב ויהי-בקר יום אחד"  
"chapter": 1, "verse": 6, "text_he": "המים ויהי מבדיל בין מים למים"  
"chapter": 1, "verse": 7, "text_he": "ין המים אשר מעל לרקיע ויהי-כן"  
"chapter": 1, "verse": 8, "text_he": "מים ויהי-ערב ויהי-בקר יום שני"  
"chapter": 1, "verse": 9, "text_he": "-מקום אחד ותראה היבשה ויהי-כן"  
"chapter": 1, "verse": 10, "text_he": "ם קרא ימים וירא אלהים כי-טוב"  
"chapter": 1, "verse": 11, "text_he": "אשר זרעו-בו על-הארץ ויהי-כן"  
"chapter": 1, "verse": 12, "text_he": "-בו למינהו וירא אלהים כי-טוב"  
"chapter": 1, "verse": 13, "text_he": "\"ויהי-ערב ויהי-בקר יום שלישי"  
"chapter": 1, "verse": 14, "text_he": "יו לאת ולמועדים ולימים ושנים"  
"chapter": 1, "verse": 15, "text_he": "השמים להאיר על-הארץ ויהי-כן"  
"chapter": 1, "verse": 16, "text_he": "הקטן למשלת הלילה ואת הכוכבים"  
"chapter": 1, "verse": 17, "text_he": "ים ברקיע השמים להאיר על-הארץ"  
"chapter": 1, "verse": 18, "text_he": "ובין החושך וירא אלהים כי-טוב"  
"chapter": 1, "verse": 19, "text_he": "\"ויהי-ערב ויהי-בקר יום רביעי"  
"chapter": 1, "verse": 20, "text_he": "פף על-הארץ על-פני רקיע השמים"  
"chapter": 1, "verse": 21, "text_he": "כנף למינהו וירא אלהים כי-טוב"  
"chapter": 1, "verse": 22, "text_he": "את-המים בימים והעוף ירב בארץ"  
"chapter": 1, "verse": 23, "text_he": "\"ויהי-ערב ויהי-בקר יום חמישי"  
"chapter": 1, "verse": 24, "text_he": "ורמש וחיתו-ארץ למינה ויהי-כן"  
"chapter": 1, "verse": 25, "text_he": "דמה למינהו וירא אלהים כי-טוב"
```

Hybrid Search

Biblical RAG

כעת נוסיף לפרויקט שימוש בחיפוש סמנטי על מנת להכין RAG שמתמש ב-Hybrid Search בעזרת TF-IDF עבור חיפוש lexical ו-FAISS עבור VectorDB (לחיפוש הסמנטי).

נקודות למחשבה:

- שחקו עם המשקולות של כל אחד מסוגי החיפושים ונסו לראות את ההבדלים (לתת יותר משקל לסמנטי, ואז לתת לו פחות ולהשוות)
- איך כדאי ליצור את chunks?

```
"chapter": 1, "verse": 1, "text_he": "ת ברא אלהים את השמים ואת הארץ"  
"chapter": 1, "verse": 2, "text_he": "ורוח אלהים מרחפת על-פני המים"  
"chapter": 1, "verse": 3, "text_he": "ויאמר אלהים יהי אור ויהי-אור"  
"chapter": 1, "verse": 4, "text_he": "יבדל אלהים בין האור ובין החושך"  
"chapter": 1, "verse": 5, "text_he": "ילה ויהי-ערב ויהי-בקר יום אחד"  
"chapter": 1, "verse": 6, "text_he": "המים ויהי מבדיל בין מים למים"  
"chapter": 1, "verse": 7, "text_he": "ין המים אשר מעל לרקיע ויהי-כן"  
"chapter": 1, "verse": 8, "text_he": "מים ויהי-ערב ויהי-בקר יום שני"  
"chapter": 1, "verse": 9, "text_he": "מקום אחד ותראה היבשה ויהי-כן"  
"chapter": 1, "verse": 10, "text_he": "ם קרא ימים וירא אלהים כי-טוב"  
"chapter": 1, "verse": 11, "text_he": "אשר זרעו-בו על-הארץ ויהי-כן"  
"chapter": 1, "verse": 12, "text_he": "בו למינהו וירא אלהים כי-טוב"  
"chapter": 1, "verse": 13, "text_he": "ויהי-ערב ויהי-בקר יום שלישי"  
"chapter": 1, "verse": 14, "text_he": "יו לאת ולמועדים ולימים ושנים"  
