

קוים מנחים לתיוג שמות פרטיים בעברית

1. כללי

1.1 רשימת התגים למשימת זיהוי ותיוג שמות פרטיים :

<PER> - ביטוי שם אדם
<LOC> - ביטוי שם מקום
<ORG> - ביטוי שם ארגון
<TIME> - ביטוי זמן
<DATE> - ביטוי תאריך
<MONEY> - ביטוי כסף
<PERCENT> - ביטוי אחוזים
<MISC> - שונות

דוגמאות :

"ממשלת קלינטון"
ממשלת <PER> קלינטון <PER/>

"יבואני ארה"ב"
יבואני <LOC> ארה"ב <LOC/>

"מחשבי אפל"
מחשבי <ORG> אפל <ORG/>

"כנסת ישראל"
<ORG> כנסת ישראל <ORG/>

"יו"ר הליכוד, אריאל שרון, ..."
יו"ר <ORG> הליכוד <ORG/>, <PER> אריאל שרון <PER/>, ...

1.2 תחיליות יכללו בתיוג

"יו"ר הליכוד"
יו"ר <ORG> הליכוד <ORG/>

1.3 ניתן לפרק ביטויי סמיכות ולתת תיוגים שונים לכל חלק

"משפחת כהן"
משפחת <PER> כהן <PER/>

1.4 במקרים מסוימים ביטויים באורך מספר מילים יכילו תתי מחרוזות שיכולות לקבל תיוג שונה. במקרים כאלה לא יהיה תיוג נפרד לתתי המחרוזות.

"עוף הנגב"
<ORG> עוף הנגב <ORG/>
(אין תיוג נפרד ל"נגב" כשם מקום).

"פסטיבל עכו"
<MISC> פסטיבל עכו <MISC/>

"מכבי חיפה מוזילה כרטיסים"
<ORG> מכבי חיפה <ORG/> מוזילה כרטיסים

1.5 בביטויי שייכות תתויג כל תת מחרוזת.

"בית הספר לניהול של אוניברסיטת בן גוריון"
<ORG> בית הספר לניהול <ORG/> של <ORG> אוניברסיטת בן גוריון <ORG/>

1.6 כינויים וקיצורים (למשל: ארה"ב, אי.בי.אם., "התפוח הגדול" (כינוי לנוו יורק) יתויגו כשמות.

"הזמרת רוני סופרסטאר תופיע הערב"
הזמרת <PER> רוני סופרסטאר <PER/> תופיע הערב

"הבית הלבן מוסר..."
<ORG> הבית הלבן <ORG/> מוסר...

1.7 מרכאות יכללו בתיוג רק אם הן מופיעות באמצע השם.

"ויטו "הסנדק" קורליאונה"
<PER> ויטו "הסנדק" קורליאונה <PER/>

PER.2

2.1 קידומות כגון: מר, גברת, יו"ר, פרופ', רב לא יתויגו כחלק מהשם. בדומה עבור סיומות כגון ז"ל.

2.2 שמות משפחה יתויגו כ-PER.

"הפרידמנים ישבו לארוחת הצהריים"
<PER> הפרידמנים <PER/> ישבו לארוחת הצהריים

"משפחת כהן יצאה לטייל"
משפחת <PER> כהן <PER/> יצאה לטייל

2.3 שמות בעלי חיים יתויגו כ-PER.

"כלבו של ראש הממשלה, רקסי, נלקח לוטרניר"
כלבו של ראש הממשלה, <PER> רקסי <PER/>, נלקח לוטרניר

2.4 דמויות בדיוניות (גיבורי טלוויזיה, סרטים מצוירים וכו') יתויגו כ-PER.

"באטמן הוא דמות מפורסמת"
<PER> באטמן <PER/> הוא דמות מפורסמת

"הלכנו לראות את הסרט באטמן"
הלכנו לראות את הסרט <MISC> באטמן <MISC/>

"גרפילד אוהב לזניה"
<PER> גרפילד <PER/> אוהב לזניה

2.5 לא יתויגו כ-PER שמות הניתנים על-שם אדם : פסקי דין, מחלות, פרסים וכו' (למשל : "פרס נובל", "הוריקן מייקל" – ראה סעיף שונות).

LOC .3

3.1 מקומות : מקומות גיאוגרפיים, פוליטיים, אסטרונומיים, מדינות, יבשות, מחוזות, ערים, כפרים, שכונות, כבישים, שדות תעופה, גשרים, בסיסים צבאיים, רחובות (יכול לכלול גם מספר+רחוב), ימים, חופים, הרים, מבנים ("מגדל אייפל"...), תעלות, אגמים.

