

Text Named Entity Recognition (NER) Task

Annotation Guidelines (Version 1.0)

1. Introduction

This document provides detailed instructions for annotating named entities in short Hindi financial sentences.

The objective of this task is **to verify candidate named entities and their assigned entity types**.

You will evaluate whether a highlighted text span is a valid named entity in the given sentence and, if so, whether the assigned entity type is correct.

This task focuses on **verification**, not free annotation. All entity candidates are pre-identified.

You should read this document carefully before beginning the annotation.

2. Onboarding

You will be given access to the annotation workspace during the annotation period. Click on the Text NER page. Here you will find the link to the platform and guidelines.

There will also be a table where you will find your name. When you hover over your name, you will find the ‘open’ option. That will lead you to your page (image below)

You will find a checklist to complete your tasks.

A screenshot of a Notion workspace titled "John". The workspace contains a checklist with the following items:

- Annotations compl...
- Annotator Feedback Empty
- Person Empty
- Segments complet... Empty
- + Add a property

Comments section:

- Vanshikaa Jani 29m
You can add any questions or comments here
- Add a comment...

From the Notion workspace, you will find:

- A link to the Text NER annotation platform with password
- This guideline document
- A checklist for tracking your assigned segments

For communication, you may:

- Email questions or completed annotations to
Vanshikaa.Jani@mbzuai.ac.ae
- Leave comments in the shared workspace (Notion)

3. Annotation Workflow

You will complete the task using a dedicated online interface. The workflow is as follows:

1. Login

- Enter your name and password on the login screen.

2. Review Guidelines

- Read this document within the interface.
- Confirm that you have read and understood the guidelines before proceeding.

3. Segment Selection

- Data are divided into predefined segments.
- Begin with the **pilot segments**.
- Complete the pilot segments first.
- Start remaining segments.

4. Entity Verification

For each annotation item:

- Read the full Hindi sentence carefully.
- Observe the highlighted text span within the sentence.
- Review the proposed entity type.
- Complete the two verification steps described in Section 4.

5. Navigation

- **Previous** allows revisiting earlier items.
- **Save & Next** records the annotation and advances.
- A **Jump to Entity** option in the sidebar enables direct navigation.
- A progress indicator shows how many items in the segment are completed.

6. Translation Option (Optional)

- Translation buttons are available for:
 - the full sentence

- the entity text span
- Use translation **only when necessary** for understanding.
- All translation usage is logged automatically.

7. Download

- After completing a segment, download the CSV file.
 - Send the file to the designated email or upload it to the workspace.
-

4. Entity Verification Task

Each annotation item consists of:

- A sentence (Sentence Under Review)

Sentence Under Review

भारतीय रिजर्व बैंक (**आरबीआई**) ने दिनांक 3 सितंबर 2025 के आदेश द्वारा दि ओटोपलम को-ऑपरेटिव अर्बन बैंक लिमिटेड, पलक्कड़, केरल (बैंक) पर आरबीआई द्वारा 'परिवक्षी कार्रवाई ढांचा (एसएफ)' के अंतर्गत जारी विशिष्ट निदेशों के अननुपालन और 'मास्टर निदेश - क्रेडिट कार्ड और डेबिट कार्ड - निर्गम और आचरण निदेश, 2022' पर आरबीआई द्वारा जारी कितिपय निदेशों के अननुपालन के लिए ₹1 लाख (एक लाख रुपये मात्र) का मौद्रिक दंड लगाया है।

- A highlighted text span (Entity Under Review – Entity text)
- A proposed entity type (Entity Under Review – Proposed type)

Entity Under Review

Entity text

आरबीआई

Proposed type:  Organization

Your task has **two steps**.

 Entity Verification

Step 1	Step 2
Is the highlighted text a named entity in this sentence?	Is the assigned entity type correct?
Entity validity	Entity type correctness
<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Uncertain	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Uncertain

[← Previous](#) [Save & Next →](#)

4.1 Step 1: Entity Validity

You must decide whether the highlighted text span is a valid named entity **in the context of the given sentence.**

Select one of the following:

- **Yes**
The span refers to a concrete, identifiable entity.
- **No**
The span does not function as a named entity in this context.
- **Uncertain**
The span may or may not be an entity, but the sentence does not provide enough clarity.

