Contents

[Introduction 2](#_Toc295208168)

[1. Named Entity Simple Annotator User Manual 3](#_Toc295208169)

[1.1. Annotator Manual 3](#_Toc295208170)

[1.1.1. Mark-up borders 5](#_Toc295208171)

[1.1.2. Automatic repositioning of mark-up borders 5](#_Toc295208172)

[1.1.3. Finishing the mark-up of NE’s 6](#_Toc295208173)

[1.2. Annotator – Disambiguator Manual 7](#_Toc295208174)

[2. NE category mark-up Guidelines 12](#_Toc295208175)

[2.1. Multi-Name and Multi-Modifier Expressions 12](#_Toc295208176)

[2.2. Numeric and Time Expressions 12](#_Toc295208177)

[2.3. Nested expressions 12](#_Toc295208178)

[2.4. Entity-Expressions that Modify Non-entities 12](#_Toc295208179)

[2.5. Entity-Expressions that "Possess" Other Entity-Expressions 13](#_Toc295208180)

[2.6. Quotation Mark Mark-up 13](#_Toc295208181)

[2.7. Specific Cases for Organizations 13](#_Toc295208182)

[2.8. Specific Cases for Products 14](#_Toc295208183)

[2.9. Specific Cases for Locations 15](#_Toc295208184)

[2.10. Specific Cases for Persons 16](#_Toc295208185)

[2.11. Specific Cases for Date and Time 17](#_Toc295208186)

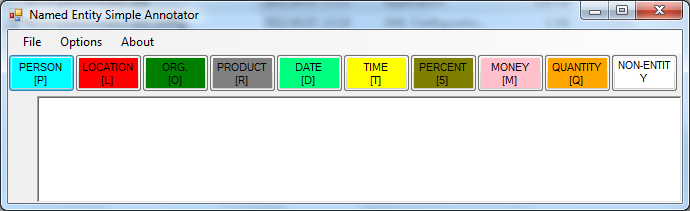
[2.12. Specific cases for Money, Percent and Quantity 19](#_Toc295208187)

Introduction

The document „” is created in order to specify Named Entity (NE) mark-up guidelines using the tool “***NESimpleAnnotator.exe***”. The guidelines specify how to process ambiguous or arguable NE border situations, in order to achieve identical mark-up results by all participating annotators. The document is intended for Estonian, Latvian and Lithuanian languages, but it can be used for any other language (which characters are supported by UTF-8) as well. The document features refined samples and explanations from the MUC-7 Named Entity task, as well as some examples in Latvian.

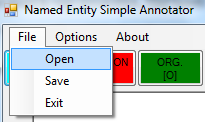
1. Named Entity Simple Annotator User Manual
   * 1. Annotator Manual

In order to mark a NE corpus, the user has to **get acquainted with the mark-up guidelines from the section 2** and then use the provided tool “*NESimpleAnnotator*”. The NESimpleAnnotator allows a user to read plaintext documents and mark them with nine NE tags (person, location, organization, product, date, time, percent, money and quantity). The application is shown in Fig. 1.1.1.



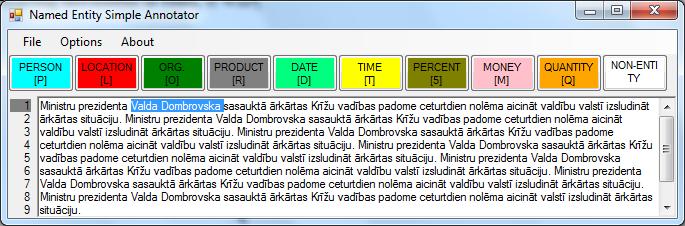
* + - 1. NESimpleAnnotator graphical user interface (GUI).

To start the mark-up process, the user has to open a plaintext document. This is done by selecting “Open” from the “File” submenu.



* + - 1. “Open” file submenu option.

When an existing document is read, its contents will be shown in the text area (see Fig. 1.1.3). If the document will contain any NE mark-up (following the MUC-7 mark-up guidelines), the existing mark-up will be preserved and marked according to corresponding NE types.



* + - 1. Opened plaintext document.

The user is given two possible ways to mark NE’s in the opened document. First of all, the user has to select the named entity (as shown in Fig. 1.1.5 and then mark the NE according to its type. The marking is done either by pressing the correct button on top of the text (as shown in Fig. 1.1.4) or by pressing the corresponding keyboard text.



