IN MEDIA RES
ANNOTATION GUIDELINE

Take into account both real and fictional entities.

Examples of Entity Types
PERSON (PER)

Patrick Stewart

Sting

Jean-Luc Picard

ORGANISATION (ORG)

United Nations

Google

Apple

Mars

LOCATION (LOC)

New York City

Manhattan

London

Fillory

TV (TV)

Star Trek: The Original Series

Seinfeld

Friends

M.A.S.H.

WORK (WORK)

Dracula / Bram Stoker's Dracula

Solar

2666

The Last Supper

Hamlet

Agatha Christie's Poirot

ROLE (ROLE)

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

Son

Mother

Descendant

MUSIC (MUSIC) – Musical Works

Purple Rain

Under Pressure

Symphony No. 9

Blade Runner (soundtrack)

EVENT (EVENT)

World War I

World War II

September 11

Super Bowl

Academy Awards

Golden Globe Awards

OTHER (OTHER)

Co-reference

His

French

Italian

Entity names in another language (e.g., Chinese, Japanese)

Dates (optional)

Everything that can't be classified with a clear type should be considered as OTHER.

FORMAT

Please record:

Doc

SurfaceForm

Type

Start

Link

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.

OVERLAPS

Nested Entities

<u>U.S.</u> 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* Committe 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)

LENSES

Please examine your initial annotations and consider the following 3 lenses:

MIN – annotation should disregard overlaps and extract minimum number of entities

Star Trek: Picard → (1) Star Trek: Picard

MAX – annotation should disregard overlaps and extract maximum number of entities

Star Trek: Picard → (1) Star Trek; (2) Star Trek: Picard

OMAX – annotation should allow for overlaps and extract maximum number of entities whenever possible

Star Trek: Picard → (1) Star Trek; (2) Star Trek: Picard