# Review 131: [Short] 28.08.23: Nougat: Neural Optical Understanding for Academic Documents https://arxiv.org/abs/2308.13418.pdf

**Paper: https://arxiv.org/abs/2308.13418v1**

אתם בטח שמתם לב כי אני אוהב לקרוא מאמרים. יש מאמרים שלא פשוט להבין אותם ואז אני מתחיל לחפש מושגים שונים או איפה ובאיזה הקשר מופיעים כל מיני רפרנסים בקובץ של המאמר. זה כמובן אפשרי רק עבור מאמרים יחסית חדשים (ב 20 השנים האחרונות). במאמרים ישנים זה בעייתי כי הם פשוט מהווים צילום של המאמר.   
  
היום ב-#shorthebrewpapereviews סוקרים מאמר שלוקח מאמר מדעי שהוא לא בפורמט הנוח והופך אותו למסמך נוח לקריאה בכיף ולחפש שם כל מיני דברים. איך הם עשו זאת? מכיוון שהמסמך מגיע בתור תמונה בשלב הראשון צריך לבצע Optical Character Recognition או OCR. משימה זאת איננה פשוטה כי מאמרים לפעמים מכילים נוסחאות די מורכבות וסביר להניח ששיטות OCR קיימות די יתקשו להתמודד עם זה.   
  
המאמר מפתח מודל לזיהוי תוכן מהתמונה של מסמך בעצמו. המודל מורכב מאנקודר שמקבל את תמונת המאמר, עושה לו עיבוד מקדים (מוריד שוליים, הופך לאותו גודל וכדומה). לאחר מכן התמונה מחולקת לפאצ'ים זרים ומכניסה אותו לרשת הטרנספורמר מסוג Swin שמטרתו להפיק את הייצוג הלטנטי של תמונת המסמך. לאחר מכן ייצוג לטנטי זה מוזן לדקודר שהוא גם טרנספומר שמטרתו לפענח את המסמך ולהציג אותו בשפת markdown, שניתן להפוך אותה ל-pdf בקלות.   
  
הדאטהסט לאימון בנוי מתמונות של מאמרים ומייצוגם בשפת markdown (למסמך קיים בצורה ״דיגיטלית״ ניתן לתרגם את ייצוג ה-tex של המסמך לשפת ה-markdown הזו). כמובן משתמשים במגוון אוגמנציות של תמונות המסמכים לאימון המודל שלהם כדי לשפר את יכולת ההכללה שלו. כמובן השיטה המוצעת עדיין מוגבלת ודורשת עיבוד מקדים לא קל של תמונות המאמרים אבל זו התחלה טובה.