**IAHLT Guidelines for Named Entity Annotation**

**Introduction**

This document aims to serve as a draft proposal for entity classification, based on two schemas:

* [Hebrew Named Entity Annotation Guidelines](https://github.com/OnlpLab/NEMO-Corpus/blob/main/guidelines/nemo_guidelines_v4.pdf) / Dan Bareket, Dafna Amit, Zef Segal, Reut Tsarfaty
* [Entity Annotation Guidelines](https://github.com/nickyringland/nested_named_entities/blob/master/entity_annotation_guildlines.pdf) / Nicky Ringland

In addition, this proposal includes guidelines and explanations from Nemo Guidelines.

We suggest to use the following categories for the IAHLT annotation of Named Entities:

**Abbreviation  Meaning**

|  |  |
| --- | --- |
| person | PER |
| organization | ORG |
| location | LOC |
| geo-political entity | GPE |
| facility | FAC |
| event | EVE |
| product | DUC |
| work of art | WOA |
| language | ANG |
| Title | TTL |
| Time expression | TIMEX |
| Miscellaneous  Informal | MISC  INF |

The examples used in this document were taken from Hebrew Wikipedia entries as well as from Ynet (an online Israeli news magazine.)

**Categories**

**PER**

Pertains to names or nicknames (including Pseudonym and Aliases) of individual people, whether fictional or real-life. (fictional non-human characters will be tagged as MISC) For example:

1. אתר הכנסייה שימש כמקום ההתרחשות של כמה מהאירועים האחרונים בחיי [ישו]**PER**.
2. ב-[17 ביוני 855] **TIMEX** מת [לאו]**PER** ב[רומא]**GPE**, ובמקומו נבחר [בנדיקטוס השלישי]**PER**.
3. [החפץ חיים]**PER** היה ככל הנראה הראשון שכתב ספר הלכה ומוסר לחיילים.
4. [הקב"ר המזמר] **PER**: אמרו לי שמדינת [ישראל]**GPE** כבר לא קיימת.

Note: A figure in a work of art should be tagged PER, whereas the name of the art work should be tagged as WOA. For example:

1. הן בספר והן בסרטים, הדמות הראשית היא [סיביל איזבל דורסט]**PER**

לעומת: בשנת 1973 פורסם הספר [סיביל] **WOA**.

In case of an ellipsis, where two first names are followed by a single surname – the first will be annotated by its own; and the second will be treated as one unit with the last name, forming a single entity:

1. עדותו של [חפץ]**PER** נמשכת, כאשר באולם נוכחים בין היתר הנאשמים [שאול]**PER** ו[איריס אלוביץ']**PER**, בעלי השליטה לשעבר ב"[בזק]**ORG**" וב"[וואלה]**ORG**".

**ORG**

Pertains to a group of people organized for a particular purpose, such as: educational institutions, government agencies, business companies, media outlets, security bodies, political parties, government offices, courts/tribunals, ensembles, sport teams, NPO’s, families, dynasties, etc. We’ll also tag partial names. For example:

1. זהו החודש הרביעי ברציפות בו [אלטשולר שחם]**ORG** הוא הגוף המוסדי ממנו מנוידת כמות הכספים הגדולה ביותר.
2. [המנכ"ל]**TTL** של [פייזר]**ORG** אמר למעשה שמנה שלישית תספק הגנה כלשהי ממחלה קשה, גם בווריאנט החדש.
3. אז כשאנשים אוהבים להגיד שהתקשורת מתנכלת ל[הארי]**PER** ו[מייגן]**PER** אבל אני מאוד לא מסכים עם זה", אמר [דן ווטון]**PER**, עיתונאי של "[דיילי מייל]**ORG**".
4. [ראש הממשלה]**TTL** ליווה את בנו [דוד]**PER** לקבל את מנת החיסון הראשונה נגד [קורונה]**MISC** בקופת החולים [כללית]**ORG** ב[הרצליה]**GPE**.
5. זה שנים ארוכות ש[איגוד עובדי חסות הנוער]**ORG** ב[כוח לעובדים]**ORG** מתריע על קריסתה של [רשות חסות הנוער]**ORG**, שאמונה על הטיפול בבני נוער בסיכון.
6. לא ברור עד כמה החקיקה מרחיקת הלכת הזו ריאלית מבחינה פוליטית, שכן כדי לשנות את החוק ייאלץ [הורוביץ]**PER** לשכנע את עמיתיו בממשלה – לרבות במפלגות הימין ו[רע"מ]**ORG** – לתמוך במהלך.
7. ב-[1964]**TIMEX** הוקלטה בגרסה המקורית בביצוע מקהלת ה-[Basílica del Socorro]**ORG** בהדרכת [חסוס גבריאל סגאדה]**PER** ורביעיית הגברים "[לוס פרונטריסוס]**ORG**" ויצאה לאור על גבי תקליט.
8. גט ומסורבות עגונות למען הפועלת עמותה היא מבוי סתום ב[ישראל]**GPE**.
9. [ועדת הכספים של [הכנסת]**ORG**]**ORG**.
10. [האגף והקרן להכוונת חיילים משוחררים]**ORG** ב[משרד הביטחון]**ORG**.

