

## תרגיל 2: מנתח לקסיקלי

- א. מנתח לקסיקלי – הגדרות ושימוש.
- ב. Lex – כלי ליצירת מנתחים לקסיקליים.

### א. מנתח לקסיקלי

#### הגדרות

1. אסימון (*token*) - יחידה בסיסית של השפה. (בהמשך נראה שמשמש כטרמינל בדקדוק של השפה).
2. לקסמה (*lexeme*) - מחרוזת בקלט אשר מתאימה לאסימון מסויים.
3. תבנית (*pattern*) - ביטוי רגולארי אשר מגדיר אסימונים או קבוצת לקסמות.

מנתח לקסיקלי – מזהה שפה רגולארית אשר מוגדרת על ידי ביטוי רגולארי (תבנית) כל מילה בשפה הזאת נקראת לקסמה ומשויכת לאסימון.

### דוגמה: מספרים טבעיים.

1. אסימון: *NUM*
2. לקסמה: 95, 7, 88...
3. תבנית:  $(0+1+2+3+4+5+6+7+8+9)^+$

### תפקידי המנתח הלקסיקלי

1. זיהוי מחרוזות בקלט והפיכתן לאסימונים. (מילים שמורות, קבועים, אופרטורים, מזהים...)
2. סינון תבניות מסוימות (למשל הערות, או white-spaces)
3. זיהוי שגיאות לקסיקליות (למשל הופעה של "\$" או "@" בתוך תוכנית, לא במחרוזת או הערה)
4. ביצוע פעולות סמנטיות (שיוך תכונות לאסימונים)

### דוגמה:

נתון קלט:  $a = x + y;$   
 חלוקה ללקסמות:  $|a| = |x| + |y|$ ;  $|;$   
 שיוך אסימונים: id, op, id, op, id, sep

אם הקלט יהיה  $d = s * t$  הפלט יהיה בדיוק אותם אסימונים.  
 לכן יש להעביר יותר מידע לפלט: id("d"), op('='), id("s"), op('\*'), id("t"), sep  
 (ל sep אין צורך להוסיף מידע, כי מפריד מוגדר להיות רק ";")

- ב. Lex – כלי ליצירת מנתחים לקסיקליים. Lex הוא לא מנתח לקסיקלי.

### בניית המנתח הלקסיקלי

קלט ל Lex: קובץ הגדרת מנתח (source.lex).  
 פלט: תוכנית בשפת C שמגדירה את המנתח הלקסיקלי.  
 את התוכנית מקמפלים (למשל ב gcc)  
 הפלט של ה gcc הוא קובץ הרצה והוא המנתח הלקסיקלי.

### עבודה עם המנתח הלקסיקלי

הקלט של המנתח הלקסיקלי הוא תוכנית שרוצים לנתח, והפלט שלו הוא אסימונים עם תכונות.

התמודדות עם קונפליקטים של המנתח הלקסיקלי:

1. אם יש שתי מחרוזות ואחת y יותר ארוכה מ x, ו x מוכלת ב y אז המנתח יבחר לזהות את y.
2. אם שתי המחרוזות באותו האורך נזהה את המחרוזת ע"פ הכלל שמופיע ראשון בקובץ lex.

