

תרגיל מס' 7 – פייתון (מתרגלת אחראי: מור)

הוראות הגשה ----- גרסת פייתון 3.6 ומעלה

שאלות בנוגע לתרגיל נא להפנות דרך פורום הקורס שנפתח במיוחד לשם כך:

<https://piazza.com/biu.ac.il/fall2018/89110/>.

אם לא נענתה תשובה תוך 24 שעות, נא לשלוח אלי (מור) מייל עם לינק לדיון הרלוונטי ואענה. המייל הוא: mor.sinay@gmail.com

בכל מייל יש לציין שם, שם משתמש, מס' קורס, וקבוצת תרגול.

- מועד פירסום: 6.1.19
- מועד אחרון להגשה: 16.1.19
- יש לשלוח את הקבצים באמצעות האתר: <http://help.cs.biu.ac.il/submit.htm>
- לפני חלוף התאריך הנקוב לעיל.
- שם ההגשה של התרגיל: ex7.py.
- יש להקפיד מאוד על כל הוראות עיצוב הקלט והפלט, כמפורט בכל סעיף וסעיף. על הפלט להיראות בדיוק כמו בדוגמאות. אין להוסיף או להשמיט רווחים או תווים אחרים ואין להחליף אותיות גדולות בקטנות או להיפך ☹️ אי-הקפדה על פרטים אלה עלול לגרור ירידה משמעותית ביותר בציון התרגיל עד כדי 0. ראו הוזהרתם!
- להזכירכם, העבודה היא אישית. "עבודה משותפת" דינה כהעתקה.
- אין להדפיס למסך שום דבר מעבר למה שנתבקש בתרגיל.
- יש לוודא שהתרגיל מתקמפל ורץ על השרתים באוניברסיטה (2u) ללא שגיאות/אזהרות.
- אתם יכולים לעבוד עם כל עורך טקסטואלי שאתם מעדיפים. להזכירכם – pico בשרתי linux שבמעבדות; notepad ב-windows; עבור פייתון מומלץ לעבוד ב-pycharm.

הקפידו על כתיבה לפי קובץ ה-Coding-Style שבאתר הקורס!!

הנחיות עבור חלק זה:

- שם קובץ מקור (source file) שיש לשלוח: ex7.py, out.txt
- קובץ ex7.py הוא קובץ התרגיל שלכם
- קובץ out.txt הוא קובץ טקסט ריק
- אין לעשות import לשום ספריה מלבד sys, os

רקע

אתרים כמו (IMDB (Internet Movie Data Base מחזיקים את כל האינפורמציה על סרטים והשחקנים שמשחקים בהם. אם אתם מחפשים סרט באתר, אתם מקבלים עמוד עם כל המידע על הסרט כולל השחקנים שמשחקים בו. אם אתם לוחצים על השם של השחקן/נית אתם מועברים לעמוד של השחקן בו יש את רשימת כל הסרטים שהם השתתפו בו. במשימה זאת תממשו גרסה מאוד פרימיטיבית של אתר זה.

המשימהחלק א – pre processing

באתר הקורס ישנו קובץ שנקרא movies.txt, שבו כל שורה היא בצורת:
Name_of_actor, Movie1, Movie2, Movie3, ...

שימו קובץ זה באותה תיקיה של קובץ הפייתון, אין צורך להגיש את הקובץ, הוא נועד לעזור לכם לבדוק את התרגיל.
עליכם לכתוב פונקציה שנקראת בתחילת התוכנית, ויוצרת את data basen שלכם (עוד על כך בהמשך).
על הפונקציה לפתוח את הקובץ ששמו ניתן input לתוכנית (בשביל שתוכלו לבדוק את עצמכם העלתי לכם את הקובץ movie.txt), לקרוא אותו וליצור ממנו בסיס נתונים כדי שתוכלו לעבוד עליו אחר כך.
אל דאגה, הכול יוסבר בהמשך. ☺
שימו לב ששמות של סרטים או שחקנים לא יכולים להתחיל עם רווחים (תוכלו להעלים את הרווחים עם **strip** ולפצל את המחרוזת עם **split**)
בשביל לקרוא את הפרמטרים שהתקבלו לתוכנית (בדומה לפרמטרים לmain שלמדנו ב-C, השתמשו בספריה sys --- sys.argv)

חלק ב – IMDB משלכם

לאחר שיצרתם את בסיס הנתונים שלכם, כתבו למסך את התפריט הבא:

Please select an option:

- 1) Query by movies
- 2) Query by actor
- 3) Insert a new movie
- 4) Save and Exit
- 5) Exit