Note: fictional organizations should also be tagged ORG:

1. [בילי]**PER** פורש מצוות [הפאוור ריינג'רס]**ORG** והופך לעוזר במרכז הבקרה.

Families referred to as a group should also be tagged ORG:

1. מלוות כל אחת בצוות ביוטי שדאג להן ללא הרף, קידמו [הקרדשיאנס]**ORG** את התוכנית החדשה בכיכובן‎.
2. [משפחת נתניהו]**ORG** עזבה הלילה את [מעון [ראש הממשלה]**TTL**]**FAC** ב[רחוב [**בלפורPER**[]**FAC**

Some organization names might contain other, nested, names, for example:

1. [הרבנות הראשית ל[ישראל]**GPE**]**ORG** היא המוסד הרבני העליון של מדינת [ישראל]**GPE**.
2. [עו"ד]**TTL** [אייל צור]**PER**, [יו"ר [העמותה ל[קרוהן]**MISC** ו[קוליטיס]**MISC** ב[ישראל]**GPE**]**ORG**]**TTL** מוסיף ואומר כי צריך להסתכל על המציאות.
3. ב-[2006]**TIMEX** שיחק שוב ב[אנסמבל [הרצליה]**GPE**]**ORG,** בהצגה "[אהבה וזעם]**WOA**" מאת [ג'ורג' פ. ווקר]**PER**, שביים [גדליה בסר]**PER**.
4. [בית הדין הצבאי ב[יפו]**GPE**]**ORG** גזר ב[יום ראשון]**TIMEX** מאסר של ארבעה חודשים וחצי והורדה לדרגת [טוראי]**TTL** על לוחם גדוד [נצח [יהודה]**PER**]**ORG** מ[חטיבת כפיר]**ORG**, [סמל]**TTL** [ש']**PER**, שהורשע במסגרת הסדר טיעון בהתעללות בעציר הפלסטיני [אחמד עארף מוחמד אבו עארה]**PER** (38) לפני כחודשיים באזור [רמאללה]**GPE**.

Religious institutions, higher education institutions, hospitals, hotels, museums and so on should be labeled ORG only when the reference is to the organization/management (otherwise, if the intention is for the building, they should be tagged FAC). For example:

1. [בריית]**PER** הפך לחבר כבוד של [האקדמיה המלכותית לאמנויות יפות ב[מינכן]**GPE**]**ORG**.
2. [שפטל]**PER** נולד בעיר [וילנה]**GPE** שב[רוסיה]**GPE** ([ליטא]**GPE**, לימים ב[פולין]**GPE**), בה הוא למד ב[ישיבת [רמיילס]**PER**]**ORG** אצל [הרב]**TTL** [הירש גרודזנסקי]**PER** אחיו של [הרב]**TTL** [חיים עוזר גרודזנסקי] **PER**, והיה בוגר המחזור הראשון של הגימנסיה העברית "[תרבות]**ORG**".
3. הוא למד מתמטיקה ב[אוניברסיטת [טומסק]**LOC**]**ORG**, עלה ל[ישראל]**GPE** ב[שנות ה-70 של המאה ה-20]**TIMEX** וקיבל את התואר מ[אוניברסיטת [בן-גוריון]**PER**]**ORG**.
4. ב [שנת 1913]**TIMEX** הפך [קוזמה]**PER** למרצה ב[בית הספר לעיצוב תעשייתי]**ORG** והקים את [סדנת [בודפשט]**GPE**]**ORG** שמטרתה הייתה לבצע רפורמה באמנות הדיור.
5. [הדסה [הר הצופים]**LOC**]**ORG** דיווח רק היום ל[משרד הבריאות]**ORG** על הזיהום בפגייה.
6. בנוסף לתפקידו כמוזיאון אנציקלופדי, שואף [מוזיאון [ישראל]**GPE**]**ORG** להישאר רלוונטי לזמן ולמקום אליו הוא משתייך.
7. בשל חוסר הוודאות - ב[אוצר]**ORG** מתנגדים לתקציב דו-שנתי עד [2024]**TIMEX**

Sports teams will be tagged ORG also when referred to by the GPE entity that they represent. Consider, for example, the two different uses of [סכנין] in this sentence:

1. לאור האירועים אמש מול [בית"ר [ירושלים]**GPE**]**ORG**, ההתנהלות של [סכנין]**ORG** במשחקים האחרונים והתפרעויות האוהדים, ב[המשטרה]**ORG** עשויים לקבל החלטה דרמטית.

