

שתי קבוצות של סטודנטים קיבלו לשיפוט 20 תוצאות ראשונות של גוגל עבור כל אחת מ-2 שאלות: דיאטת אטקינס ומחשוב ענן. עבור כל שאלתא: הם נתבקשו למיין את 20 התוצאות שלה לקטגוריות רלוונטיות, היו 10 קטגוריות רלוונטיות שלהן לשאלתא, ה-1 - הכי רלוונטי, ה-10 - הכי לא רלוונטי לשאלתא. יש 2 קבוצות סטודנטים (A,B) ו-2 שאלות שכל אחת מהן שפטה - אז סך הכל 4 גיליונות אקסל עם שיפטים סטודנטים לפי שם השאלתא. כך בכל קובץ השורות אלו הסטודנטים - בכל שורה יש שיפטים של סטודנט אחד, עבור סט אחד של 20 תוצאות. העמודות אלו התוצאות (יש 20 כאלה בסט). בתוך כל תא רואים את מספר הקטגוריה שהסטודנט שייך תוצאה זו אליה (1-10). הסטודנטים נתבקשו לחזור שוב על תהליך השיפוט חודש מאוחר יותר ותוצאות השיפטים אלו נמצאות בעמודות 21-40 בקובץ ושוב חודש נוסף מאוחר יותר, ותוצאות שיפטים אלו נמצאות בעמודות 41-60 בקובץ. כלומר היו 3 סבבים של שיפוט. אז אם עוברים על שורה מס' 3 בעמודות 1-20 בגיליון ניתן לראות כיצד סטודנט (בשם ישעיהו פישמן בגיליון cloudB) מיין את תוצאות השאלתא: בקטגוריה הראשונה הוא הכניס את תוצאות 3,4,6,20, בקטגוריה 2 את 5,8,14,15,16, בקטגוריה 3 את 1,9,13,18, בקטגוריה 4: 2,7,10,14,19, ובקטגוריה 5 - 10 - לא הכניס אף תוצאה. כלומר הוא השתמש רק ב-4 קטגוריות. מטרתנו היא לבחון האם אנשים משנים את דעתם לאורך זמן לגבי רלוונטיות של תוצאות חיפוש. ולשם כך אנו רוצים לחשב את הדברים הבאים:

עבור כל גיליון עבור כל סבב שיפטים (בנפרד - 1) מספר הקטגוריות לכל סטודנט, (2) עבור כל קטגוריה - מספר הסטודנטים שהכניסו לפחות תוצאה אחת בקטגוריה נתונה (סך הכל יש 10 קטגוריות), (3) גודל הקטגוריות - מספר התוצאות שהוכנסו לכל קטגוריה על ידי כל סטודנט. בנוסף, בכדי להשוות שיפטים בסבבים שונים - נרצה לחשב JACCARD MEASURE עבור כל קטגוריה - מהי מידת השוני בין התוצאות שהוכנסו לתוכה על ידי אותו הסטודנט בסבבים השונים (יש 3 סבבים, יש להשוות כל זוג סבבים - 1 ו-2, 2 ו-3, 1 ו-3). בשלב הבא: נחשב את מידת הדמיון בין שיפוט הסטודנטים לדירוג של גוגל. כל התוצאות בגיליונות מסודרות לפי הסדר שלהם בגוגל (עמודה 1 - תוצאה מס' 1 ברשימת גוגל, עמודה 2 - תוצאה מס' 2 בגוגל וכד' עד 20). נחשב עבור כל סטודנט וכל קטגוריה - מהו דירוג גוגל הממוצע של התוצאות שהושמו בקטגוריה זו על ידי סטודנט זה. אנו מצפים לראות דירוג ממוצע נמוך יותר עבור קטגוריה נמוכה יותר. נחזור על חישוב זה על כל אחד מ-3 הסבבים של שיפטים. בנוסף, נחשב מתאם פירסון בין מספרי התוצאות (בגוגל) (בסקלה 1-20) לבין מספרי הקטגוריות שאליהן הושמו על ידי הסטודנט (סקלה 1-10). גם זאת עבור כל סטודנט ועבור 3 הסבבים בנפרד. את התוצאות יש לייצא לקובץ אקסל מסודר.