* + - 1. NE mark-up buttons.

The possible keyboard combinations for NE mark-up are:

* P – Person;
* L – Location;
* O – Organization;
* R – Product;
* D – Date;
* T – Time;
* 5 – Percent;
* M – Money;
* Q – Quantity;
* “-”, “Space”, “Esc” or “Backspace” – Non-entity.

Note that for this mark-up task “Percent” and “Quantity” have been disabled, therefore, these two NE types should not be marked.



* + - 1. NE selected for mark-up.

Once the NE has been selected and an appropriated mark-up has been applied (by pressing the correct button or key), the NE is marked in a corresponding color (as can be seen in Fig. 1.1.6).



* + - 1. NE marked as a person.
         1. Mark-up borders

**The user should not mark NE’s partially**. The user has to mark the whole NE as a single named entity, for instance, the person “Valdis Dombrovskis” should not be marked as “Valdis” and “Dombrovskis” (see Fig. 1.1.1.1. below). This would cause the NE to be divided into two separate named entities, which is wrong (Valdis Dombrovskis, for instance, is one person and not two separate persons). The correct mark-up is shown in Fig. 1.1.6.



NE borders marked wrongly (example 1).

Another wrong (maybe even accidental) NE mark-up could be mark-up of a longer than necessary NE fragment (for an example see Fig. 1.1.1.2). Such a situation can either be fixed by repeated mark-up of the correct NE (shown in Fig. 1.1.1.3) or by marking the whole NE as a non-entity (shown in Fig. 1.1.1.4).



NE borders marked wrongly (example 2).



Wrong mark-up fixed (example 1).

To remove mark-up from a wrong NE, the user may also position the text editing cursor “|” within the NE and trigger (press the appropriate key) non-entity mark-up.



Wrong mark-up fixed (example 2).

* + - * 1. Automatic repositioning of mark-up borders

The *NESimpleAnnotator* repositions the borders of the users selection if it includes whitespaces (whitespaces at ends of a NE are not allowed), therefore, the mark-up can be done faster. Examples of a correct selection and the corresponding auto-trimming are shown in Fig. 1.1.2.1.

🡪

🡪

Correct selection leads to correct NE borders.

If, though, the user selects the NE wrong, trimming may not remove all wrong characters. For instance, punctuation marks are not trimmed (as shown in Fig. 1.1.2.2). It is, therefore, up to the user to make a correct selection, not including false characters (other than whitespaces).

🡪

Bad selection leads to wrong NE borders.

The system does not trim punctuations because some NE’s require punctuations as a part of the NE’s, for instance dates.

Another automatic repositioning is the expansion of the user’s selection, if only partial NE is selected. The Fig. 1.1.2.3 shows some examples of the expansion when the text positioning cursor is placed inside or before a NE and when the NE is selected only partially.

🡪

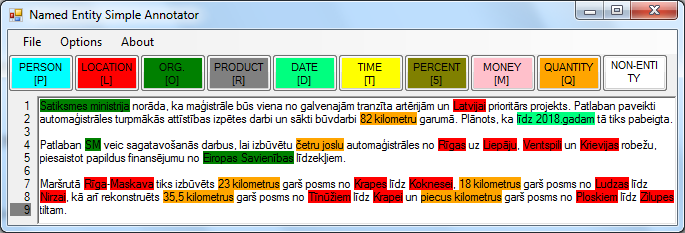
🡪

🡪

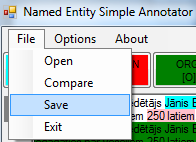
Expansion of a partially selected single token (word) NE.

* + - * 1. Finishing the mark-up of NE’s

Once all named entities are selected (for an example see Fig. 1.1.3.1), the document has to be saved. To save the document, the user has to select the “*Save*” option from the “*File*” submenu (see Fig. 1.1.3.2).

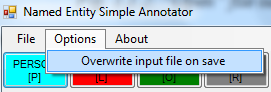


Completed mark-up example of a document.