לעומת: לאורך התקופה ההלניסטית, הרומית והביזנטית התקיימה במקום העיירה היהודית [סכנין]**GPE**.

This category also includes government departments or ministries. Thus, in mentions of ministers or secretary of departments, the ministry or the department is to be tagged ORG within the span of the title.

1. בין השנים [2004]**TIMEX**–[2008]**TIMEX** כיהן [סטוב]**PER** כחבר [הפרלמנט האירופי]**ORG**, בין [2008]**TIMEX** [2011]**TIMEX** כ[שר [החוץ]**ORG**]**TTL** של [פינלנד]**GPE**, ומ-[2011]**TIMEX** עד [2014]**TIMEX** כ[שר[לענייני [אירופה]**LOC** ולסחר חוץ]**ORG**] **TTL**.
2. [השרה לשוויון חברתי]**ORG**]**TTL**מביעה ב[אולפן [ynet]**ORG**]**FAC** התנגדות לחובת חיסונים.
3. תוקם וועדה בשיתוף משרדי [האוצר]**ORG**[הפנים]**ORG**ו[ראש הממשלה]**ORG**

"הממשלה” or “המשטרה” will not be tagged as “ORG”, as it’s not their name, but only a definite noun. Also, entities that appears in an informal way will not be tagged (“הצבא הישראלי”, “כוחות אסד”) as opposed to “הצבא הסורי” that it’s the entity’s partial name. Wikipedia and Google can be a good test

**LOC**

Places that are lacking a defined political component, such as: celestial objects, continents, geographical regions, waterbodies, straits/isthmuses, nature reserves, mountains/hills, mountain ranges, etc. For example:

1. אומנות שכזאת נפוצה ב[חצי האי ערב]**LOC**.
2. בישראל שישה יישובים אדומים: [גוש חלב]**GPE** ([ג'ש]**GPE**), [גבעת זאב]**GPE**, [קרני שומרון]**GPE**, [אחיהוד]**GPE**, [קצר א-סר]**GPE** ו[קשת]**GPE** שב[רמת הגולן]**LOC**.
3. [אי [חג המולד]**EVE**]**LOC** נמצא בין [אוסטרליה]**GPE** ל[אינדונזיה]**GPE** ומתגוררים בו כ-1,500 בני אדם.
4. בעבר הם חיו גם לאורך [אירואסיה]**LOC** ואף ב[אמריקה הצפונית]**LOC**.
5. כתבה אודות [רצועת [עזה]**GPE**]**LOC**

In some cases, the name of a geographically-defined entity also contains the name of a larger geographical area in which it is being located.

For example:

1. [טטוון]**GPE** ממוקמת על גדות הנהר [ויט]**LOC** במורדות [הרי ה[בלקן]**LOC**]**LOC**.

**GPE**

An entity who is defined by having a shared (1) ruling authority, (2) geographic location, and (3) population; including: nation states, counties, cities, towns, villages, regional councils,    
(cooperative) settlements, kibbutzim, neighborhoods, etc. For example:

We will not tag nations as “העם היהודי”\”האומה הפלסטינית”

1. ל[אירן]**GPE** שגרירות ב[בגדאד]**GPE**וארבע קונסוליות ב[בצרה]**GPE**, [סולימאניה]**GPE**, [ארביל]**GPE** ו[כרבלא]**GPE**.
2. הצגה המבוססת על הסרט, בבימויה של [דדי ברון]**PER**, הועלתה ב[מרץ 2011]**TIMEX** בתיאטרון של [דיסלדורף]**GPE**.

Note: appositions will not be tagged (in the example below the apposition is: המועצה האזורית)

1. [ראש המועצה האזורית [מטה אשר]**GPE**]**TTL**

Plural demonyms such as "The Russians", "the Americans", would be tagged GPE whenever they refer to the political entity or its government: the definite article should be within the span.

1. לאחר המלחמה הועבר [פרסטר]**PER** על ידי [האמריקנים]**GPE** ל[רוסים]**GPE**, והוא שהה בשבי הרוסי עד ל[שנת 1955]**TIMEX**.
2. [הנאצים]**GPE** כבשו את פולין.
3. השלמת קטע ראשון זה אפשרה העברה של אספקה לכוחות שתקפו את [הדיוויזה ה-18] היפנית, שהגנה בכוחות גדולים על השטחים שכבשו היפנים בצפון [בורמה]**GPE**, ובכללם הערים קאמאינג]**GPE**, מוגאונג]**GPE** ו[מיטקינה]**GPE**.

