

Task - Entity Linking

What is Entity Linking?

Entity linking is the process of connecting all mentions of a specific person, place, object, etc., in the text to a unique identifier in a dictionary or knowledge base.

Why is this important?

This helps to **remove confusion** when the same name can refer to different people or things, or when the same person is mentioned using different names.

Example 1: One person, many names

The same person may appear with different names:

- अर्जुन (Arjuna)
- धनञ्जय (Dhanañjaya)
- सव्यसाचिन् (Savyasācin)

These all refer to **Arjuna**, the hero from the Mahābhārata.

Example 2: Same name, different people

The name अर्जुन may refer to:

- अर्जुन, the Pāṇḍava prince
- अर्जुन, the Haihaya king (Kārtavīrya Arjuna)
- अर्जुन, a ṛṣi in the palace of Yama

Entity linking helps us find out **which अर्जुन** the text is talking about.

In such cases, entity linking helps us clarify which **अर्जुन** is meant by connecting it to a unique ID in the dictionary.

Another Example from Mahabharata

Verse:

यदाश्रीषं देवराजं प्रवृष्टं शरैर्दिव्यैर्वारितं चार्जुनेन ।
अग्निं तथा तर्पितं खाण्डवे च तदा नाशंसे विजयाय संजय ॥

In this verse, the word **देवराजं** may refer to any of four different characters called "devarāja" which are listed in the dictionary. One of them is **Indra**, whose ID is 3264. The annotator's task is to choose the correct one based on the context.

देवराज

देवराज [Printed book page 237-1]

Devarāja¹ = Indra, q.v. [ID=3264]

Devarāja² = Nahuṣa, q.v. [ID=3265]

Devarāja³ = Vasiṣṭha: XIII, 6256 (V°). [ID=3266]

[Printed book page 237-2]

Devarāja⁴, an ancient king. § 267 (Yamasabhāv.): II, 8, 337 (in the palace of Yama).

[ID=3267]

Basic Concepts

Entity:

A real-world thing or person. Example: कृष्ण, द्वारका, नृपः

Named Entity:

A named entity is a specific name used to identify a single or group of person, place, object, time, concept etc. These names usually refer to one specific thing in the world and do not change meaning based on context. Named entities are the main focus of annotation.

- For example, कृष्ण (if it refers to the person Kṛṣṇa) and द्वारका (if it refers to the city) are named entities.
- However, general words like नृपः (king) or शार्दूलः (tiger) are not named entities because they do not point to a specific, unchanging individual or group.

Examples:

- **Person:** वसुदेवस्य, हिरण्यकशिपु
- **Location:** मथुरायां, यमुनां
- **Object:** सुदर्शनास्त्रं
- **Time:** कल्पायुषो
- **Miscellaneous:** गायत्री (a mantra)

Mentions:

A mention is the actual occurrence of the name in the text.

Example: In the verse bellow:

Verse:

यदाश्रीषं देवराजं प्रवृष्टं शरैर्दिव्यैर्वारितं चार्जुनेन ।
अग्निं तथा तर्पितं खाण्डवे च तदा नाशंसे विजयाय संजय ॥

Entity	Mentions
अर्जुन	च-अर्जुनेन (Only अर्जुनेन)
देवराज	देवराजं
अग्नि	अग्निं
खाण्डव	खाण्डवे
संजय	संजय

Mentions are all are inflected forms (like अर्जुनः, अर्जुनेन, अर्जुनम्) found in text, but they refer to the same entity.

Named Entities in Sanskrit

Sanskrit names appear in several ways:

1. **Proper names:** अर्जुन
2. **Names based on characteristics:** सव्यसाचिन् (ambidextrous)
3. **Names based on relationships:** पाण्डव (son of Pāṇḍu), कृष्णसारथि (Kṛṣṇa's charioteer)

More examples of Name Variants in Mahabharata the we will consider as Named Entities.

Entity	Name Variant	Meaning in English	Type
अर्जुन	अर्जुन	-	Proper Name
	धनञ्जय	Winner of wealth	Defined by Characteristics
	सव्यसाचिन्	Ambidextrous	
	भारतश्रेष्ठ	Best of the Bhāratas	
	पाण्डव	Son of Pāṇḍu	Defined by Relation
	पाण्डवेय	Descendant of Pāṇḍu	
	पाण्डुपुत्र	Son of Pāṇḍu	
	भीमसहानुज	Younger brother of Bhīma	
	कृष्णसारथि	Charioteer of Kṛṣṇa	
द्रौपदी	श्यामा	Dark-complexioned	Defined by Characteristics
	कृष्णा	Dark-skinned or feminine of Kṛṣṇa	

