

## פרויקט במדעי הרוח הדיגיטליים – בותרת צד בחקיקה

בפרויקט זה נתקשנו לבתוב יישום המקובל בקלט קבצי וואח המכנים מידע על חוקי הכנסת, ומאפשר לבודק באלו חוקים מופיעים סעיפים מסוימים על ידי בותרת צד מסוימת.

קבצי הואה שמתקיים בקלט הנם חצי מובנים, זאת מכיוון שלא בכל החוקים קיימים ערכים לתגים השונים, בלומר אין זה בהכרחSCP של Tag בקובץ ימולא ע"י ערך.

הפענו אלגוריתם חיפוש בקבצים שאפשר לנו ללקחת את המידע, במידה וקיטם, ולשמור אותו במבנה נתונים מסווג dictionary כך שלכל ערך יש מפתח משלו.

בעבר הנו מקבעו הואה לבניונה הנתונים לא שמרנו את המבנה המתואג אלא את תוכן הסעיפים הרלוונטיים לתוכנית, בלומר שמרנו רק את תוכן הסעיף שיש לו בותרת צד בצורה בזאת השומרת על מבנה הסעיף והוא מוניניות להציג למשתמש (בותרת, מספר סעיף ותוכן).

מכיוון שלעתים הגדרות זרות או דומות מופיעות בהבדלים נתונים הפענו אלגוריתם של מתייג חלקו דבר (YAP) העושה ניתוח מורפולוגי על כל בותרת צד. מכך קיבלנו את:

segmented text – אשר מכיל את תוכנת הרצף שבו מילה מוצגת בצורה הבסיסית שלה.

בהתנן בותרת צד אותה המשמש הכוון לטיבת החיפוש אנו מפעילות אלגוריתם למציאת

בותרות צד דומות.

האלגוריתם מפעיל את מתייג חלקו הדבר גם על מילת החיפוש, ובודק האם יש התאמה בין מילת החיפוש ותוצאות המתואג שלה ובין בותרת צד נתונה ותוצאות המתואג שלה.

את התוצאות אנו מציגות בצורה טבלה המאפשרת השוואת בין הסעיפים השונים. מבנה הטבלה הוא בצורה צדו בה יש חלוקה לבותרת הצד, שם החוק בה נמצא, שנה בה הוא פורסם ותוכן הסעיף עצמוו.

בתוכן הסעיף קיים כפתרון המאפשר הצגה מלאה/חלקי של התוכן ובמו כן קיימת אפשרות להציג את התוצאות ממיניות לפי תאריך פרסום החוק.