In some cases, the name of a geographically-defined entity also contains the name of a larger geographical area in which it is being located.

1. אנשי היישוב [רחלים]**GPE** קיוו להמשיך את הרצף הטריטוריאלי בין [גוש [שילה]**GPE**-[עלי]**GPE**]**GPE** לבין העיר [אריאל]**GPE**.

**FAC**

(Wo)man-made constructions, roughly generalized as objects within the responsibility of civil architecture; such as: buildings, stadiums, towers, prisons, museums, parking lots, airports, docking ports, roads/highways, train/bus stations, bridges, tunnels, streets, borders etc.  For example:

1. משאית התנגשה בגשר ליד [נתב"ג]**FAC**.
2. המהנדסים קבעו: [גשר המוגרבים]**FAC** בסכנת קריסה על [הכותל]**FAC**.
3. [ראש העיר]**TTL** [לאון]**PER**: "[טדי]**FAC** יהיה האצטדיון המוביל ב[ישראל]**GPE**".
4. בזמן הסגרים הפעיל מטבח רפאים בשם [פרייבט קיטשן]**FAC** שפעל באזור [פתח תקווה]**GPE**, [קריית אונו]**GPE** והסביבה עם תפריט מנות מצומצם.
5. במקביל הוא משיק ב[קניון הזהב]**FAC** מתחם ענק של [פקטורי 54]**ORG**.

We’ll also include names of ships/airplanes etc(when a name of an airplane model or other vehicel is mentioned we will tag it DUC

1. [אנולה גיי]**FAC** הוא מטוס [חיל האוויר האמריקאי]**ORG** מסוג [בואינג B-29 סופרפורטרס]**DUC** שממנו הוטלה פצצת האטום "[ילד קטן]**MISC** "
2. ב[[רחוב [אלנבי]**PER**]**FACי** 50]**FAC** - [54] **FAC**

The common noun will not be tagged. For example, if [קניון הזהב] is mentioned as just "הקניון" further down the road, the latter will not be annotated unless it is commonly used to refer to that specific entity, as is the case regarding the auditorium in Haifa, which is commonly referred to as "האודיטוריום"; or the Western Wall, which is known as "הכותל":

1. מדי שנה נערכים ב[האודיטוריום]**FAC** טקס הפתיחה וטקס הנעילה של [פסטיבל הסרטים הבינלאומי ב[חיפה]**GPE**]**EVE**.
2. לקראת סוף השבוע, במסגרת ההכנה לקראת [הפסח]**EVE** – אחד משלושת הרגלים שבהם נצטוו [ישראל]**GPE** לעלות לרגל – פונו הפתקים שהוטמנו בחודשים האחרונים בין אבני [הכותל]**FAC**.

Sometimes an entity could be both, FAC or ORG, we should tag be context:

1. [הדסה [הר הצופים]**LOC**]**FAC** הוא בית חולים הניצב על פסגת [הר הצופים]**LOC** ב[ירושלים]**GPE**.

לעומת: [הדסה [הר הצופים]**LOC**]**ORG** דיווח רק היום ל[משרד הבריאות]**ORG** על הזיהום בפגייה.

1. במתחם [מוזיאון [ישראל]**GPE**]**FAC** נמצא ביתה של אספנית האומנות והנדבנית [שרלוט ברגמן]**PER**.

לעומת: בנוסף לתפקידו כמוזיאון אנציקלופדי, שואף [מוזיאון [ישראל]**GPE**]**ORG** להישאר רלוונטי לזמן ולמקום אליו הוא משתייך.

Reminder: we will not tag appositions (in the example below the apposition is: בית החולים)

1. מתקפת סייבר על בית החולים "[הלל יפה]**FAC**" ב[חדרה]**GPE**.

Note: a neighborhood is considered a geopolitical entity, whereas streets should be treated as facilities. For example:

1. לאורך ה[מאלקון]**FAC** נבנו בתים ססגוניים בסגנון האר-דקו, שהעניקו לרחוב חזות המזכירה את רחובותיה של [מיאמי]**GPE**.

In cases of facilities that are named after a person, a geopolitical entity or a place it is necessary to tag also the nested entities. For example:

1. לפי סיפור זה, במקום בו התפלל [עומר]**PER** נבנה [מסגד [עומר]**PER**]**FAC**, השוכן עד היום ליד [כנסיית הקבר]**FAC**.
2. [תחנת הרכבת [[רעננה]**GPE** מערב]**FAC** נפתחה בטקס חגיגי בתחילת [יולי 2018]**TIMEX**.

