# Review 103: [Short] Meta-Transformer: A Unified Framework for Multimodal Learning, 22.07.23

**Paper: https://arxiv.org/abs/1902.03356v3**

https://huggingface.co/papers/2307.10802

זוכרים את ImageBind של מטא עם המודל שלהם מקבל דאטה מ-6 סוגים שונים. זה היה מרשים נכון? אז!   
  
עכשיו ב-shorthebrewpapereviews# קבלו את Meta-Transformer שיודע לעבוד עם לא פחות מ 12 סוגים שונים של דאטה כולל דאטה טבלאי, סדרות עתיות, קרינה אינפרא רד, גרפים, ואפילו דאטה מצילומי רנטגן. בגדול הם הצליחו לבצע טוקניזציה ״יוניפורמית״ לכל סוגי הדאטה האלו ואחריה בא המקודד שמשכן (embed) אותם לאותו המרחב. נראה לכם דמיוני?   
  
גם לי האמת אז בואו נצלול(טיפה) לפרטים של טוקניזציה של כמה סוגים של דאטה: לתמונות ולטקסט זה די ברור ומובן אבל איך נעשה טוקניזציה לענן נקודות למשל? כאן משתמשים בשיטת Farthest Point Sampling (FPS) שדוגם תת קבוצה של נקודות הענן באופן אחיד ולאחר מכן מקלסטרים את הנקודות סביב נקודות אלו לפי k-nearest neighbors. כל קלסטר יהווה טוקן שיוכנס למקודד.   
  
בנוסף לטוקנים מכניסים לאנקודר את הטוקן המיוחד המכיל את סוג הדאטה. מכיוון שמשתמשים בטרנספורמטורים מכניסים בנוסף גם את קידוד תלוי המיקום (positional encoding). בקיצור הרחבה נחמדה של ImageBind.