3.2 שם מקום יכול להיכלל בשם הארגון ולא יתויג בנפרד

"אוניברסיטת בן גוריון בנגב"

<ORG/> אוניברסיטת בן גוריון בנגב

"אפריקה ישראל להשקעות בע"מ היא חברת השקעות בינלאומית"

<ORG/> אפריקה ישראל להשקעות בע"מ היא חברת השקעות בינלאומית

3.3 בביטוי המורכב משמות מקומות עוקבים (מופרדים בפסיק או לא) יתויג כל מקום בנפרד.

"ירושלים, ישראל"

<LOC/> ירושלים , <LOC/> ישראל

"רח' ביאליק 114 באר שבע"

<LOC/> רח' ביאליק 114 באר שבע <LOC/>

3.4 שמות עצם :

"רחוב" "שדרות", "שכונה", "צומת" ההכרחיים לזיהוי המקום יתויגו כחלק מהביטוי ("עיר", "ישוב" וכו' לא יתויגו).

(אם ניתן להשמיט את שם העצם ושם המקום עדיין נשאר ברור, הוא אינו חלק מהביטוי)

"שדרות רוטשילד"

<LOC/> שדרות רוטשילד

"ר"ח שאול המלך"

<LOC/> ר"ח שאול המלך

"צומת גילת"

<LOC/> צומת גילת

"בסמוך לעיר באר שבע...."

בסמוך לעיר <LOC/> באר שבע....

"היישוב רהט"

<LOC/> רהט

ישנם שמות עצם שנהוג להצמידם לשם המקום (צורת דיבור) – גם אלה יכללו כחלק מהביטוי.

"קיבוץ עין גדי"

<LOC/> קיבוץ עין גדי <LOC/>

שמות עצם המזוהים עם שם המקום יתויגו יחד עם השם (חלק מהשם הרשמי, או צורת דיבור).

"ים המלח"

<LOC/> ים המלח <LOC/>

"עמק יזרעאל"

<LOC/> עמק יזרעאל <LOC/>

"נהר הירדן"

<LOC/> נהר <LOC/> הירדן <LOC/>

"כדור הארץ"

<LOC/> כדור הארץ <LOC/>

"ארץ ישראל"

<LOC/> ארץ ישראל <LOC/>

3.5 שמות כבישים יתויגו כביטוי אחד הכולל את שם העצם "כביש".

"כביש ירושלים תל-אביב"

<LOC/> כביש ירושלים תל-אביב <LOC/>

3.6 לא יתויגו שמות עצם המתארים את המקום

"חופי הים התיכון"

<LOC/> חופי הים התיכון <LOC/>

3.7 ביטויי כיוון (צפון, דרום וכו') יתויגו רק כאשר הם חלק מהשם הרשמי

"מזרח אירופה"

<LOC/> מזרח אירופה <LOC/>

"איזור הדרום"

איזור הדרום (ללא תיג)

"המזרח התיכון"

<LOC/> המזרח התיכון <LOC/> (כינוי)

"מרכז אסיה"

<LOC/> מרכז אסיה <LOC/>

3.8 שמות מקומות בצורת תואר השם לא יתויגו כשם מקום (ראה סעיף 9 - שונות).

"חיל האוויר האמריקאי"

<ORG/> חיל האוויר <ORG/> האמריקאי <MISC_AFF/>

"סטודנטים ישראלים"

<MISC_AFF/> סטודנטים <MISC_AFF/> ישראלים <MISC_AFF/>

3.9 כינויים וקיצורים יתויגו כשמות (ראה 1.6).

"מהומות בשטחים"
<LOC> מהומות <LOC/> בשטחים

"ישובי הבקעה"
ישובי <LOC> בשטחים <LOC/>
(קיצור לבקעת ירדן)

3.10 ארצות הברית ו-"ברית המועצות" יתויגו כשמות מקומות (לקוח מההנחיות באנגלית
.[1])

3.11 כאשר שם מקום מופיע לבד השם תמיד יתויג כשם מקום.