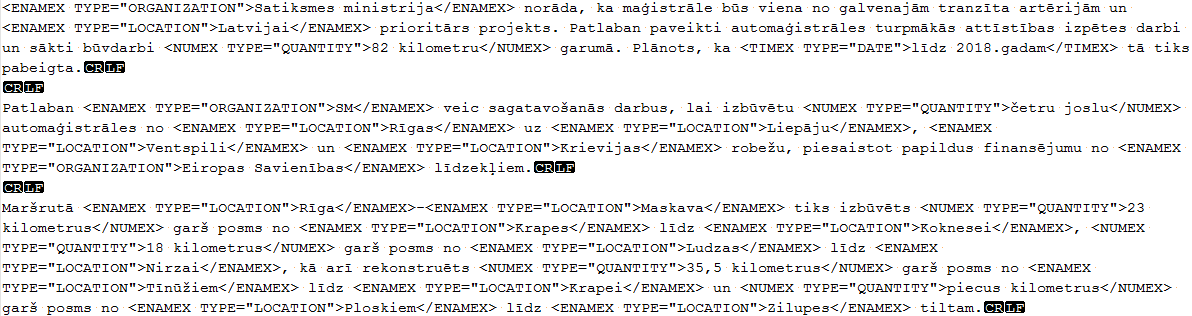
“Save” file submenu option.

If the user wants to save the data back to the input document (which makes the mark-up process faster when many documents have to be marked), the user may select the “*Overwrite input file on save*” submenu option. The selection will create a tick () left to the option to indicate that it is selected. The choice will be saved when the application will be closed and the user won’t have to select the choice again.



“Overwrite input file on save” submenu option.

If the “*Overwrite input file on save*” option will not be selected, the user will have to save the document by specifying an output file in the file system. An example of a marked text is shown in Fig. 1.1.3.4.

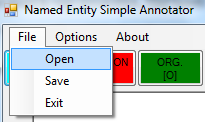


Text marked for NE’s.

* + 1. Annotator – Disambiguator Manual

When two annotators have completed annotation of a document, the document mark-up has to be disambiguated and checked for false annotations. This task is done by a third annotator.

To start the disambiguation procedure, the annotator has to **get acquainted with the mark-up guidelines from section 2** and then open the first document by selecting “*Open*” from the “*File*” submenu (see Fig. 1.2.1). The order of opened documents is not important if the user has not selected the “*Overwrite input file on save*” submenu option (see Fig. 1.1.3.3). If the user has selected the overwrite feature, the user must make sure that the **first document is always selected from the same directory** (or from the same previous annotator). **Otherwise, on save the files processed files will be mixed up!** It is not advised to use the overwrite feature by the third annotator as he/she may want to re-compare the annotation!



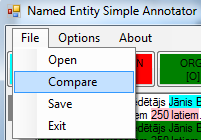
* + - 1. “Open” file submenu option.

Once the document is opened, the mark-up from the first annotator will be shown (see Fig. 1.2.2).



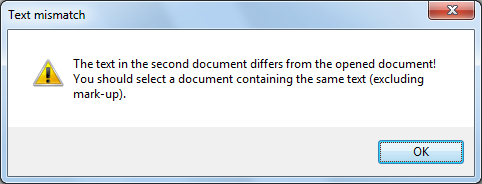
* + - 1. NESimpleAnnotator after opening the first annotator’s document.

The second step is to compare the mark-up of the first annotator to the second annotator’s mark-up. To do this the user has to select “*Compare*” from the “*File*” submenu (see Fig. 1.2.3) and select the second annotator’s document.



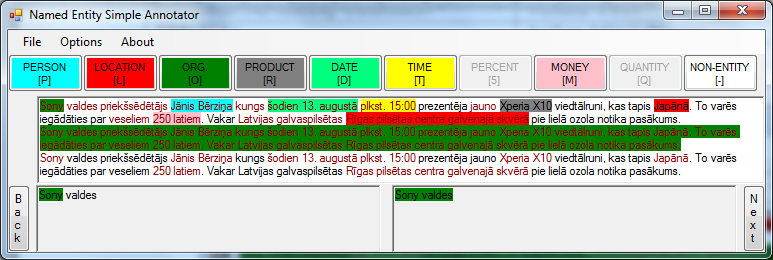
* + - 1. “Compare” file submenu option.

If the user accidentally will select a wrong document, the program will show an exception if the text without mark-ups won’t be identical in both documents (See Fig. 1.2.4).



