

# **RANLP2019**

## **ANNOTATION GUIDELINE**

### **Entity Types**

OKE2016 - PER, LOC, ORG, ROLE

REUTERS128 - PER, LOC, ORG

Record the following:

### **Missing Entities**

Doc

SurfaceForm

Type

Comment

Here record all entities that are missing from current Orbis visualization, regardless of the reason.

Only PER, LOC, ORG, ROLE should be annotated for now.

### **Overlap**

Doc

SurfaceForm

Comment

Here record the long surface forms that can lead to various overlap incidents (e.g., U.S. Embassy in United Kingdom, U.S. Embassy, etc), regardless of if they are nested entities or multiple entities.

Only PER, ORG, LOC should be annotated for now.

### **Nested Entities**

U.S. Embassy in *United Kingdom*

### **Multiple Entities**

Loyola's University in Belgium, Economics

The **Comment** field should only be used to record additional spans for the mentioned entity.

E.g., if *U.S.* is annotated six times in a document as a separate entity (e.g., do not take into account nestedness here), but it appears 7 times, the span of the annotation that is not displayed should be recorded.

## **POTENTIAL OVERLAP**

Several cases need to be considered:

*Geolocation entities separated by Comma (even when that is the standard name of the entity in a KB)*

Nagoya, Japan

King's College, Cambridge

Blackpool, Lancashire, England

*Multiple names (entities) without comma between them*

**Southeastern Louisiana State University** *Louisiana State University*

*Potential collision due to change of names or otherwise*

**Port Huron** *Border Cats*

*Nested Entities*

**U.S. Coast Guard** *Reserve*

**Yale** *School of Law*

*Long entities with at least 3 names that can have some potential confusion due to nested entities*

**Legislative Assembly** *of Manitoba*

**U.S. Senate** *Committee on Budget*

**MIT** *Radiation Laboratory*

## **What should not be reported as overlap**

*Cases in which the name is effectively a single entity, even if in some cases there might be confusion*

Federal Reserve  
Amsterdam bourse (Euronext Amsterdam)

## **MISSING ENTITIES**

*PER, ORG, LOC*

Same rules used in published annotation guidelines

*Co-reference*

**His**

French

Italian

*Roles*

**M.D.** (if it is part of the name as in Dr. Marcus Welby, M.D.)

Son

Mother

Descendant