05 ינואר, 2024

**פגישה 3 - פרויקט גמר - קוד המס בארה"ב**

משתתפים: ד"ר תמיר שאנן, ד"ר אור אנידג'ר, תומר סורוז'ון, אורי בראל, רויטל יוסופוב ויפעת גסטפריינד

* הוצגה מצגת (ראו נספח) המפרטת מספר גרפים להצגה ויזאולית של הדמיון וההבדלים, בהיבטים שונים, בין קוד המס לשנת 2017 לבין קוד המס 2018. כמו כן, בהתאם לסיכום הפגישה מיום 22.12.24, הוצגו תוצאות ניתוח המבנה ההיררכי של הקוד ברמת כותרות וכמות המילים הנמצאת תחת כל היררכיה.
* מהתוצאות עולה, כי על אף שהמבנה נשאר דומה בין השנים, כמות התוכן והפירוט ב-2018 גדלה, דבר שעשוי להצביע על מורכבות רבה בקוד המס משנת 2018, ביחס לזה של 2017.
* במסגרת ההשוואות בין קודי המס לשנים 2017 ו-2018, זוהה פער גדול באופן יחסי בין השנים בהיבט של כמות המילים וכן, בתוצאות ההתאמות עבור ההיררכיות השונות, דבר המחזק את הסברה כי אכן קוד המס לשנת 2018 מורכב יותר. צוין, כי הפער וה"קפיצה" המשמעותית שהוצגה בניתוח ראשי פרקים מסוג "Part", ככל הנראה, מקורה בכפילות מסוימת שנוצרה עת שליפת המידע מקובץ ה- PDF של קוד המס לשנת 2018. הנושא בטיפול ובבדיקה.
* תמיר ציין, כי עלייה בחלק מהביטויים שהוצגו בניתוח שבוצע (N-Grams), יכולה להעיד על שינויים אותם ציפו לראות בקוד המס, לאור ניסוחי החוקים החדשים והרפורמות שנוספו בו. לבקשת הצוות, תמיר ודורון יעבירו דוגמאות נוספות לביטויים / רצפי מילים ספציפיים, אשר יהוו "עוגנים" בניתוח נוסף דומה שיבוצע על תוכן הפסקאות, בין קודי המס לשנים השונות.
* אור ציין, כי ש לנרמל את תוצאות הניתוח הסמנטי בוצע על כותרות דומות, על בסיס מודל LaBERT, ולהציג התוצאות במונחי cosine similarity, המייצגים את הקשרים בין הביטויים בצורה טובה יותר.
* סוכם כי כתיבת החלק במאמר המתייחס למודלי השפה ולניתוח שבוצע, ייכתב בלשון שאינה טכנולוגית לגמרי ותוצג באופן שיאפשר לקורא הסביר, גם כזה שאינו בקיא בתחומי האלגוריתמים בהם אנו עוסקים, להבין את המודל ברמה הכללית.
* המשך עבודה
  + טיוב ופתרון סופי לשליפת המידע מקוד המס של שנת 2018 והרצה מחדש של הניתוחים עד כה (על בסיס הטיוב שיבוצע).
  + נרמול תוצאות הניתוח הסמנטי שבוצע, על בסיס מודל LaBERT, והצגתן במונחי cosine similarity.
  + המשך עבודה בהתאם ללוחות הזמנים שפורטו במצגת, בדגש על התחלת כתיבת החלק הרלוונטי של הניתוח המבוצע למאמר ולספר הפרויקט, על בסיס האפיון הקיים והתוצאות החדשות.
  + תמיר ודורון -
    - דוגמאות נוספות לביטויים / רצפי מילים ספציפיים, אשר יהוו "עוגנים" בניתוח נוסף דומה שיבוצע על תוכן הפסקאות, בין קודי המס לשנים השונות
    - ניסוח טענות פרטניות לצורך המחקר, אשר אנו מבקשים להוכיח בעבודתנו, על מנת שהצוות יוכל לוודא שאל מול כל אחת מהן קיים ניתוח שלם העונה על הטענה.

**נספח - מצגת לפגישה 3**

****