

הגשה דרך moodle. אפשר להגיש בזוגות.

מה צריך להגיש: קובץ zip הכולל את הקוד שאתם כתבתם ודוגמא לקלט עם הפלט המתאים. אין להגיש את הקובץ שיצר flex.

התוכנה של flex נמצאת בתיקיה flex ב-moodle.

יש לכתוב בעזרת flex תוכנית שכותבת לפלט תאור של האסימונים שהיא מזהה בקלט שלה.

הקלט כולל מידע על נצחונות של שחקני טניס בטורנירים שונים. הקלט יהיה בקובץ טקסט שינתן כ- command line argument לתוכנית שלכם. תאור מפורט יותר של הקלט עם דוגמא מופיע בהמשך.

פלט התוכנית תכתוב לפלט (ל- standard output) תאור של האסימונים שהיא מוצאת בקלט.

עבור כל אסימון שמופיע בקלט יופיע בפלט שם האסימון, ה- lexeme (המחרוזת שהופיעה בקלט שמהווה את האסימון) והערך הסמנטי של האסימון (במקרה ויש לו כזה). התאור של כל אסימון יופיע בשורה נפרדת.

התוכנית תשתמש במנתח לקסיקלי - הפונקציה yylex() -- שתכתב ע"י flex. הממשק ל- yylex יהיה כמו שנלמד בכיתה:

yylex() תחזיר מספר המציין את סוג האסימון שמצאה. במקרה שלאסימון יש ערך סמנטי היא תכתוב אותו למשתנה הגלובלי yylval שיוגדר כ- union.

המנתח הלקסיקלי מיועד להיות חלק מתכנית גדולה יותר (אותה אינכם צריכים לכתוב) שמטרתה להדפיס את שם השחקן (להבדיל משחקנית) שזכה במספר הגבוה ביותר של טורנירים באליפות הפתוחה של אוסטרליה. לאור זאת יש להחליט לאיזה אסימונים צריך להיות ערך סמנטי. לרוב האסימונים לא יהיה צורך להגדיר ערך סמנטי.

אתם צריכים להחליט לבד מהם סוגי האסימונים.

white space (רווחים, טאבים ו- newlines) לא יוגדרו כאסימונים. אבל כל תו אחר שמופיע בקלט יהיה חלק מאסימון.

אחד האסימונים יהיה GENDER. ה- lexemes המתאימים לאסימון זה יהיו <Man> ו- <Woman>.

במקרה שמופיע בקלט תו שאינו חלק מאסימון חוקי יש לכתוב (ל- standard error) הודעת שגיאה המפרטת את השורה בה נפלה השגיאה ואת התו השגוי. לאחר מכן התוכנית תמשיך ותזהה אסימונים.

שימו לב שכאן מתבקש לכתוב רק מנתח לקסיקלי. הסדר שבו מופיעים האסימונים אינו מענינו של המנתח הזה.

תאור הקלט

הקלט כולל רשימה של שחקני טניס. עבור כל שחקן או שחקנית מופיעים הנתונים הבאים: שם השחקן או השחקנית, המגדר (גבר או אשה), השנים בהן זכה או זכתה בכל אחד מהטורנירים: ווימבלדון, אליפות אוסטרליה הפתוחה ואליפות צרפת הפתוחה.

שני הפרטים האחרונים הם אופציונליים. למשל בדוגמא שתובא בהמשך, עבור Simona Halep חסר החלק של אליפות אוסטרליה הפתוחה (שבה מעולם לא זכתה).

שנות הזכייה מופיעים כרשימה של שנים מופרדים ע"י פסיקים.

בתור קיצור עשוי להופיע גם טווח של שנים למשל 1982-1987

זה קיצור של 1982, 1983, 1984, 1985, 1986, 1987

בקלט יכול להופיע גם whitespace (רווחים, טובים ו- newlines).

הנה דוגמא לקלט:

**** Winners ****

```
<name> "Martina Navratilova" <Woman>
  <Wimbledon> 1978, 1979, 1982-1987, 1990
  <Australian Open> 1981,1983,1985
  <French Open> 1982, 1984

<name> 'Roger Federer' <Man>
  <Wimbledon> 2003-2007, 2009, 2012, 2017
  <Australian Open> 2004, 2006, 2007, 2010, 2017, 2018
  <French Open> 2009

<name> "Novak Djokovic" <Man>
  <Wimbledon> 2011, 2014, 2015, 2018, 2019, 2021
  <Australian Open> 2008, 2011-2013, 2015, 2016, 2019-
2021
  <French Open> 2016, 2021

<name> 'Simona Halep' <Woman>
  <Wimbledon> 2019
  <French Open> 2018
```

לא ניתנה כאן הגדרה מדויקת של מה יכול להופיע בקלט. למשל לא נאמר אם השם של שחקן תמיד כולל שם פרטי ושם משפחה והאם כל אחד מהשמות חייב להתחיל באות גדולה. עליכם לקבוע את ההגדרות המדויקות. כל הגדרה סבירה תתקבל. (וכן, אין כאן הגדרה של מהי "הגדרה סבירה"). שימו לב שהשמות עשויים להיות מוקפים מכל צד בגרש בודד או בגרשיים.