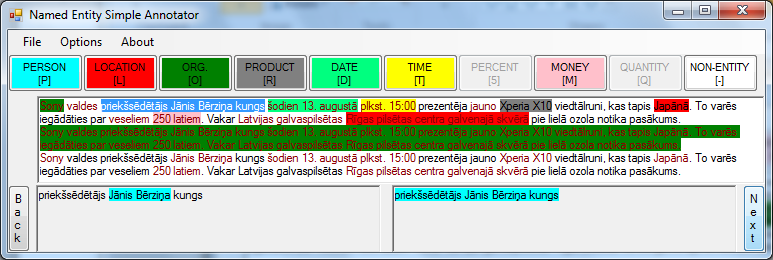
* + - 1. Text mismatch error message.

If such an exception occurs, the user has to select “*Compare*” again (as long as the correct document is not selected). Once the correct document is selected, the application will automatically analyze the mark-ups and highlight with darker red mark-up mismatching positions (see Fig. 1.2.5) and open the comparison toolbar below the text.

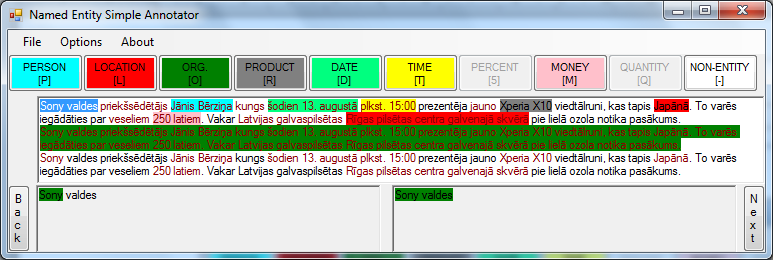


* + - 1. Comparison of two marked documents.

The comparison toolbar consists of two buttons (“*Next*” – to move to the next difference (see Fig. 1.2.6) and “*Back*” – to move to the previous difference (see Fig. 1.2.7)) and two text fields (one to the left – representing the first annotator’s annotation; one to the right – representing the second annotators annotation).



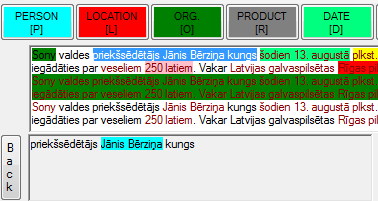
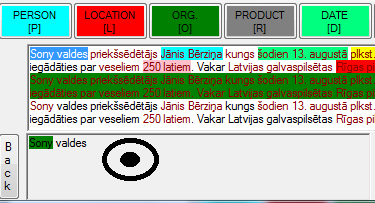
* + - 1. Button “Next” shows the next difference.



* + - 1. Button “Back” shows the previous difference.

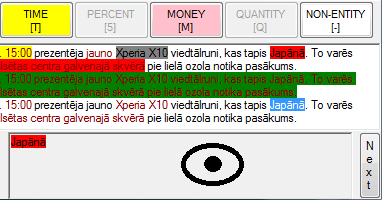
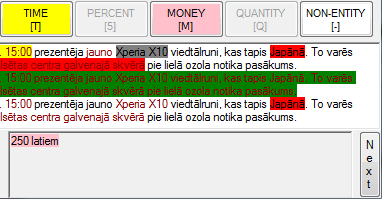
The task of the third annotator is to select the correct annotation in all difference sections, but not only, as the third annotator can also re-annotate a position, if both annotators have made a mistake in the annotation or a text position lacks an annotation at all. **The re-annotation is performed following the guidelines from the section 1.1.** When re-annotation is done within a difference section (the dark red colored text), the difference is removed.

If the **first annotator’s annotation is correct**, the third annotator has to **double click on the left text field** as shown in Fig. 1.2.8 or press the keyboard **key “1”**.



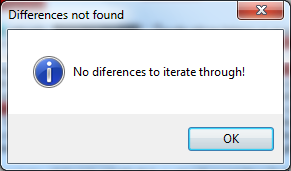
* + - 1. Selecting the first annotator’s annotation by double-clicking the left text field.

If the **second annotator’s annotation is correct**, the third annotator has to d**ouble click on the right text field** as shown in Fig. 1.2.9 or press the keyboard **key “2”**.