"ארה"ב תקפה את עיראק"
<LOC> ארה"ב <LOC/> תקפה את <LOC> עיראק <LOC/>

"במשחק הכדורגל בשבת סכנין ניצחה את בני יהודה"
במשחק הכדורגל <DATE> בשבת <DATE/> <LOC> סכנין <LOC/> ניצחה את
<ORG> בני יהודה <ORG/>

"ישראל תשחרר אסירים"
<LOC> ישראל <LOC/> תשחרר אסירים

"מדינת ישראל עומדת בפני הכרעה היסטורית"
<ORG> מדינת ישראל <ORG/> עומדת בפני הכרעה היסטורית

ORG .4

4.1 שמות ארגונים : עסקים, תחנות רדיו/טלוויזיה, עיתונים, ארגונים פוליטיים, מפלגות, דתות, קבוצות דתיות, ארגונים דתיים, בנקים, תזמורות, משרדים ממשלתיים, קבוצות ספורט, צבאות, חטיבות צבאיות, חיילות, ארגונים כלכליים, מלונות, אוניברסיטאות, בתי חולים, ועדות.

"ממשלת צפון קוריאה"
<ORG> ממשלת צפון קוריאה <ORG/>

"מכבי תל אביב"
<ORG> מכבי תל אביב <ORG/>

"מאה ושתיים אפ.אם."
<ORG> מאה ושתיים אפ.אם. <ORG/>

"ערוץ 2"
<ORG> ערוץ 2 <ORG/>

"נסדאק"
<ORG> נסדאק <ORG/>

"הפצועים פונו לבית החולים הדסה בירושלים"
<ORG> הפצועים פונו <ORG> לבית החולים הדסה <ORG/> <LOC> בירושלים <LOC/>

"אוניברסיטת תל-אביב"
<ORG/> אוניברסיטת תל-אביב <ORG/>

"משרד העבודה והרווחה"
<ORG/> משרד העבודה והרווחה <ORG/>

"אלוף פיקוד דרום"
<ORG/> אלוף פיקוד דרום <ORG/>

4.2 סיומות ארגונים (למשל "בע"מ") יתויגו כחלק מהשם.

4.3 בביטויים מורכבים תתויג כל תת מחרוזת (ראה 1.5).

"בית המשפט המחוזי בירושלים"
<ORG/> בית המשפט המחוזי <ORG/> <LOC/> בירושלים <LOC/>

"ועדת חוץ וביטחון של הכנסת"
<ORG/> ועדת חוץ וביטחון <ORG/> של <ORG/> הכנסת <ORG/>

4.4 שמות אירועים לא מתויגים כארגונים (ראה סעיף 9 - שונות), אולם ועדות וארגונות הקשורים אליהם יתויגו כשם ארגון.

"הועד האולימפי הישראלי"
<ORG/> הועד האולימפי הישראלי <ORG/>

4.5 במקרים בהם שם יצרן ומוצר מוזכרים, שם היצרן לא יתויג בנפרד (ראה סעיף 9 - שונות).

"פורד פוקוס"
<MISC/> פורד פוקוס <MISC/>

4.6 שמות כללים כגון "הממשלה", "הצבא", "המשטרה" לא יתויגו (לעומת "ממשלת ישראל", "צה"ל", "משטרת ישראל", שיתויגו כ-ORG).

4.7 שמות עצם יתויגו רק אם הם חלק מהשם הרשמי של הארגון.

"מפלגת הליכוד"
<ORG/> מפלגת הליכוד <ORG/>

"ערוץ 7"
<ORG/> ערוץ 7 <ORG/>

"הארגון הצבאי הלאומי"
<ORG/> הארגון הצבאי הלאומי <ORG/>

"ארגון החמאס"
<ORG/> ארגון החמאס <ORG/>

4.8 כינויים וקיצורים יתויגו כשמות (ראה 1.6).

"ההסתדרות תשבית את הנמלים"
<ORG/> ההסתדרות <ORG/> תשבית את הנמלים
(קיצור של הסתדרות העובדים)

"האוצר החליט על קיצוצים..."

<ORG/> האוצר <ORG/> החליט על קיצוצים...
(קיצור של משרד האוצר)

4.9 הביטוי "עם ישראל" יתויג כשם ארגון.

TIME .5

5.1 ביטוי מדויק של זמן: "10 דקות אחרי 5" (לעומת "מספר דקות אחרי 5"), "חצות" (לעומת "בוקר", "ערב").

