

החוקים של ביטויים רגולאריים (Extended Regular Expression Style)

תו בודד כלשהו מתאים לעצמו. לדוגמה a מתאים ל- a .
קבוצת תווים בתוך סוגריים מרובעים מתאימה לתו בודד מתוך הקבוצה.
לדוגמה $[a-c6-8]$ מתאימה ל אחד מהתווים $a,b,c,6,7,8$.
קבוצת תווים בתוך סוגריים מרובעים שמתחילים בסימן $^$ מתאימה לתו בודד שאינו בתוך הקבוצה. לדוגמה $^[a-c]$ מתאימה לתו בודד שאינו a או b או c .
• (נקודה) מתאימה לתו בודד כלשהו (אבל לא למחרוזת ריקה).
 $^$ (שאינו בתוך סוגריים מרובעים) מציין התאמה לתחילת המחרוזת.
 $\$$ מציין התאמה לסוף המחרוזת.
חשוב לזכור: אם לא מופיעים $^$ ו- $\$$ בביטוי הרגולארי אז מספיקה התאמה לחלק (רציף) של המחרוזת הנבדקת. לדוגמה: הביטוי a מתאים למחרוזת bab אבל הביטוי $a^{\$}$ לא מתאים למחרוזת bab (אלא לתו a בלבד).

* מציין 0 או יותר חזרות של מה שלפני ה- *
+ מציין 1 או יותר חזרות של מה שלפני ה- +
? מציין 0 או 1 חזרות של מה שלפני ה- ?
 $\{n\}$ מציין בדיוק n חזרות של מה שלפני ה- $\{n\}$
 $\{n,m\}$ מציין מספר חזרות גדול שווה n וקטן שווה m של מה שלפני ה- $\{n,m\}$
 $\{,n\}$ מציין מספר חזרות קטן שווה n (כולל 0 חזרות) של מה שלפני ה- $\{,n\}$
 $\{n,\}$ מציין מספר חזרות גדול שווה n של מה שלפני ה- $\{n,\}$

בכל הכללים הנ"ל מה שלפני יכול להיות:

- תו בודד כלשהו. לדוגמה בביטוי ab^*c מה שלפני ה- * הוא b .
ולכן הכוונה ל- a אחד לאחר מכן 0 או יותר חזרות של b ולבסוף c אחד.
- קבוצת תווים בתוך סוגריים מרובעים (עם או בלי $^$ בהתחלה). לדוגמה בביטוי $a[0-9]^+c$ מה שלפני ה- + הוא $[0-9]$. ולכן הכוונה ל- a אחד לאחר מכן 1 או יותר חזרות של ספרות ולבסוף c אחד.
- ביטוי רגולארי בתוך סוגריים עגולים. לדוגמה בביטוי $a(bc)^2d$ הכוונה ל- a אחד לאחר מכן פעמים הרצף bc ולבסוף d אחד.

| מציין or לוגי בין ביטויים לדוגמה $cd | ab^{\$}$ יתאים למחרוזת ab או למחרוזת שמכילה את הרצף cd באיזשהו מקום.