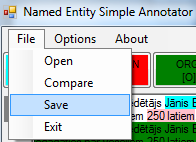
* + - 1. Selecting the second annotator’s annotation by double-clicking the right text field.

If all differences will be processed, the user will receive a message as shown in Fig. 1.2.10.



* + - 1. No differences to iterate through.

Once the differences, false annotations and missing annotations have been resolved, the document has to be saved. To save the document, select “*Save*” from the “*File*” submenu (see Fig. 1.2.11) and choose where to save the file (press “*Save*” when done). Make sure you do not overwrite the initial documents as you could want to re-compare/ change your judgment.



* + - 1. “Save” file submenu option.

To start a new comparison, simply open a new document as shown in the beginning of section 1.2.

1. NE category mark-up Guidelines

The guidelines are based (mostly copied, but adapted and with some ambiguities and contradictions resolved) on MUC-7 NE annotation guidelines[[1]](#footnote-1).

* + 1. Multi-Name and Multi-Modifier Expressions

“North and South America” (LOCATION)

“U.S. Fish and Wildlife Service” (ORGANIZATION)

“Latvijas Republikas Finanšu un Veselības ministrijas” (ORGANIZATION)

“Latvijas Vides, ģeoloģijas un meteoroloģijas centrs”(ORGANIZATION)

Conjoined multi-name expressions (expressions, which are combined of multiple entities) and single entities with conjunctions have to be treated as one entity.

"10- and 20-dollar bills"

“5 un 10 latu naudas zīmes”

* + 1. Numeric and Time Expressions

“175 to 180 million Canadian dollars” (MONEY)

“projekta izmaksas lēstas no 175. līdz 180. miljoniem igauņu kronu.” (MONEY)

“the 1986-87 academic year” (DATE)

“2010./2011. akadēmiskais gads” (DATE)

“from 1990 through 1992” (DATE)

“līdz pat 2011. gada 17. janvārim” (DATE)

The subparts of time, date, money, quantity and percentage range expressions should be marked up as parts of a single expression, even if there is no elision of the numeric "units".

* + 1. Nested expressions

“8:24 a.m. Chicago time” and not “8:24 a.m. Chicago time”

“U.S. $10 million” and not “U.S. $10 million”

“the U.S. Customs Service” and not “the U.S. Customs Service

Nested expressions are not to be tagged as separate NE’s.

“Arthur Anderson Consulting”

“SIA Jānis Bērziņš un partneri”

In some cases, multi-word strings that are proper names will contain entity name substrings; such strings are not decomposable; therefore, the substrings are not to be tagged separately.

* + 1. Entity-Expressions that Modify Non-entities

“IK Baudas virsotne piedalījās conferencē” (ORGANIZATION)

“Gospel Singers Ltd. joined the choir.” (ORGANIZATION)

“Gospel singers joined the choir.” (NON-ENTITY)

“Gaiziņkalna virsotne nav visai stāva” (LOCATION)

“Uzņēmumā Apavu dakteris Jūs varat …” (ORGANIZATION)

“Apavu dakteris salaboja manas kurpes” (NON-ENTITY)

“Bērziņu ģimene …” (PERSON)

“Strawberry family …” (PERSON)

“Strawberry fruits are tasty.” (NON-ENTITY)

Entity names used as modifiers are to be tagged only when it is clear to the annotator from context or the annotator's knowledge of the world that the name is that of an organization, person, or location.

* + 1. Entity-Expressions that "Possess" Other Entity-Expressions

“Temple University's Graduate School of Business” (ORGANIZATION)

“Shearson Lehman Hutton's OTC department” (ORGANIZATION)

“California's Silicon Valley” (LOCATION)

“Canada's Parliament” (LOCATION, ORGANIZATION)

“Latvijas Republikas Saeima” (ORGANIZATION)

In a possessive construction, the possessor and possessed LOCATION, ORGANIZATION, PERSON and PRODUCT substrings should be tagged as one entity if not separated by non-entity words (“Latvijas Republikas likumdevējs Saeima” or “Canada’s legislator the Parliament”).

* + 1. Quotation Mark Mark-up

“Vito “The Godfather” Corleone” (LOCATION)

“SIA “Jūras vārti”” (LOCATION)

Quotation marks have to be included in the NE mark-up if they are included within the NE. Otherwise do not tag quotation marks!

