**Regex – ל"ע**

**הגדרה -**  ביטוי בשפה רגולרית.

**הסבר** – רגקס הוא אוסף תווים וחוקים תחביריים שישמשו אותנו לחיפוש דפוסים בטקסט.

בתוך עולם הדאטה ובפרט בהתעסקות עם מידע גדול אנו צריכים לדעת לשלוט במידע שלנו בצורה היעילה והטובה ביותר – היום נלמד כיצד להשתמש ברגקס – נוכל לשלוף בSQL רק מידע שמתאים לדפוס שנכתוב או לקרוא לפונקציה שכתבנו בפייתון על כל המידע שלנו שעונה על תבנית מסוימת.

החוקים והנוסח של רגקס נשארים דומים עם שינויים מינוריים בכלים השונים בהם נשתמש ברגקס ולכן חשוב להכיר טוב את העקרונות והכתיבה הנכונה ומאוחר יותר יהיה קל ללמוד ליישם בתוכנות השונות.

[באתר](https://regexone.com/lesson/introduction_abcs) הבא תלמדו על התחביר הבסיסי ועל אופציות מתקדמות יותר ברגקס, שימו לב שכדי להפנים את הנושא צריך לתרגל אותו ולכן ענו על כל השאלות, כמו כן על חלק מהשאלות תוכלו לענות במספר דרכים – נסו לענות על כל שאלה בכמה דרכים, זה יפתח לנו יותר אופציות בהמשך כשנשתמש ברגקס על מידע אמיתי.

**תווים מיוחדים**

* נוכל לכתוב כל אות או מספר וכך לחפש הופעה שלו בטקסט, אם נכתוב א' יתאימו כל ההופעות של האות א' בטקטסים שונים.
* \d יתאים לתו שהינו מספר
* \D יתאים לתו שאינו מספר
* \w יתאים לתו שהינו אותיות (גדולות או קטנות) ו\או מספרים ובנוסף מקף תחתון, זהה לטווח הבא [A-Za-z0-9]
* \W יתאים לתו שאינו אות או מספר
* \s יתאים לרווח (Space tab) "hey friend" מכיל רווח בין 2 המילים ולכן טקסט זה מתאים
* \S יתאים לכל תו שאינו רווח
* נקודה "." מסמנת הופעה של כל תו (מספר, ספרה או תווים שונים - @! וכו')
* .\ יסמן נקודה
* סוגריים () משמשים לתפוס קבוצות בתוך טקסט – לדוגמא (שלום) מתוך הטקסט שלום לך יחזיר רק את המילה שלום, אם יש מספר קבוצות שמתאימות לדפוס שכתבנו נוכל לפנות לכל אחת מהקבוצות שנתפסו (נלמד בהמשך).
* ^ מסמן תחילת טקסט - ^שלום יתאים לטקסט שמתחיל אך ורק במילה שלום
* $ מסמן סיום טקסט – להתראות$ יתאים לכל טקסט שמסתיים אך ורק במילה להתראות
* \* אפס או יותר חזרות
* + אחת או יותר חזרות
* ? תו אופציונלי

**טווחים ומספר הופעות**

* סוגריים מרובעים [] מסמנים הופעה רק של תווים מסוימים – התווים אשר נמצאים בתוך הסוגריים.

[אב] יתאים לתווים שמכילים את האות א' ו/או את האות ב')

* סוגריים מרובעים שמתחילים עם כובע [^...] מסמן הופעה של כל תו אשר לא נמצא בתוך הסוגריים, לדוגמא – [^052] יתאים לכל תו שאינו 0, 5 או 2
* טווח בסוגריים מרובעות [א – ד] יתאים לכל תו שמכיל אותיות בין א' לד' (א,ב,ג,ד), ניתן לעשות זאת גם עם מספרים ובאנגלית עם אותיות גדולות [A-Z] או קטנות [a-z]
* מספר חזרות יסומן על ידי סוגריים מסולסלים {}:

1. א{2} יתאים לקטע מטקסט שמכיל את האות א' בדיוק פעמיים
2. א{1,3} יתאים לקטע מטקסט שמכיל את האות א' לפחות פעם אחת ולא יותר מ3 פעמים
3. [אבג]{6} יתאים לקטע מטקסט שמכיל את האותיות א' ב' או ג' 6 פעמים ברצף (לדוגמא: 'אבגאבג')

**שימו לב**

כעת תוכלו לחזור לקובץ זה כדי לעבור על החוקים כל פעם שתשתמשו ברגקס עד שתזכרו אותם בעל פה, חשוב לתרגל הרבה ולהשתמש בכמה שיותר שיטות כדי לעבוד בצורה היעילה ביותר.