**EVE**

Events that are known by a common, widely-used and/or official name, such as: sporting events, festivals, peace treaties, wars, battles, natural disasters, affairs, Holidays, commemorative dates, periodic events and so on. For example:

1. הוא הסתגל למקום בצורה יוצאת מן הכלל, אך נאלץ לחזור ל[טרינידד]**GPE** ב-1916 עקב פריצת [מלחמת העולם הראשונה]**EVE**.

Please make sure to tag also the internal entities in cases of events whose names contain other (nested) names, such as the name of the month/year in which they took place; or the place where they occurred. For example:

1. [שמחה רותם]**PER** היה אחד מהלוחמים הבולטים של [מרד [גטו [ורשה]**GPE**]**FAC**]**EVE**.
2. מבנה חשוב נוסף ברובע [ודאדו]**GPE**, על גבעה המשקיפה על ה[מאלקון]**FAC**, הוא [הוטל נסיונל]**FAC**, בית מלון מפואר שהוקם ב-1930 ושימש בעבר את עשירי התיירים האמריקאים שנהרו ל[הוואנה]**GPE** עד [המהפכה הקובנית]**EVE** ב-1959.
3. [הפיכת [18 בברימר]**TIMEX**]**EVE**, אף שאיש לא ידע זאת בשעתו, סימנה את קצה של [המהפכה הצרפתית]**EVE** בכלל.

1. בעונת [9–2008]**TIMEX** הם סיימו מחוץ לעשיריה הראשונה בשתי אליפויות אלו, אך בעונה האולימפית הם זכו להצלחה רבה, ובכלל זה מקום שביעי ב[אליפות [אירופה]**GPE**]**EVE** (הישג שיא עבורם) ומקום עשירי ב[אולימפיאדת [ונקובר**GPE**] **EVE** [**TIMEX** [2010].
2. [פרשת ילדי [תימן]**GPE**]**EVE**היא פרשת היעלמותם של פעוטות בני עולים חדשים, כשני שלישים מהם מיהדות [תימן]**GPE**, בין השנים [1948]**TIMEX**עד [1954]**TIMEX**

If the name of the holiday includes a name of a month\year, the month\year should be tagged as TIMEX

1. [תשעה ב[אב]**TIMEX**]**EVE** הוא יום תענית מדרבנן ושיא תקופת האבלות של [ימי בין המצרים]**EVE**
2. "[4 ב[יולי]**TIMEX**]**EVE** " הוא שם כינוי ל"[יום העצמאות האמריקאי]**EVE** "
3. המלונות חוזרים לתפוסה של לפני [הקורונה‎]**EVE**

לעומת: אילון מאסק חלה ב[קורונה]**MISC**

**DUC**

Any product that has a brand name, such as: electronic devices, vehicles, weapons, food products, medicines, applications, newspapers etc. For example:

1. [DB2]**DUC** הוא אחד מבסיסי הנתונים הנפוצים בעולם.
2. הסינכרון עם המחשב האישי מתבצע באמצעות כבל [USB]**DUC**.
3. סרטון ההדרכה של [צה"ל]**ORG** חושף כיצד פועלת התוכנה "[זאב כחול]**DUC**".
4. טיסת הבכורה של ה-[V8]**DUC** נערכה ב-9 ב[יולי]**TIMEX** 1961.
5. התרופה [PrEP]**DUC** מונעת הדבקה בנגיף ה-[HIV]**MISC** ומפחיתה משמעותית את שיעור ההדבקות.
6. עישנתי 25-20 סיגריות [מרלבורו]**DUC** ביום כששיחקתי.
7. הבובה החדשה, שכבר מכונה "[ברבי]**DUC** הנורמלית", כוללת עודף משקל קל של כמה קילוגרמים נוספים בעיקר סביב הירכיים, ואפילו צלוליטיס.
8. היעד של [טרה]**ORG** ב[שבועות]**TIMEX** הקרוב הוא להגדיל את המכירות של גבינת [טוב טעם]**DUC** פי 6 מהמכר שבשבועיים שלפני החג.

Note: many brands/companies have a product that it is the brand/company name, and in that case, we need to decide due to context if it's an ORG or DUC

1. [נתניהו]**PER**, שהחליט לצייץ על זה ב[טוויטר]**DUC**

לעומת: [טוויטר]**ORG**הודיעה כי תעביר ל[ביידן]**PER**את פרטי החשבון הרשמי של [נשיא [ארה"ב]**GPE**]**TTL**

1. היו פרסומים ב[הארץ]**DUC** שעומדת להיות חקירה נוספת.