“The movie “Avatar” was not shown yesterday.” (PRODUCT, DATE)

* + 1. Specific Cases for Organizations

“SIA Tilde” (ORGANIZATION)

“Bridgestone Sports Co.” (ORGANIZATION)

“The LTV1 program is great” (ORGANIZATION)

“I was watching BBC yesterday” (ORGANIZATION, DATE)

Corporate designators such as "Co.", “Ltd.”, “Inc.”, etc. are considered part of an organization name.

The ORGANIZATION NE’s include **companies, stock exchanges, multinational organizations** (such as “EU”, “European Union”)**, political parties, orchestras, unions, non-generic** (“police”, “hospital” is a generic entity and should not be tagged.) **governmental entity names** such as “Congress” or “Chamber of Deputies”, **sports teams and armies, churches, embassies, factories, hospitals** (“General Hospital”, “Gaiļezers”), **hotels, museums, universities, TV channels**.

In expressions where organizations may be used as products (metonymic expressions), tag the expression as organization (“I have an LG; what phone have you got?”, “I was watching BBC1 yesterday.”).

“the Pan-American Games” (Non-Entity)

“the U.S. Olympic Committee” (ORGANIZATION)

“China Film Festival” (LOCATION)

**Do not tag event names** (sports, concerts, festivals – none shall be tagged), even if they refer to events that occur on a regular basis and are associated with institutional structures. However, the institutional structures themselves - steering committees, etc. - should be tagged.

“US government” (LOCATION, NON-ENTITY)

“US Air Force” (ORGANIZATION)

Do tag governmental institutions **only if such really (officially) exist**. If the expression is only a general indication towards an institution, do not tag such instances. If the expression contains only common words, make a strong decision, whether the expression may also be a modifier or a common word expression. If “yes” then do not tag such instances as organizations.

* + 1. Specific Cases for Products

“Ford Fiesta ST” (PRODUCT)

“Samsung Galaxy Tab” (PRODUCT)

“Samsung’s newest invention the Galaxy Tab” (ORGANIZATION, PRODUCT)

“Apple iPhone” (PRODUCT)

“Microsoft Windows XP Professional” (PRODUCT)

“Laimas konfektes “Serenāde”” (ORGANIZATION, PRODUCT)

“Microsoft jaunā Windows 7 operētājsistēma” (ORGANIZATION, PRODUCT)

“Espresso coffee” (NON-ENTITY)

“Tematiskā paka” (NON-ENTITY)

“I also buy the thematic package” (NON-ENTITY)

Tag company names together with their products. Only if it is absolutely clear that the product does not include the company name in it, tag the company name separately. If the company name is separated with the product by non-entity words, tag the company as ORGANIZATION and the product as PRODUCT.

Do not tag general expressions that are widely used as non-entities as products. For instance, “coffee”, “milk”, “car”, “internet connection” all are considered general expressions. Only if the expression is capitalized (“Our newest product the “Espresso Coffee” is …”, “Lattelecom piedāvājums – “Tematiskā paka” tika …”) and emphasized (by quotes, for instance) tag the expression as a product. Overgeneralized product names (lowercase) create a lot of noise in the data, therefore, do consider, whether the expression in the context of a sentence may also refer to a general object and not only to a specific company’s product. If there is doubt that the expression may be a general expression, do not tag the expression as a product.

**Do not tag expressions as products if they are not capitalized** and include only generic names in the product part (ignoring the optional company name). For instance, “Lattelecom tematiskā paka ir lieliska” or “Lattelecom thematic pack is awesome”. Non capitalized product names create noise in the training data as these might actually refer to a general expression and the company name could be substituted with something else, for instance “The game’s thematic pack is awesome” (which does not include a specific product).

“DVD, PC, VGA, WUXGA, LCD, HDD, SSD, HD, SD, XD” (NON-ENTITY)

Do not tag common technical abbreviations as products, nor organizations as these do not refer to concrete products (as we do not tag “bread”, “maize” as well, tagging these would create noise in the data). Only tag such expressions if it is clear from the context that these refer to a concrete/unique product or are subpart of a product name.

* + 1. Specific Cases for Locations

“Germany invaded Poland in 1939.” (LOCATION, DATE)

Metonyms that reference political, military, athletic, and other organizations by the name of a city, country, or other associated location should be tagged as LOCATION.