	द्रौपदी	Daughter of Drupada	Defined by Relation
	द्रुपदकन्या	Daughter of Drupada	
	द्रुपदपुत्री	Daughter of Drupada	
	द्रुपदात्मजा	Born of Drupada	
	कौरवराजपत्नी	Wife of the Kaurava princes	
	पाञ्चालपुत्री	Daughter of Panchāla	
	पाञ्चाली	Princess of Panchāla	
	पार्षती	Granddaughter of Prṣata	
	पावकात्मजा	Daughter of fire	
	सुतसोममातृ	Mother of Sutasoma	
	याज्ञसेनी	Daughter of Yājñasena (Drupada)	
जनमेजय	जनमेजय	-	Proper Name
	पारिक्षित	Tested one – tested by gods while in womb	Defined by Characteristics
	कुरुप्रवीर	Hero among the Kurus	Defined by Relation
	पाण्डवनन्दन	Joy of the Pāṇḍavas	
	भारत	Descendant of Bharata	
Khāṇḍava	खाण्डव	[Name of a forest]	Proper Name

Rules of Thumb for Annotating Names

- The word is or contains a **proper noun**.
- The name can include other parts (e.g. relations)
- It **refers to a specific person/place/object etc** in the story.
- Its **reference remains the same over time**.
- See the [index_data.xlsx](#) for all the names in Mahabharata

Do not annotate generic titles or roles that change (like “king” or “warrior”).

Transliteration and SLP1 Format

In this annotation task, you will be provided with:

- **Devanagari text**
- **SLP1 text** (Sanskrit transliterated using English characters)
- **English translation**

Example from annotation platform:

1	upahArErvaYcayitvA varmanyeva suyoDanaH .
2	varadaM taM varaM vatre sAhAyyaM kriyatAM mama ..
3	*****
4	[--METADATA-DO NOT ANNOTATE:
5	Verse_ID: Volume_1.Chapter_2.Subchapter_1.Verse_200
6	Devanagari: उपहारैर्वञ्चयित्वा वर्मन्येव सुयोधनः । वरदं तं वरं वत्रे साहाय्यं क्रियतां मम ॥
7	Translation: Having deceived him on the way by present. Duryodhana induced him to grant him a boon and for that boon he asked his help in the war.
8	--]

What is SLP1?

SLP1 (Sanskrit Library Phonetic Basic Encoding) is a standard way of writing Sanskrit using only English characters. It is especially useful for computational processing and precise dictionary search.

See more: [SLP1 - Wikipedia](https://en.wikipedia.org/wiki/SLP1)

SLP1 is case-sensitive and uses a one-to-one mapping between Devanagari characters and Roman letters. For example:

Example: यदाश्रीषं - yadASrOzaM [य - ya दा - dA श्री - SrO षं - zaM]

Example

SLP1:

yadASrOzaM devarAjaM pravfzwaM SarErdivyErvAritaM cArjunena .
agni taTA tarpitaM KARqave ca tadA nASaMse vijayAya saMjaya ..

Devanagari:

यदाश्रीषं देवराजं प्रवृष्टं शरैर्दिव्यैर्वारितं चार्जुनेन ।
अग्निं तथा तर्पितं खाण्डवे च तदा नाशंसे विजयाय संजय ॥

Translation:

"I had no hope of success, O Sanjaya, when I heard that Arjuna had satisfied Agni by giving him the forest of Khandava, preventing at the same time by his celestial arrows the downpour made by Indra, the king of the celestials."

Important Instructions

Dashboard

admin - New entity annotation project 0001_Volume_1_Chapter_1_Subchapter_0_Verse_121.txt 1-8 / 8 ser

(NamEnt...)

1 yadASrOzaM devarAjaM pravfzwaM SarErdivyErvAritaM cArjunena .
2 agni taTA tarpitaM KARqave ca tadA nASaMse vijayAya saMjaya ..
3 *****
4 [--METADATA-DO NOT ANNOTATE:
5 Verse_ID: Volume_1.Chapter_1.Subchapter_0.Verse_121
6 Devanagari: यदाश्रौष देवराजं प्रवृष्टं शरैर्दिव्यैर्वीरितं चाजुनेन । अग्निं तथा तर्पितं खाण्डवे च तदा नाशसे
7 Translation: I had no hope of success, O Sanjaya, when I heard that Arjuna had satisfied
8 --]

[1] devarAja (3264)
https://www.sanskrit-lexicon.uni-koeln.de/scans/INMScan/2020/web/webtc/indexcaller.php?input=slp1&output=deva&key=devarAja&id=3264&sub_entry=0
ALIAS OF: indra (4618_0) -- Devarāja1 = Indra, q.v.