לעומת: מעיתון [הארץ]**ORG** נמסרה תגובה.

1. התכנית משודרת ב[קשת]**DUC.**

לעומת: זו העונה ש[קשת]**ORG** מתכננת לנו.

**WOA**

Artworks, such as: book titles, songs/poems, drawings/paintings, sculptures, movies, TV shows, scientific works/papers, theories, etc. For example:

1. הספר פורסם בשנת 1973 ונקרא [סיביל]**WOA**.
2. סרטים מוזיקליים רבים, כמו "[אוקלהומה!]**WOA**", "[צלילי המוסיקה]**WOA**" ו"[שיער]**WOA**" הועברו ישירות ממחזות-זמר מצליחים.
3. השיר "[Mr. Lonely]**WOA**" המוכר מהסרט "[אסקימו לימון]**WOA**" יעורר בצופים רגשות אמפתיה כלפי הגיבור המאוהב והבודד.
4. [ברכת אלוהי נשמה]**WOA** היא אחת מברכות השחר הנאמרות בכל יום קודם תפילת [שחרית]**WOA**.
5. כך נולדו [הגאומטריות הלא אוקלידיות]**WOA** שאינן קשורות למציאות הפיזיקלית כפי שהיא נתפסת על ידי האדם.

The name of an artwork may also include internal entities. For example:

1. [טירופו של [המלך]**TTL** [ג'ורג']**PER**]**WOA** הוא סרט קולנוע בריטי שיצא ב]שנת 1994]**TIMEX**

**ANG**

Pertains to language names. For example:

1. איך אומרים את זה ב[הולנדית]**ANG**?
2. [אצ'בה]**PER** כתב את יצירותיו ב[אנגלית]**ANG** והוא החל לכתוב עוד בהיותו סטודנט.
3. [אליעזר בן יהודה]**PER**הוא מחייה [השפה העברית]**ANG**
4. הניב המצרי מדובר ב[מצרים] - לא מתוייג,

לעומת: שבטים אלו אינם דוברים [ערבית מצרית]**ANG**

**TTL**

Any type of honorifics, such as: military ranks, religious titles, academic titles, legislative titles, executive titles, aristocratic titles, etc.; whether referring to a real-life human-being, or a made-up character. We are not including adjectives in the TTL tag, but we will include specifically those: בדימוס, אמריטוס, במיל’(במילואים)

It should be noted that there is no nesting TTL in another TTL.

In Hebrew it was decided not to tag :אדון,גברת,מר

A title is to be tagged (TTL) only when it precedes or follows the name of the person who carries it - stand alone titles are not to be tagged, even when they have an "obvious" referent.

.

For example:

1. [בית הדין הצבאי ב[יפו]**GPE**]**ORG** גזר ב[יום ראשון]**TIMEX** מאסר של ארבעה חודשים וחצי והורדה לדרגת [טוראי]**TTL** על לוחם גדוד [נצח [יהודה]**PER**]**ORG** מ[חטיבת כפיר]**ORG**, [סמל]**TTL** [ש']**PER**, שהורשע במסגרת הסדר טיעון בהתעללות בעציר הפלסטיני [אחמד עארף מוחמד אבו עארה]**PER**.
2. [גוטמחר]**PER** השיגה דרגת [אמן בינלאומי]**TTL** לנשים ב[שנת 2004]**TIMEX**.
3. ב-[1416]**TIMEX** זכו אצילי המשפחה לקבל את התואר [דוכס]**TTL**.
4. ב[שנת 1860]**TIMEX** הוא הועלה לדרגת [סרג'נט מז'ור]**TTL**, וב-[1865] הוא הועלה לדרגת [לוטננט משנה]**TTL**.
5. [פרופסור]**TTL** [אליעזר רבינוביץ']**PER** יכהן בין שנה לשלוש שנים כנשיא המרכז הגדול בעולם לחקר מאיצי חלקיקים.
6. בהתבטאויות אלו הפר [הרב]**TTL** [יוסף]**PER** את הדין המשמעתי החל עליו.
7. לשיחה עם [נשיא [סין]**GPE**]**TTL** [שי ג'ינפינג]**PER**, התלוו אל [ביידן]**PER** [מזכיר המדינה]**TTL**, [אנתוני בלינקן]**PER** וכמה עוזרים.
8. "עמדת [ירדן]**GPE** ברורה מאוד, אנחנו אומרים לא מוחלט" אמר [המלך]**TTL** [עבדאללה]**PER** במהלך ביקור שהתקיים ב[עקבה]**GPE**.