The **LOCATION** NE’s include **named heavenly bodies** (stars, constellations, etc.), **continents, multi-country sub-continental regions** (for instance, “Eastern Europe”), **multi-country trans-continental regions** (“the Middle East”) **countries, sub-national regions** (“Amazonia”), **provinces, counties, cities, regions, districts, towns, villages, neighborhoods, airports, highways, street names, street addresses, oceans, seas, straits, bays, channels, sounds, rivers, islands, lakes, national parks, mountains, fictional or mythical locations, and monumental structures**, such as the Eiffel Tower and Washington Monument, that were built primarily as monuments.

Do not tag general location modifiers or historic time modifiers, such as, “South”, “West”, “Upper”, “Former”, “Ancient” when these are defined without a definite location attached or they are used as ad hoc modifiers that are readily separable from the name (for instance, do tag “West London”, “Ancient Rome”, but do not tag “upper section of Daugava”).

“University of Latvia in Riga” (ORGANIZATION, LOCATION)

“Latvijas Universitāte Rīgā” (ORGANIZATION, LOCATION)

“Rīga, Latvija, Brīvības gatve 50a, LV-1001” (LOCATION)

“10 Downing Street, London, UK, 3BX CB2” (LOCATION)

Compound expressions in which place names are separated by a comma are to be tagged as separate instances of LOCATION. In the case if locations are defined only by numeric indexes (with an exception to the postal code) (for instance “10, 23, 10, 34 and 5 Downing street”) do not tag the indexes as these will create noise in the data. Tag only the last index, which is combined with the location.

“Mississippi River” (LOCATION)

“Valka district” (LOCATION)

“In Valka’s special district” (LOCATION)

“Valkas īpašajā rajonā” (LOCATION)

“Valka region” (LOCATION)

“Riga City” (LOCATION)

“Rīgas pilsēta” (LOCATION)

“Velna ala Salacgrīvā” (LOCATION)

“Alūksnes ezers”

“… such as Lake Baikal and …” (LOCATION)

Designators that are integrally associated with a place name are to be tagged as part of the name. For example, include in the tagged string the word "River" in the name of a river, "Mountain" in the name of a mountain, "City" (“pilsēta”) in the name of a city, etc., if such words are contained in the string.

“Brīvības piemineklis” (LOCATION)

“Turaidas pils” (LOCATION)

“Freedom monument” (LOCATION)

“Marianburg fortress” (LOCATION)

**Sightseeing objects, landmarks** are to be tagged as locations.

Tag an expression as a location only if it contains a proper noun **and is capitalized** (at least the first word, for instance, in Latvian). Do not tag common words as locations (for instance, “mājā”, “in the house”, “here”, “there”, “in the North”, “up”, “down”, “first floor”, “basement”, “in the fridge”, “gravā”, “ielejā”, “in the valley” – an expression given in the locative case does not necessarily mean that it is a LOCATION NE).

Do not tag expressions as locations if the proper noun can also be a common word and is misspelled with lowercase letters (“šodien saulē satikās ministri”, “Today I visited new York” – “new” not tagged as location as it is a common word and may also be used only as a modifier of York).

* + 1. Specific Cases for Persons

“Mr. Harry Schearer” (PERSON)

“Dr. Jānis Bērziņš” (PERSON)

“Mr. President” (NON-ENTITY)

“Secretary Robert Mosbacher” (PERSON)

“Sers Šons Konerijs”

“John Doe, Jr.” (PERSON)

“Jānis Bērziņš Juniors” (PERSON)

[Janis.Berzins@mail.com](mailto:Janis.Berzins@mail.com) (NON-ENTITY)

Titles such as "Mr." and role names such as "President" are \*not\* considered part of a person name. However, appositives such as "Jr.", "Sr.", and "III" \*are\* considered part of a person name.

Do not tag e-mail addresses as person names (nor locations, nor organizations, nor any other entity!)

“the Kennedy family and the Kennedys” (PERSON)

“Bērziņu ģimene un Bērziņi” (PERSON)

Family names are to be tagged.

Miscellaneous types of proper names that are not to be tagged as PERSON include political groups, laws named after people, diseases/prizes named after people, and saints (because removal of a saint's title leaves a non-unique name).

* + 1. Specific Cases for Date and Time

The TIME sub-type is defined as a temporal unit shorter than a full day, such as second, minute, or hour. The DATE sub-type is a temporal unit of a full day or longer.