[2] devarAja (3265)
https://www.sanskrit-lexicon.uni-koeln.de/scans/INMScan/2020/web/webtc/indexcaller.php?input=slp1&output=deva&key=devarAja&id=3265&sub_entry=0
ALIAS OF: nahuza (7485_0) -- Devarāja2 = Nahuṣa, q.v.

[3] devarAja (3266)
https://www.sanskrit-lexicon.uni-koeln.de/scans/INMScan/2020/web/webtc/indexcaller.php?input=slp1&output=deva&key=devarAja&id=3266&sub_entry=0
ALIAS OF: vasiṣṭha (11659_0) -- Devarāja3 = Vasiṣṭha: XIII, 6256 (V*).

[4] devarAja (3267)
7 items found

devarAja

Annotate Only the SLP1 Text

- Always annotate using the **SLP1 line**, not the Devanagari line.
- The Devanagari and Translation lines are for reference only.

Use SLP1 for Dictionary Search

- Use the SLP1 version of the text to perform dictionary lookups.
- First, **select the word or span** in the SLP1 text by highlighting it from the start character to the end character.

How to Search in the Dictionary

- After selecting the text, type the **base form** (root form) of the word in the search box.
- Do not include grammatical inflections.

Example:

If the word in the text is **arjunena**, search for **arjuna**.

This helps in identifying the correct entry in the dictionary.

Choosing the Correct Dictionary Entry

- The dictionary will show one or more meanings with the label “ALIAS OF” and an ID.
Example results:
 - ALIAS OF Indra [ID=3264]
 - ALIAS OF Nahuṣa [ID=3265]
 - ALIAS OF Vasiṣṭha [ID=3266]
- Choose the one that is correct based on the context of the verse.

Handling Sandhi in Annotation

In Sanskrit, *Sandhi* is the phonetic merging of two words at their boundaries. This often causes proper names or entities to be joined with conjunctions, prepositions, or other elements. When this occurs, **only the named entity portion** should be annotated — **not the entire Sandhi-joined word**.

Key Rule:

Mark only the portion of the word that refers to the named entity, not any prefix or suffix joined via Sandhi.

Examples:

1. cArjunena

- **SLP1:** cArjunena (चार्जुनेन)
- **Breakdown:** ca (and) + arjunena (by Arjuna)
- **Annotate:** arjunena
- **Do not annotate:** cArjunena

Nested Mentions

Sometimes, named entities appear as part of a longer compound that refers to a different specific entity. In such cases, **do not annotate internal subparts** that are themselves named entities if the full compound clearly refers to **a different named entity**.

Key Rule:

Do not annotate subparts of a compound if the entire phrase refers to another distinct entity.

Example:

drupadaputrI

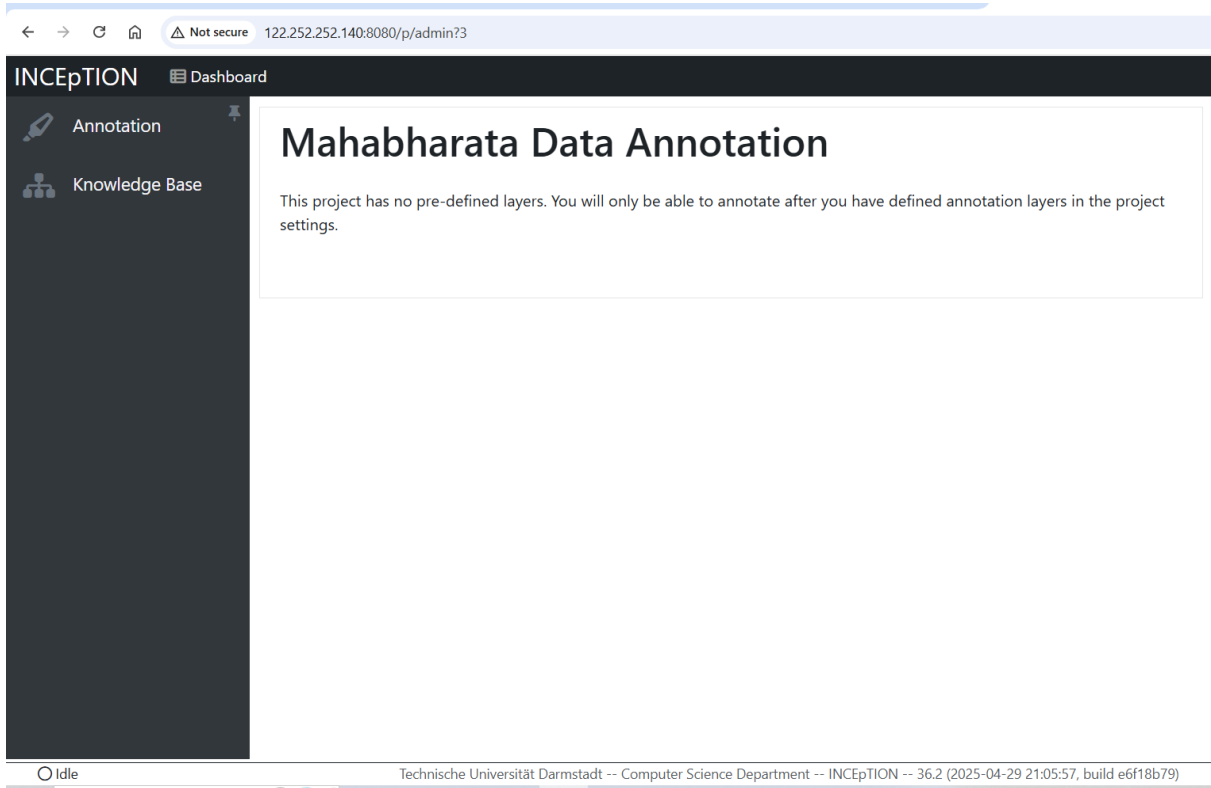
- **SLP1:** drupadaputrI (द्रुपदपुत्री)
- **Breakdown:** drupada (Drupada) + putrI (daughter)
- The compound refers to **Draupadī**.
- **Annotate:** drupadaputrI
- **Do not separately annotate:** drupada

