# Review 148: Knowledge Graph Prompting for Multi-Document Question Answering, 17.09.23

**Paper: https://arxiv.org/abs/2308.11730v3**

https://arxiv.org/abs/2308.11730  
סקירה זו נכתבה על ידי עדן יבין

אנחנו יודעים שמודלי שפה יודעים לענות על שאלות על גבי מסמכים אם בצורת zero shot או בצורת few shot אבל מה עם המקרה שבו צריך להסתכל על מסמכים שונים כדי לקבל תשובה? או להשוות בין שני מסמכים בעלי מבנה שונה (טבלה וטקסט חופשי)? ספוילר, לא טוב. היום ב-#shorthebrewpapereviews קבלו את הסקירה הבאה:

החוקרים מראים ששיטות קיימות כגון chain of thought שיכולות להתמודד עם הבעיה מביאות איתן מחיר לא קטן. הם אומרים (ולדעתי בצדק) ששיטות אלו לא יעילות כי הן צורכות המון קריאות יקרות למודל השפה. כתוצאה מכל אלו, החוקרים פיתחו שיטה חדשה לה קראו KGP המשתמשת בגרף ידע. הם בונים את גרף הידע(Knowledge Graph) שבו קודקודים מסמלים פסקאות / מבנה מסמך וקשת מסמך קשר סמנטי או שייכות מבנה (למשל מסמך A הוא \*מסוג\* טבלה). מכאן כאשר מגיעה שאלה חדשה נחלק אותה למקרים. כל מקרה צורך לוגיקה קצת שונה.

שאלה על מבנה העמוד, למשל ״על מה מדובר בעמוד 1?״ כדי לענות על מקרה זה, השיטה משתמשת במודל שפה כדי לסווג את סוג העמוד. למשל במקרה זה מדובר במסמך טקסט, כלומר אוסף של פסקאות. לאחר שנדע את סוג המסמך נוכל לבצע הליכה בגרף מקודקוד העמוד אל כל הקודקודים התואמים את סוגו (למשל כלל הפסקאות במסמך הטקסט).  
  
החוקרים מציינים שהם התייחסו לשני סוגי טקסט שיכולים להופיע בעמוד והתייחסו לכל אחד באופן שונה: – מסמך טקסט: אוסף של פסקאות, כל פסקה היא קודקוד. – טבלה: התוכן הוא הטבלה עצמה מיוצגת כ-markdown בקודקוד. צוין שניתן יהיה להרחיב את השיטה לעוד סוגים של טקסט.

מבנה על תוכן עמוד מסוים, למשל: באיזו שנה נולד יוצר הסדרה הסימפסונים? כדי לענות על שאלה זו מבוצע התהליך הבא: מאחזרים את הקודקוד עם המידע הרלוונטי ביותר לשאלה. מוסיפים את המידע כקונטקסט לשאלה ומבקשים ממודל השפה לייצר את המידע הבא שנצטרך כדי לענות על השאלה.  
  
נסתכל על השכונה של הקודקוד בו אנו נמצאים כרגע ונחפש את הקודקוד הדומה ביותר למידע שמודל השפה ייצר ״כצעד הבא״. כך נמשיך עד שיש בידינו את כל המידע כדי לענות על השאלה. נשים לב שניתן לבחור יותר מקודקוד אחד בשלב החיפוש, כלומר נוכל לאחזר מידע משני קודקודים בצעד אחד. זוכרים שבהתחלה דיברנו על בעיה נוספת הינה מספר קריאות רב למודל השפה (chain of thought)?  
  
ובכן כדי להתמודד עם הבעיה הוצע לבצע אימון נוסף (fine tune) שבו מודל שפה אומן על מעקב אחר הוראות. במקרה הזה, ההוראה הינה לחזות את הראיה הבאה הרלוונטית כדי לענות על השאלה בהינתן המידע עד כה. אני יודע שמעניין אתכם הפיל שבחדר, איך נבנה הגרף עצמו? ואובכן נחלק זאת למקרים:

סוג: באמצעות Extract-PDF מחלצים את סוג המידע במסמך ומוסיפים זאת בתור קודקוד.

טבלה: קודקוד של טבלה מכיל את סוגו ובנוסף את כל הטבלה כ-markdown.

במקרה של טקסט: ראשית נחלק אותו לפסקאות. כל פיסקה מקודדת לשיכון באמצעות מודל אמבדינג או באמצעות TF-IDF. לאחר מכן נחבר בקשת מסמכים דומים. 4. קשתות: בנוסף לקשתות בין פסקאות דומות קיימות קשתות נוספות בין קודקוד העמוד לפסקה הראשונה בו / הטבלה שהוא מכיל בקשר של ״מכיל״.

התוצאות נראות טוב ובנוסף מביאות פחות הזיות ממודל השפה. מודלי שפה נוסו מסוגים שונים: T5, Llama, MDR ודווקא T5 הביא המון פעמים לחיפוש הטוב ביותר בגרף.