99. [רה"מ]**TTL**  [יאיר לפיד] **PER** נפגש השבוע עם [ראש הממשלה]**TTL** הצרפתי [מקרון] **PER**

100. [ראש [עיריית [רחובות]**GPE**]**ORG**] **TTL** [רחמים מלול]**PER** התייחס הבוקר לפרשה.

1. תפנה הקבוצה לכיוונו של בלם [צפרירים [חולון] **GPE**] **ORG** ו[קפטן הנבחרת האולימפית] **TTL**, [טל חן]**PER**

**TIMEX**

Pertains to the following types of time expressions:

1. Specific dates that have a name or a numeral representation, such as: the days of the week, months of the year, a certain day in a month (7 במאי), seasons (not hours!). The tag will include the words “יום”,”שנה”,”שנת” etc.
2. Holidays, commemorative dates and periodic events will be tagged as EVE, and will contain TIMEX if they have a time expression (**EVE** [**TIMEX** חנוכה [2020]],   
   **EVE** [**TIMEX** אולימפיאדת טוקיו [2020]])
3. Words referring to a non specific part of the time unit, such as "beginning of (2018)", "late (October), will not be tagged within the span.

It should be noted that there is no nesting timex in another timex.

For example:

1. לאחר הפסקה של שנה בשל ה[קורונה]**MISC**, אחד הכנסים השנתיים הוותיקים ב[ישראל]**GPE** [יום העיון לצפרות]**EVE**, שנערך זו השנה ה-42 - יתקיים ב-[6 בדצמבר]**TIMEX** ([יום ב']**TIMEX**).
2. נתונים נוספים על אוכלוסיית [ישראל]**GPE**בתחילת [המאה ה-21]**TIMEX**
3. סוף [שנות ה-90]**TIMEX** היו...
4. הצילומים החלו ב [חודש יולי]**TIMEX** וצפויים להימשך עד [דצמבר]**TIMEX**.
5. ה[גרנד פרי]**EVE** הראשון נערך ב[טורקיה]**GPE** ב[אוגוסט 2005]**TIMEX** כמרוץ ה-14 בעונה.
6. היא הגיעה לחצי גמר [הגביע הספרדי]**EVE**.
7. במקרים רבים דווקא בעונת ה[חורף]**TIMEX**, גם במהלך גשמים שוטפים ורוחות עזות התקיימו התחרויות כסדרן.
8. ב-[1969]**TIMEX** ייצגה את [צרפת]**GPE** בתחרות ה[אירוויזיון]**EVE** שנערכה ב[מדריד]**GPE** עם השיר "[Un jour, unenfant]**WOA".**
9. [טקס פרסי [גלובוס הזהב]**MISC**]**EVE** הראשון נערך לראשונה ב-[1944]**TIMEX** ב[אולפני פוקס [המאה ה-20]**TIMEX**]**ORG**, ומאז הוא נערך במקומות שונים רבים.
10. [חוק יסודות התקציב]**MISC**, [התשמ"ה]**TIMEX** [1985]**TIMEX**.

Note: holidays, even though they occur regularly, are to be tagged EVE rather than TIMEX.

1. היעד של [טרה]**ORG** ב[שבועות]**EVE** הקרוב הוא להגדיל את המכירות של גבינת [טוב טעם]**DUC** פי 6 מהמכר שבשבועיים שלפני החג.
2. רק אתמול חגגו בני הזוג דעבול עם חברים את הדלקת [נר שלישי של [חנוכה]**EVE** ]**EVE** באווירה שמחה.
3. השירים המוקלטים הושמעו במסגרת שידורי התחנה ב[יום הזיכרון לחללי מערכות [ישראל]**GPE**]**EVE**.

1. ב[יום ו']**TIMEX**, ה[1 באפריל 2022]**TIMEX** חל "[יום הכזבים הבינלאומי]**EVE**"
2. ב[קיץ 2018]**TIMEX** היה חם מהרגיל.
3. ב[קיץ]**TIMEX** ב[שנת 2018]**TIMEX** היה חם מהרגיל.
4. בסוף [אוגוסט]**TIMEX** תחל ירידה בטמפרטורות.

**MISC**

Any type of entity that cannot be classified into any of the existing categories.

