

תרגיל 3 בעיבוד שפה טבעית לעברית – Parsing וביטויי זמן

אפשר להגיש בזוגות. להגשה עד 30.6.2020.

1. מטרת התרגיל היא, בהינתן תאריך ייחוס, לתרגם ביטויי זמן יחסיים בעברית לתאריך חדש. כל התאריכים לועזיים.
לדוגמה, אם תאריך הייחוס הוא ה-11 ביוני 2020, והביטוי הוא "מחר.", יש לתרגם את הביטוי ל-12 ביוני 2020 – ולהדפיס 20200612 .

2. אוצר המילים שיש לכסות:
היום, אתמול, מחר, בעוד, לפני,
מספרים בין 1 ל-31,
יום, שבוע, חודש.
ימים, שבועות, חודשים.
יומיים, שבועיים, חדשיים.
(לדוגמה: בעוד שלושה ימים.).

3. כל ביטויי זמן חוקי מסתיים בנקודה.

יש להגיש קובץ פייתון (מודול), בשם date_translator.py, המכיל:

```
class DateTranslator:
    def __init__(self, init_date=None):
        #if init_date is None, use date of the run
        #otherwise, expect a date in format yyyyymmdd

    def translate_date(self, date_tokens):
        #return a string with the translated date in format yyyyymmdd

    def tokenizer(self, date_phrase):
        #return list of tokens

    def run(self, date_phrase):
        #return translated date in format yyyyymmdd

def main():
    #driver function
    date_translator=DateTranslator("20200611")
    phrase="בעוד שלושה ימים."
    print(date_translator.run(phrase))
#20200614
```

בנוסף, יש להגיש קובץ בשם date_grammar.txt שיכיל את הדקדוק שמתאר ביטויי זמן.

ספריות מועילות:
time
arrow
delorean