This decision should be made **before** considering the entity type.

4.2 Step 2: Entity Type Correctness

This step applies **only if you selected “Yes” in Step 1.**

You must verify whether the assigned entity type correctly describes the entity.

Select one of the following:

- **Yes**
The assigned entity type is correct.

- **No**

The assigned type is incorrect. You must select the correct type from the list.

Step 2

Is the assigned
entity type correct?

Entity type
correctness

- Yes
 No
 Uncertain

Select correct
entity type

Organ... ▾

- **Uncertain**

The entity is valid, but the sentence does not clearly indicate its type.

5. Entity Types

Each valid entity belongs to exactly one of the following categories:

- **Organization**

Banks, institutions, regulators, companies, authorities

- **Person**

Named individuals

- **Location**

Cities, states, regions, countries

- **Date**

Explicit dates or time references

- **Amount**

Monetary values, quantities with financial meaning

- **Financial_Product**

Financial instruments, accounts, products

- **Scheme**

Named schemes, policies, regulatory frameworks, directives

- **Miscellaneous**

Named entities that do not fit the above categories

6. Use of “Uncertain”

The **Uncertain** option is an important part of this task.

You should select **Uncertain** when:

- The sentence lacks sufficient context
 - Multiple interpretations are plausible
 - The entity or type cannot be confidently verified
-

7. Principles for Annotation

1. Base all decisions strictly on the sentence provided.
Do not use external knowledge.
 2. Evaluate entities in context.
Do not judge spans in isolation.
 3. Do not infer unstated information.
 4. Treat abbreviations as valid entities **only if clearly linked** to a named entity.
 5. Use translation as a comprehension aid only.
 6. Apply criteria consistently across the dataset.
-

8. Interface Reference Guide

8.1 Sentence Under Review

- Full Hindi sentence displayed at the top
- Highlighted entity span
- Sentence translation button

8.2 Entity Under Review Card

- Displays the entity text
- Displays the proposed entity type
- Entity translation button

8.3 Verification Panel

- Step 1: Entity validity
- Step 2: Entity type correctness
- Correction dropdown (if applicable)

8.4 Sidebar Tools

- Jump to Entity
- Segment progress indicator

8.5 Segment Completion

- Download CSV file after completion
 - Partial progress can be downloaded if needed
 - Multiple CSVs may be submitted if annotation is split across sessions
-

9. Quality Expectations

You are expected to:

- Apply entity definitions consistently
 - Avoid guesswork
 - Use Uncertain where appropriate
 - Complete all assigned segments
-

10. Contact and Support

For questions or clarification, contact:**Vanshikaa.Jani@mbzuai.ac.ae**

Or use the comment section in the shared workspace on Notion

11. Completion

Click on Download CSV to download after each segment (**your work will not get saved automatically**).

All changes are saved locally in the session until explicitly downloaded.

A manageable size of 30 per segment is chosen but if you have to stop early then you may download the current progress. When you return, use the jump option to skip to the last annotated fact in the segment. In such cases you can upload multiple CSVs.

Upload the datasets here

Example:

 pilot_segment_1_Vana.csv 6.8 KiB

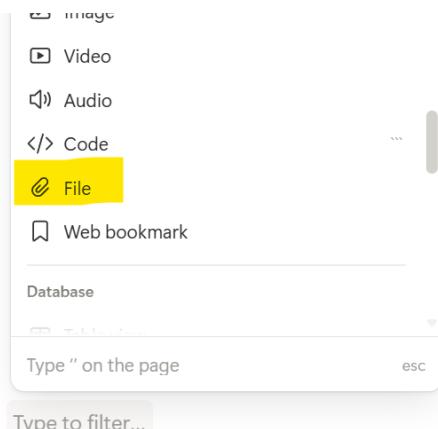
Or email it.

To upload it on Notion:

1. Click on + option



2. Then scroll down till you find “file”



Appendix: Examples and Boundary Cases (Illustrative)

This section provides illustrative examples to clarify how entity validity and entity type decisions should be made.

Each example is shown in **Hindi (original)** and **English (for clarity only)**.

All decisions must be based strictly on the **Hindi sentence**. The English translation is provided only to aid understanding.