תזכורת: הכנת תוכנית בעזרת flex.

(ההערות מתייחסות ל- Windows ול- Linux)

נניח שברשותנו קובץ קלט ל- flex שהכנו בעזרת text editor (למשל Notepad++). נקרא לקובץ tennis.lex.

נריץ את הפקודות הבאות בחלון המריץ command prompt (או shell אם עובדים על linux).

```
1. מריצים את flex
flex tennis.lex
נוצר קובץ lex.yy.c
```

הערה: כדי לקבל הודעות דיבוג כשמריצים את yacc אפשר להריץ את flex עם האופציה -d בצורה כזאת:

```
flex -d tennis.lex
```

2. יש לקמפל את קובץ ה- C ש- flex יצרו עבורנו. (כמובן שאם התוכנית שלנו כוללת קבצים נוספים יש לקמפל גם אותם). לצורך כך ניתן להשתמש בכל קומפיילר לשפת C.

אם נשתמש בקומפיילר gcc (קומפיילר פופולרי של GNU) הפקודה היא:

```
gcc -o tennis.exe lex.yy.c
```

כאן האופציה -o מציינת את שם הקובץ שהוא התוצר של הקומפילציה (במקרה זה שם הקובץ הוא tennis.exe. אם עובדים על linux לא צריך סיומת .exe)

3. נכין קובץ טקסט שנקרא לו test_tennis.txt ובו נכתוב קלט לדוגמא למשל

```
<name> "Martina Navratilova" <Woman>
<Wimbledon> 1978, 1979, 1982-1987, 1990
<Australian Open> 1981,1983,1985
<French Open> 1982, 1984
```

. . .

נריץ את הפקודה

```
tennis test_tennis.txt
```

ב- linux זה יראה כך:

```
./tennis test_tennis.txt
```

ובפלט תופיע רשימה של אסימונים.

הפלט יכלול למשל בין היתר את השורות הבאות (השורה הראשונה היא כותרת שנועדה להקל על קריאת הפלט):

TOKEN	LEXEME	SEMANTIC VALUE

...		
GENDER	<Woman>	WOMAN

בדוגמא הזאת הערך הסמנטי נכתב באותיות גדולות באופן שרירותי. לעיתים

קרובות הערך הסמנטי יכול להיות זהה או דומה ל- lexeme. בתוכנית עצמה יתכן שהם ייוצגו בצורה שונה. למשל ה- lexeme יהיה מחרוזת והערך הסמנטי יכול להיות enum type:

```
enum gender { MAN, WOMAN };
```

דוגמאות לתוכניות שהוכנו בעזרת flex

השאלה הראשונה בכל בחינה ישנה (ראו בתיקה "בחינות ישנות" ב- moodle) עוסקת ב- flex & bison. יש שם הרבה דוגמאות לקוד שמיועד ל- flex (התעלמו לעת עתה מהקוד שמיועד ל- bison).

יש גם בתיקה של bison ב- moodle דוגמאות לתוכניות שנכתבו בעזרת flex (ו- bison).

הערות אחרונות

בתיקת ה- flex ב- moodle יש גם סיכום בעברית על flex. (וגם manual באנגלית).

אל תשכחו לכתוב את הפקודה הבאה בחלק הראשון של קובץ הקלט ל- flex שתכינו:

```
%option noyywrap
```

זה אומר ל- flex שאנו לא משתמשים בפונקציה yywrap (למי שמעוניין, אפשר למצוא ב- manual הסבר על הפונקציה).

זכרו ש- flex "רגיש" לרווחים בקלט שלו. את כל הביטויים הרגולריים יש להצמיד לתחילת שורה. אם הביטוי הרגולרי שלכם צריך לכלול רווח אז יש להקיף אותו בגרשיים או בסוגריים מרובעות או לשים לפניו backslash וזה כדי שהרווח ייצג את התו רווח ולא יתפרש בטעות כסוף הביטוי הרגולרי.

כאמור מקודם, ניתן להפעיל את flex עם אופציה -d כדי לקבל הודעות דיבוג כש- yylex רצה. לדוגמא

```
flex -d tennis.lex
```

הודעות הדיבוג מפרטות (עבור כל התאמה לביטוי רגולרי) איזה ביטוי רגולרי הותאם ומה היתה המחרוזת בקלט שהתאימה לו.

אם אתם מעתיקים קבצים מ- linux ל- Windows (או להיפך) אז שימו לב שעל Windows יש מוסכמה שסופי שורות מיוצגים ע"י שני תווים: \r\n (התו carriage return ואחריו התו newline) בעוד שעל linux המוסכמה היא סוף שורה מיוצג ע"י newline בלבד.

בכל מקרה לא יזיק לשים בקובץ הקלט ל- flex כלל עם ביטוי רגולרי \r שיתאים לתווי carriage return ולא יעשה איתם כלום.

בהצלחה!