"הכנס יתחיל בשעה 8:10"

<TIME/> 8:10 <TIME/> בשעה

"8:00 25/11/04"

<TIME/>8:00 <TIME/> <DATE/>25/11/04 <DATE/>

"המסיבה מתחילה בחצות"

<TIME/> בחצות <TIME/> המסיבה מתחילה

5.2 במקרים של ביטויי שייכות הביטוי יתויג כביטוי אחד

"ארבע לפנות בוקר"

<TIME/> ארבע לפנות בוקר <TIME/>

5.3 המילה "שעה" על כל צורותיה תתויג כחלק מהביטוי במידה והיא מופיעה בצמוד לביטוי שעה.

"קבענו לשעה 12:00"

<TIME/> 12:00 <TIME/> לשעה קבענו

5.4 אזורי זמן יתויגו כחלק מהביטוי

"GMT 5:00"

<TIME/> GMT 5:00 <TIME/>

"23:00 לפי זמן ישראל"

<TIME/> 23:00 לפי זמן ישראל <TIME/>

"בישראל, ביום שלישי בשעה שמונה..."

<LOC/> בישראל <LOC/>, <DATE/> ביום שלישי <DATE/>, <TIME/> בשעה
... <TIME/> שמונה

DATE .6

6.1 ביטוי תאריך מדייק: ציון יום מדויק ("ראשון"), תאריך מדויק ("1 בנובמבר" לעומת "ראש חודש"), עונה ("אביב"), שנה מדויקת ("2005" לעומת "שנה הבאה", "השנה"), עשור ("שנות ה-80" לעומת "עשר השנים האחרונות" "המאה ה-20" לעומת "המאה הנוכחית").

"ינואר 2005"

<DATE/> 2005 ינואר <DATE/>

"בסביבות ד' בתשרי"
<DATE/> בסביבות ה- <DATE> ד' בתשרי

"אוגוסט שנה שעברה"
<DATE/> אוגוסט <DATE/> שנה שעברה

"במהלך החורף השנה ירדו משקעים רבים"
<DATE/> החורף <DATE/> השנה ירדו משקעים רבים

"האביב הקרוב"
<DATE/> האביב <DATE/> הקרוב

6.2 שמות עצם לא יתויגו כחלק מהשם ("חודש", "עונה" וכו'), פרט למילה "יום" המופיעה בצמוד לציון אחד מימות השבוע.

"יום ג' הקרוב"
<DATE/> יום ג' <DATE/> הקרוב

"החגיגות יערכו במשך חודש יולי"
<DATE/> החגיגות יערכו במשך חודש <DATE/> יולי

6.3 במקרים של ביטויי שייכות מדויקים הביטוי יתויג כביטוי אחד

"רבע שלישי של 1990"
<DATE/> רבע שלישי של 1990 <DATE/>

"שלישי שעבר"
<DATE/> שלישי <DATE/> שעבר

6.4 ימים מיוחדים, חגים למשל, יתויגו כתאריך. במקרה של ביטוי תאריך המציין חג, שם העצם "חג" יתויג כחלק מהביטוי.

"ביום העצמאות האחרון התקיימה תהלוכה"
<DATE/> ביום העצמאות <DATE/> האחרון התקיימה תהלוכה

"כוננות גבוהה לקראת חג הפורים"
<DATE/> כוננות גבוהה לקראת <DATE/> חג הפורים

"יום הזיכרון לחללי צה"ל"
<DATE/> יום הזיכרון לחללי צה"ל <DATE/>

6.5 ביטויי תאריך בהתאם ללוחות השנה השונים (סיני, עברי וכו') יתויגו.

MONEY .7

דוגמאות :

"עשרים מיליון שקלים חדשים"
<MONEY/> עשרים מיליון שקלים חדשים <MONEY/>

"\$100"

<MONEY/> \$100 <MONEY>

"הפרויקט מוערך בשווי של מיליוני דולרים"
<MONEY/> מיליוני דולרים <MONEY> של

7.1 ביטויי כסף ביחידות נפרדות יתויגו בנפרד

"\$20 (90 ש"ח)"
<MONEY/> \$20 <MONEY> (<MONEY/> 90 ש"ח <MONEY>)

7.2 הערכות לא יתויגו (פרט לתחיליות-ראה סעיף 1.2)