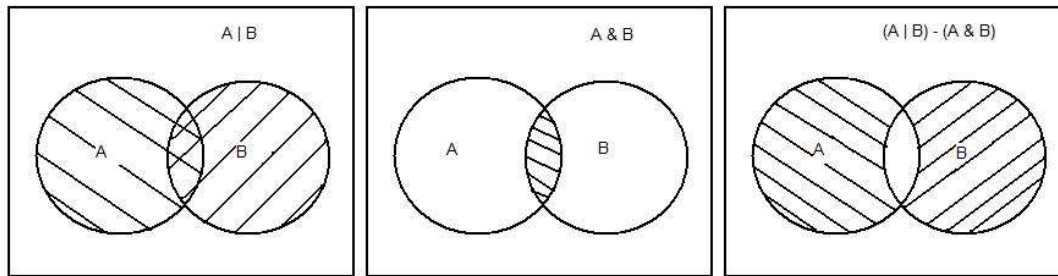
אם המשתמש בחר 1, עליו להכניס שמות של שני סרטים מופרדים בפסיק ולאחר מכן אופרטור מבין השלושה: & | ^ . גם מופרד בפסיק.

לדוגמא: *Runaway Bride, Ocean's Eleven, &*

לפי האופרטור עליכם למצוא קבוצת שחקנים כך ש:

1. **אם קיבלתם &** - הדפיסו למסך את כל השחקנים שישחקו גם בסרט הראשון וגם בסרט השני (קבוצת החיתוך בין שני הסרטים).

2. **אם קיבלתם** | - הדפיסו למסך את כל השחקנים ששיחקו בשני הסרטים (איחוד קבוצות השחקנים בשני הסרטים).
3. **אם קיבלתם** ^ - הדפיסו את רשימת השחקנים ששיחקו באחד מהסרטים אך לא בשניהם (ההפרש הסימטרי בין קבוצות השחקנים בשני הסרטים).



אם **הסרט** לא קיים במסד הנתונים/ האופרטור שגוי / שהוכנסו פחות נתונים אדי הקלט לא תקין.
ראו טיפול בהמשך.

אם המשתמש בחר 2, עליו להכניס שם של שחקן, עליכם להדפיס למסך את כל שחקני המשנה ששיחקו בסרט כלשהו עם השחקן.
אם השחקן לא קיים במסד הנתונים אדי הקלט לא תקין.
ראו טיפול בהמשך.
אם אין שחקנים ששיחקו איתו בשום סרט, אדי הקלט תקין אבל ראו הדפסה בהמשך.

אם המשתמש בחר 3, עליו להכניס סרט חדש למבנה או לעדכן סרט קיים.
עליו להכניס שורה בצורה הבאה (השתמשו ב-input):
Movie_name, Actor1, Actor2, Actor3, ...

חייב להיות לפחות שחקן אחד, אם אין, אדי הקלט לא תקין.
ראו טיפול בהמשך.
אם הסרט קיים כבר במסד הנתונים – עדכנו את השחקנים להיות האיחוד של השחקנים. (מה שקיים ומה שקיבלנו בקלט)

אם המשתמש בחר 4, שימרו את מסד הנתונים שלנו בקובץ וסיימו את התוכנית.
את שם הקובץ קיבלתם בקלט לתוכנית.
נרצה לשמור את מסד הנתונים שלנו כך שנוכל לפתוח אותו בריצה הבאה, לכן כל שורה תהיה בדיוק באותה צורה בה קראנו את המידע מלכתחילה
Name_of_actor, Movie1, Movie2, Movie3, ...

כדי שתהיה תאימות, שמות השחקנים צריכים להיות ממיינים לקסיקוגרפית, כנ"ל גם שמות הסרטים.
(השתמשו בפונקציית sort)

את השם של קובץ הפלט גם נקבל כארגומנט לתוכנית (sys.argv), הקובץ קיים כי שלחתם אותו, אז אתם צריכים לדרוס אותו.

לפני שאתם פותחים את הקובץ, אנחנו צריכים לשנות לו את ההרשאות (שלא יהיה קריאה בלבד)
תוסיפו את השורה os.chmod(output_file, 0o777)
כאשר output_file זה שם הקובץ שלכם

אם המשתמש בחר 5, סיימו את התוכנית.

רלוונטי לכל המשימות

שמות השחקנים והסרטים הם **CASE SENSITIVE** כלומר **Brad Pitt** שונה מ**brad pitt**.