Absolute TIME and DATE examples:

“20 minutes after 10” (TIME)

“10 minūtes pēc pieciem” (TIME)

“twelve o'clock noon” (TIME)

“midnight” (TIME)

“pusnakts” (TIME)

“5 p.m. EST” (TIME)

“Monday” (DATE)

“10th of October” (DATE)

“Easter” (DATE)

“Līgosvētki” (DATE)

“autumn” (DATE)

“fourth quarter” (DATE)

“1995” (DATE)

“the 20th century” (DATE)

“1980s” (DATE)

“80-ie gadi” (DATE)

“the autumn report” (DATE)

“the fourth quarter ended Sept. 30” (DATE)

“ceturtais ceturksnis līdz 30. septembrim” (DATE)

“fiscal 1989's fourth quarter” (DATE)

“1989. fiskālā gada ceturtais ceturksnis” (DATE)

Temporal expressions are to be tagged as a single item. Full day holidays have to be tagged as dates.

Relative TIME and DATE examples:

“last two months” (DATE)

“yesterday” (DATE)

“morning” (TIME)

“Līgo vakars” (TIME)

“Christmas eve” (TIME)

“last night” (TIME)

“1999. gada 5. februāra pusnaktī” (DATE, TIME)

“February 12, 8 A.M.” (DATE, TIME)

“yesterday evening” (DATE, TIME)

If an expression consists of two separate types (DATE and TIME together), then these are tagged separately. Holiday mornings, evenings, afternoons have to be tagged as time.

“by 9 o'clock Monday” (TIME, DATE)

“līdz 9 no rīta, pirmdien” (TIME, DATE)

“around the 4th of May” (DATE)

“ap ceturto maiju” (DATE)

“shortly after May 4” (DATE)

“neilgi pēc 4. maija” (DATE)

Determiners that introduce the expressions are to be tagged only if these express a period of time. This will be done to balance between too long and too short date expressions, and to allow time periods to be tagged as well!

Words or phrases modifying the expressions (such as "around" or "about") also will not be tagged if these express indefinite time periods and none of the period ends are definable. Modifiers that are separated by an article will not be tagged.

“4:15 p.m. Tuesday local time” (TIME)

“early Friday evening” (TIME)

“Tuesday, 4:15 p.m. local time” (DATE, TIME)

“The morning after the July 17 disaster” (TIME, DATE)

“July last year” (DATE)

“1:30 p.m. Chicago time” (TIME)

“Japan time, 19 February, 8:00 A.M.” (TIME)

“February 19th, 8:00 A.M., Japan” (DATE, TIME, LOCATION)

If an expression consists of an embedded expression (also LOCATION) within another expression, the whole expression has to be tagged as a single tag. Be sure that the expression is really embedded in the other expression and not just adjacent to it! When a time expression contains both relative and absolute elements, the entire expression is to be tagged.

Indefinite or vague date expressions with non-specific starting or stopping dates will not be tagged. Non-taggable expressions include: “now”, “recently”, “after some time”, “for the past few years”, “since”.

“Ziemassvētkos” (DATE)

“All Saints' Day” (DATE)

“In Ziemassvētku vakarā” (TIME)

Special days, such as holidays, which are referenced by name, should be tagged.

Exceptional Cases

* + 1. Specific cases for Money, Percent and Quantity

“20 million New Pesos” (MONEY)

“$42.1 million” (MONEY)

“million-dollar conferences” (MONEY)

“several million dollars” (MONEY)

“30 million plus New Pesos” (MONEY)

“U.S. $700 million” (MONEY)

“U.S. $700 million per month” (MONEY)

“U.S. $700 million/month” (MONEY)

“about U.S. $700 million” (MONEY)

“55 Igaunijas kronas” (MONEY)

“15 pct” (PERCENT)

“15%” (PERCENT)

“Two and a half percent” (PERCENT)

“minus 15 percent” (PERCENT)

“55 km, 100 m and 13 cm long” (QUANTITY)

The entire string expressing the monetary, percentage or quantitative value is to be tagged.

1. MUC-7 Named Entity Task Definition - <http://www-nlpir.nist.gov/related_projects/muc/proceedings/ne_task.html> [↑](#footnote-ref-1)