"יותר ממאה שקלים"
<MONEY/> יותר <MONEY> ממאה שקלים <MONEY/>

"כמאה שקלים"
<MONEY/> כמאה שקלים <MONEY/>

7.3 כמתים יתויגו רק אם ניתן להחליפם במספר

"כמה אלפי דולרים"
<MONEY/> כמה אלפי דולרים <MONEY/>

PERCENT .8

8.1 המילה "אחוז" על כל צורותיה תתויג כחלק מהביטוי

"שלושת רבעי אחוז"
< PERCENT/> שלושת רבעי אחוז < PERCENT/>

"10%"
< PERCENT/> 10% < PERCENT/>

"התוצאות השתפרו משמונים אחוזים בשנה שעברה לתשעים השנה"
< PERCENT/> התוצאות השתפרו <PERCENT> משמונים אחוזים < PERCENT/>
< PERCENT/> בשנה שעברה <PERCENT> לתשעים < PERCENT/> השנה

"סיכויים של 50-50"
סיכויים של < PERCENT/> 50 < PERCENT/> - < PERCENT/> 50
<PERCENT/>
(המקף לא מתויג כחלק מהביטוי שכן הוא מפריד בין שני ביטויים שונים.)

"40-30 אחוזים"
< PERCENT/> 40 < PERCENT/> - < PERCENT/> 30 < PERCENT/>

8.2 המילה מינוס (סימן -) תתויג כחלק מהביטוי

"מינוס 10 אחוז"
< PERCENT/> מינוס 10 אחוז < PERCENT/>

8.3 הערכות לא יתויגו (פרט לתחיליות-ראה סעיף 1.2)

”בערך 70%“
< PERCENT /> 70% < PERCENT > בערך

”כ70%“
< PERCENT /> 70% כ < PERCENT >

MISC .9

9.1 ביטויים המציינים שמות אירועים יתויגו כשונות (<MISC_EVENT>). שמות עצם יתויגו כחלק מהשם רק כאשר הם חלק מהשם הרשמי.

”גביע אירופה לאלופות“
< MISC_EVENT /> גביע אירופה לאלופות < MISC_EVENT >

”פסטיבל ערד“
< MISC_EVENT /> פסטיבל ערד < MISC_EVENT >

”מלחמת המפרץ“
< MISC_EVENT /> מלחמת המפרץ < MISC_EVENT >

”רם וארליך העפילו לחצי גמר ווימבלדון“
< PER /> רם < PER /> < PER /> וארליך < PER /> העפילו לחצי גמר
< MISC_EVENT /> ווימבלדון < MISC_EVENT >

”האינתיפאדה אשמה בהפסדי המשק“
< MISC_EVENT /> האינתיפאדה < MISC_EVENT /> אשמה בהפסדי המשק

9.2 ביטויי השתייכות לקבוצה (למקום או לארגון) יתויגו כשונות (<MISC_AFF>) .

”אלפי מוסלמים משתתפים בתפילות בהר הבית“
אלפי < MISC_AFF /> מוסלמים < MISC_AFF /> משתתפים בתפילות < LOC > בהר
< LOC /> הבית

”הפרדה בין יהודים לערבים“
הפרדה בין < MISC_AFF /> יהודים < MISC_AFF /> < MISC_AFF /> לערבים
< MISC_AFF />

”עלייה במספר הפלשתינים התומכים בשלום“
עלייה במספר < MISC_AFF /> הפלשתינים < MISC_AFF /> התומכים בשלום

”המורדים הכורדים“
המורדים < MISC_AFF /> הכורדים < MISC_AFF />

”האיחוד האירופי“
< ORG > האיחוד האירופי < ORG >

”יורש העצר האוסטרי“
< MISC_AFF /> האוסטרי < MISC_AFF /> יורש העצר

"מאות חרדים הפגינו..."
מאות <MISC_AFF> חרדים <MISC_AFF/> הפגינו...

9.3 שמות סרטים, תכניות טלוויזיה, ספרים, שירים, דוחות, הסכמים, פסקי דין וחוקים יתויגו כשונות (<MISC_ENT>).

"לונדון את קירשנבאום"
<MISC_ENT> לונדון את קירשנבאום <MISC_ENT/>

"טולסטוי כתב את מלחמה ושלו"
<PER> טולסטוי <PER/> כתב את <MISC_ENT> מלחמה ושלו
<MISC_ENT/>

"בסיום הטקס שרנו את 'התקווה'"
' <MISC_ENT> התקווה <MISC_ENT/> ' <MISC_ENT/>

"דו"ח מצא"
"דו"ח מצא" (ללא תיוג)

"חוק פינוי פיצוי"
"חוק פינוי פיצוי"

"הסכם אוסלו"
"הסכם אוסלו"

"תכנית 'מפת הדרכים'
"תכנית 'מפת הדרכים'"