"chapter": 1, "verse": 15, "text_he": "השמים להאיר על-הארץ ויהי-כן"  
"chapter": 1, "verse": 16, "text_he": "הקטן למשלת הלילה ואת הכוכבים"  
"chapter": 1, "verse": 17, "text_he": "ים ברקיע השמים להאיר על-הארץ"  
"chapter": 1, "verse": 18, "text_he": "ובין החושך וירא אלהים כי-טוב"  
"chapter": 1, "verse": 19, "text_he": "ויהי-ערב ויהי-בקר יום רביעי"  
"chapter": 1, "verse": 20, "text_he": "פף על-הארץ על-פני רקיע השמים"  
"chapter": 1, "verse": 21, "text_he": "כנף למינהו וירא אלהים כי-טוב"  
"chapter": 1, "verse": 22, "text_he": "את-המים בימים והעוף ירב בארץ"  
"chapter": 1, "verse": 23, "text_he": "ויהי-ערב ויהי-בקר יום חמישי"  
"chapter": 1, "verse": 24, "text_he": "ורמש וחיתו-ארץ למינה ויהי-כן"  
"chapter": 1, "verse": 25, "text_he": "דמה למינהו וירא אלהים כי-טוב"
```


Hybrid Search

Biblical RAG

סדר ביצוע מומלץ:

- מימוש RAG עם TF-IDF על התנ"ך בלבד
- הרחבה לRAG עם חיפוש סמנטי
- הוספה של שאר מקורות המידע (חשבו איך להתמודד עם ההבדל בפורמט של המידע עבור החלוקה לchunks)

```
"chapter": 1, "verse": 1, "text_he": "ת ברא אלהים את השמים ואת הארץ"  
"chapter": 1, "verse": 2, "text_he": "ורוח אלהים מרחפת על-פני המים"  
"chapter": 1, "verse": 3, "text_he": "ויאמר אלהים יהי אור ויהי-אור"  
"chapter": 1, "verse": 4, "text_he": "יבדל אלהים בין האור ובין החושך"  
"chapter": 1, "verse": 5, "text_he": "ילה ויהי-ערב ויהי-בקר יום אחד"  
"chapter": 1, "verse": 6, "text_he": "המים ויהי מבדיל בין מים למים"  
"chapter": 1, "verse": 7, "text_he": "ין המים אשר מעל לרקיע ויהי-כן"  
"chapter": 1, "verse": 8, "text_he": "מים ויהי-ערב ויהי-בקר יום שני"  
"chapter": 1, "verse": 9, "text_he": "-מקום אחד ותראה היבשה ויהי-כן"  
"chapter": 1, "verse": 10, "text_he": "ם קרא ימים וירא אלהים כי-טוב"  
"chapter": 1, "verse": 11, "text_he": "אשר זרעו-בו על-הארץ ויהי-כן"  
"chapter": 1, "verse": 12, "text_he": "-בו למינהו וירא אלהים כי-טוב"  
"chapter": 1, "verse": 13, "text_he": "\"ויהי-ערב ויהי-בקר יום שלישי"  
"chapter": 1, "verse": 14, "text_he": "יו לאת ולמועדים ולימים ושנים"  
"chapter": 1, "verse": 15, "text_he": "השמים להאיר על-הארץ ויהי-כן"  
"chapter": 1, "verse": 16, "text_he": "הקטן למשלת הלילה ואת הכוכבים"  
"chapter": 1, "verse": 17, "text_he": "ים ברקיע השמים להאיר על-הארץ"  
"chapter": 1, "verse": 18, "text_he": "ובין החושך וירא אלהים כי-טוב"  
"chapter": 1, "verse": 19, "text_he": "\"ויהי-ערב ויהי-בקר יום רביעי"  
"chapter": 1, "verse": 20, "text_he": "פף על-הארץ על-פני רקיע השמים"  
"chapter": 1, "verse": 21, "text_he": "כנף למינהו וירא אלהים כי-טוב"  
"chapter": 1, "verse": 22, "text_he": "את-המים בימים והעוף ירב בארץ"  
"chapter": 1, "verse": 23, "text_he": "\"ויהי-ערב ויהי-בקר יום חמישי"  
"chapter": 1, "verse": 24, "text_he": "ורמש וחיתו-ארץ למינה ויהי-כן"  
"chapter": 1, "verse": 25, "text_he": "דמה למינהו וירא אלהים כי-טוב"
```



THE INSTITUTE
תודה!