בכל מקרה של **קלט לא תקין** יש להדפיס Error לרדת שורה ולחזור לתפריט הראשי.

בכל מקרה שמתקבלת **קבוצת שחקנים ריקה** יש להדפיס:

There are no actors in this group

לרדת שורה ולחזור לתפריט הראשי.

מסד הנתונים

אז כפי שאתם כבר מבינים, יש חשיבות לצורה בה אתם מחזיקים את הנתונים.

שמות הסרטים הם ייחודיים ואנחנו רוצים למצוא שחקנים לפי איזה קריטריון מתוך הסרטים.

בpython ישנו מבנה אשר נקרא dictionary. מבנה זה בנוי על זוגות של key,value כאשר כל key הוא ייחודי ולא יכול לחזור פעמים בdictionary, אך אותו ערך יכול להופיע כמה פעמים למפתחות שונים. ולכן

מבנה זה הוא בחירה טובה עבורנו, כך שהסרטים הם המפתחות ורשימת השחקנים היא הערכים של המילון. (שני סרטים לא יהיו בעלי אותו שם אבל שחקן יכול להופיע בכמה סרטים)

הדבר הבא שעליכם להחליט הוא באיזה מבנה להשתמש כדי להחזיק את רשימת השחקנים בכל ערך במילון. במקרה זה ישנם שני אפשרויות:

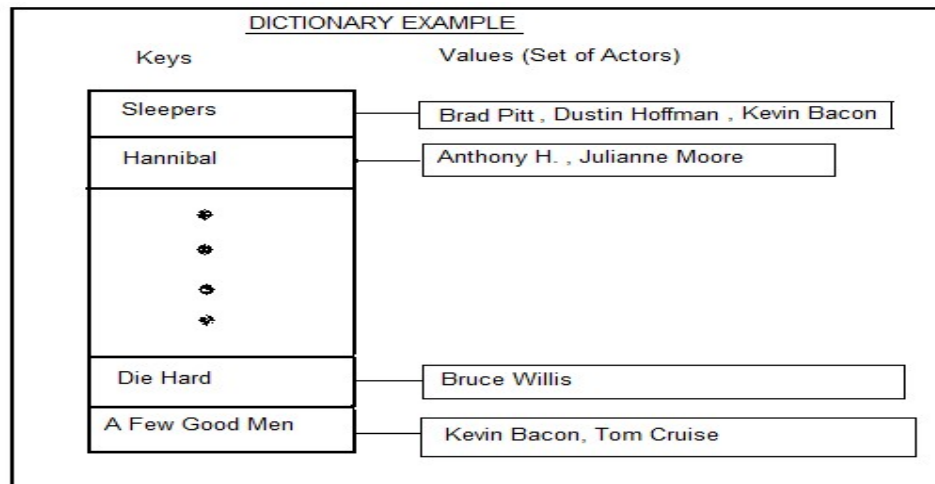
1. List – מבנה אשר דומה למערך.

2. Set – מבנה אשר גם הוא דומה למערך, אך בניגוד למערך הערכים בו הם ייחודיים, כלומר אף ערך

לא יכול להופיע בו יותר מפעם אחת. יש למחלקה שלו כבר מימוש לפונקציות union,

intersection ו-symmetric_difference כך שזה יחסוך לכם זמן בביצוע הפעולות הנ"ל.

מסד הנתונים שלכם צריך להראות משהו כמו:



השתמשו בקישור הבא כדי ללמוד יותר על מבני נתונים בpython:-

<https://docs.python.org/3.6/tutorial/datastructures.html>

עוד כמה טיפים והערות

- על התוכנית שלכם להכיל לפחות חמש פונקציות (אבל אני מציעה לכם שזה יהיה קצת יותר. ☺)
- פונקציית **add** של **set** יכולה לשמש כדי להוסיף שחקנים לרשימת שחקנים של כל סרט במילון.
- כדי למצוא את שחקני המשנה יהיה עליכם לעבור על הסרטים במילון ולבדוק האם השחקן נמצא ב-**set** השחקנים של הסרט. מילת המפתח **in** יכולה לעזור לכם בזה.
- בכדי להריץ את התוכנית שלכם עם קובץ הטקסט **movie.txt** ושידפיס את מסד הנתונים ל-**out.txt** כתבו את השורה הבאה במובקס\הסביבה איתה אתם עובדים
python ex7.py movie.txt out.txt
- מצורף לכם קובץ ה- **IO** של הבדיקה. (אתם יכולים לפתוח בכל עורך טקסט)



בהצלחה!