1. Valid Entity with Correct Type

Hindi sentence:

“भारतीय रिज़र्व बैंक (आरबीआई) ने नई गाइडलाइन्स जारी की हैं।”

English translation:

The Reserve Bank of India (RBI) has issued new guidelines.

Highlighted span:

“भारतीय रिज़र्व बैंक”

Proposed type:

Organization

Step 1: Is this a valid named entity?

→ Yes

Step 2: Is the entity type correct?

→ Yes

Explanation:

This is a formally named regulatory institution. The entity is explicit and unambiguous.

2. Valid Entity with Incorrect Type

Hindi sentence:

“सरकार ने प्रधानमंत्री जन धन योजना के तहत खातों की संख्या बढ़ाई।”

English translation:

The government increased the number of accounts under the Pradhan Mantri Jan Dhan Yojana.

Highlighted span:

“प्रधानमंत्री जन धन योजना”

Proposed type:

Organization

Step 1: Is this a valid named entity?

→ Yes

Step 2: Is the entity type correct?

→ No

Corrected type:

Scheme

Explanation:

This refers to a government scheme, not an organization. The entity exists, but the assigned type is incorrect.

3. Not a Named Entity

Hindi sentence:

“बैंक ने ग्राहकों को बेहतर सेवाएं प्रदान कीं।”

English translation:

The bank provided better services to customers.

Highlighted span:

“ग्राहकों”

Proposed type:

Person

Step 1: Is this a valid named entity?

→ No

Explanation:

“ग्राहकों” refers to a generic group (customers), not a specific identifiable entity.

4. Ambiguous Entity (Use Uncertain)

Hindi sentence:

“सरकार ने नई डिजिटल पहल की घोषणा की।”

English translation:

The government announced a new digital initiative.

Highlighted span:

“डिजिटल पहल”

Proposed type:

Scheme

Step 1: Is this a valid named entity?

→ Uncertain

Explanation:

The phrase could refer to a named scheme, but no official name is provided. The sentence does not offer sufficient clarity.

5. Abbreviations and Short Forms

Hindi sentence:

“आरबीआई ने बैंकों को निर्देश जारी किए।”

English translation:

The RBI issued directions to banks.

Highlighted span:

“आरबीआई”

Proposed type:

Organization

Step 1: Is this a valid named entity?

→ Yes

Step 2: Is the entity type correct?

→ Yes

Explanation:

Well-established abbreviations referring to formally named entities should be treated as valid entities when clearly used in context.

6. Date and Amount Entities

Hindi sentence:

“दिनांक 3 सितंबर 2025 को ₹1 लाख का जुर्माना लगाया गया।”

English translation:

A fine of ₹1 lakh was imposed on 3 September 2025.

Highlighted span:

“3 सितंबर 2025”

Proposed type: Date

→ Yes / Yes

Highlighted span:

“₹1 लाख”

Proposed type: Amount

→ Yes / Yes

Explanation:

Explicit dates and monetary values are valid entities when they carry factual or financial significance.

7. Location vs Organization

Hindi sentence:

“केरल ने सहकारी बैंकों के लिए नई नीति लागू की।”

English translation:

Kerala implemented a new policy for cooperative banks.

Highlighted span:

“केरल”

Proposed type:

Location

Step 1: Is this a valid named entity?

→ Yes

Step 2: Is the entity type correct?

→ Yes

Explanation:

Even when a location functions as an administrative authority in the sentence, it should be labeled as a Location unless explicitly referring to an organization.

8. Policy or Scheme Name Not Explicit

Hindi sentence:

“सरकार ने नई नीति की समीक्षा की।”

English translation:

The government reviewed a new policy.

Highlighted span:

“नई नीति”

Proposed type:

Scheme

Step 1: Is this a valid named entity?

→ **Uncertain**

Explanation:

The sentence does not specify whether “নई নীতি” is a formally named policy or a generic reference.

9. Summary Decision Guide

Scenario	Step 1: Entity Validity	Step 2: Type Check
Clearly named entity	Yes	Verify type
Generic term or role	No	NA
Entity exists but type unclear	Yes	Uncertain
Ambiguous reference	Uncertain	NA
Type clearly incorrect	Yes	No → Correct

10. Reminder on Annotation Quality

- Prefer **Uncertain** over guessing
- Do not infer official names or types
- Judge entities strictly **within sentence context**
- Apply criteria consistently across all segments