For example:

a. Diseases:

1. פאנל מומחי ה-[FDA]**ORG** המליץ לאשר את גלולת ה[קורונה]**MISC** של "[מרק]**ORG**" - עם שורת אזהרות.
2. לאחר שנתיים של מגפה ובעוד העולם טרוד מהתפשטות זן ה[אומיקרון]**MISC** – התרופה הראשונה ל[קורונה]**MISC** מגיעה.
3. [תסמונת אירוקנג'י]**MISC** נגרמת כתוצאה של מגע זרועות המדוזה בעור והחדרת הארס לגוף.
4. למרות הדמיון בשם, [הפרעת אישיות סכיזואידית]**MISC** איננה תת-סוג של [סכיזופרניה]**MISC**.
5. [עו"ד]**TTL** [אייל צור]**PER**, [יו"ר]**TTL** [העמותה ל[קרוהן]**MISC** ו[קוליטיס]**MISC** ב[ישראל]**GPE**]**ORG**, מוסיף ואומר כי צריך להסתכל על המציאות.
6. אחרי 2 אשפוזים ובירור רפואי אצל 3 מומחים נאמר לה שהיא חולה ב[פיברומיאלגיה]**MISC** ושאין מה לעשות עם זה  - היא צריכה ללמוד לחיות עם הכאבים.

b. Awards:

1. כמו כן קיבלה [פריקר]**PER** על התפקיד גם מועמדות לפרס [גלובוס הזהב]**MISC**.
2. הוא עיצב לה את השמלה ל[טקס פרס [אמי]**MISC** 2019]**EVE**.

c. Game names:

Note: we will tag MISC only for game names who do not have a brand name:

1. אם כל ילד ישראלי ילמד לשחק [שמחט]**MISC.**

לעומת: כל חשבון ב[מייפל סטורי]**DUC** יכול להתחתן במשחק פעם אחת.

d. Names of rules/laws:

1. חוק-יסוד: [כבוד האדם וחירותו]**MISC** הוא חוק יסוד, שנועד להגן על זכויות האדם העיקריות במדינת [ישראל]**GPE**.

e. Scientific names of animal/plant species:

1. יאם הוא שמם המשותף של כמה מיני צמחים בסוג [דיאוסקוריאה]**MISC**.
2. תפוצתו הטבעית של הנקר הייתה ביערות [מחטניים]**MISC**.
3. האדם הנבון המודרני נחשב לתת-המין היחיד ששרד עד ימינו מבין תת-המינים של [הומו ספיינס]**MISC**, מבין הסוג [הומו]**MISC** ומבין תת-השבט [הומינינים]**MISC**.

f. Zodiac signs:

1. הקנטאורים היו קשתים מעולים, ולכן דמותם מסמלת את מזל [קשת]**MISC** בגלגל המזלות.

g. Non human animates objects (animal’s name – pets, racehorses – or fictional characters who are not human:

1. דמות של כלב בשם [בלק] **MISC**

h. Ethnic groups:

Note: usually only the name of the people\tribe will be included in the span, with a few exceptions like in example 141.

1. מבחינה אתנית הוא משתייך לשבט [איגבו]**MISC**.
2. ישלמו את המחיר ילדיהם ו[עם ישראל]**MISC**.

i. Geographical lines:

1. קו הגבול בין ישראל לירדן היה חלק מקו הגבול המכונה "[הקו הירוק]MISC."

j. Flight numbers\ Train or bus lines:

1. נוסעי [טיסת 002]**MISC**

**INF**

This tag is only an added tag to information that is not informal. In the case of tagging something informal we will have two tags on the same span. For example:

1. אתם בארץ [ישראל] אבל במדינת [[ישרהל]]**GPE-INF**!

**General notes on tagging borders:**

* If the phrase includes an adposition, it should be included in the tagging (if it’s a part of the entity’s name) or be left out of it (otherwise).

1. [מוזיאון [ישראל] **GPE**] **FAC**ב[ירושלים] **GPE**
2. [הקרן החדשה ל[ישראל] **GPE**] **ORG**
3. [ועדת הכספים של [הכנסת]**ORG**]**ORG**
4. [מנהלת [אגף הכספים]**ORG**]**TTL** ב[נעמ"ת]**ORG**]

* Ellipsis: tag only if the phrase is legitimate.

1. הרשויות ל[הגנת הפרטיות] **ORG** ו[הגנת הצרכן] **ORG** ( "הרשויות להגנת הפרטיות" is not coherent)
2. משרדי [האוצר] **ORG** ו[הרווחה] **ORG** הם האחראיים על הרפורמה ( [משרד האוצר] opposed to as)
3. [שרה]**PER** ו[בנימין נתניהו] **PER**.
4. הקוריאות [הצפונית] **GPE** ו[הדרומית] **GPE**

* Include in span the word that crucial for understanding the specific entity’s name. A good way of determination – if we are not including the word, will the meaning stay the same?

1. העיר [חיפה] **GPE** היא העיר השלישית בגודלה ב[ישראל] **GPE**
2. [אוניברסיטת [תל אביב] **GPE**] **FAC** נמצאת בצפון [תל אביב] **GPE**