Data retriever

שלב א'

[יתכן שיש לחלק זאת לשני שירותים – קליטת נתונים והטענתם]

1. שליחת קבצי התיקייה:

מעבר על (כל ה)קבצים בתיקייה קיימת – המעבר יתבצע על הקבצים הממוספרים. – ככל הנראה באמצעות לולאה שרצה על גודל (קבצי) התיקייה והשמת המספר במיקום המתאים.

\*\*\* נמצאה פונקציה שנכנסת לתיקייה וקוראת את הנתיבים מתוכה

1. פרטי המעטפת:
2. כיצד שולפים
3. שליפה של פרטים נחוצים תוך מחשבה עניינית (לכאורה גודל הקובץ ותאריך היצירה שלו)

פונקציות:

manager

Get\_meta\_data

Service:

שלב ב'

שלב ד':

חישוב טקסט מסוכן:

המקרה הגבוה ביותר (100%): כל המשפט מכיל את כל המילים ברשימה המסוכנת יותר

היינו: אחוז (מספר הופעות מילים בעיתיות חלקי מספר המילים במשפט – ספירה נפרדת לכל רשימה – הספירה שהתקבלה מהרשימה הפחות מסוכנת – האחוז נחתך בחצי) המילים הבעיתיות כפול אחוז (המילים השונות מתוך כלל המילים ברשימה הרלוונטית להם) הגיוון

בעיית צמדי המילים:

באמצעות חיפוש :re (אם החיפוש מתבצע על הטקסט כרשימה – נעלמות צמדי המילים)

אם אעבור על הטקסט ואחפש אותו במאגר – לא אתפוס צמדי מילים

לעומת זאת אם אעבור על המאגר – כולל צמדי מילים אוכל למצוא בטקסט

השימוש בre למקרה שהטקסט יכיל המשך של מילה מהמאגר