Accessing the Annotation Platform

1. Open your browser and visit: <http://122.252.252.140:8080/>
2. Log in using the **username and password** provided to you.

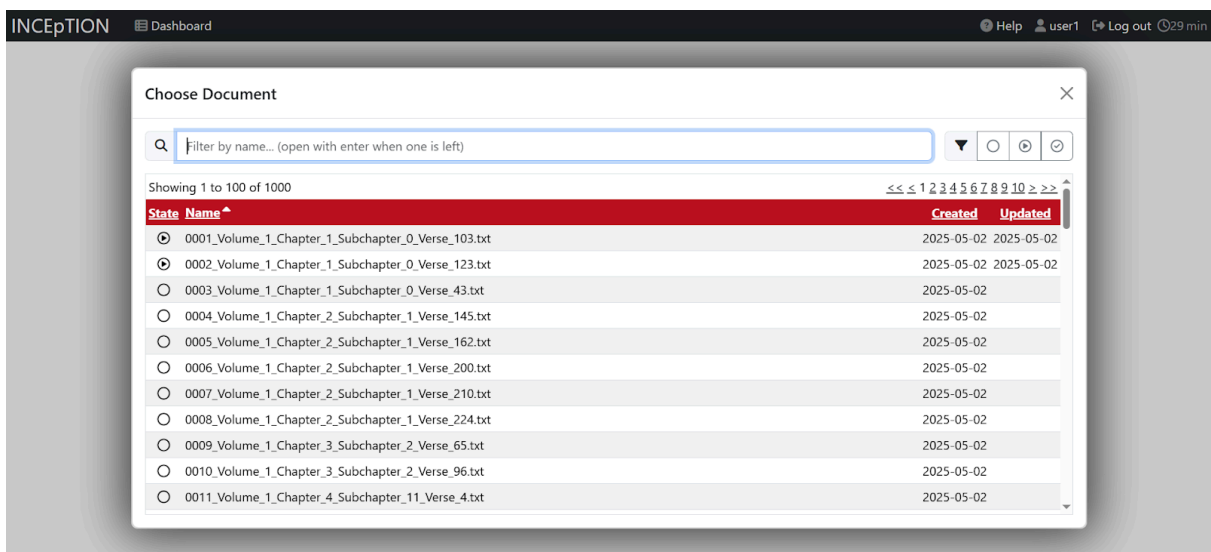
Getting Started After Login

- Click on **Knowledge Base** to explore the dictionary of possible named entities from the *Mahābhārata*.
- Click on **Annotation** to begin annotating the verses.



Selecting a Verse

- You can either select a specific verse number to begin from, or start from the first and navigate using the arrows.
- The annotation window will display **one verse at a time**.
- Annotate only the part in **SLP1** (Sanskrit in Latin script); **do not annotate metadata**.



Performing Annotations

1. **Highlight** the text (in SLP1) you think is a named entity — from the **start to end character**.

INCEpTION Dashboard

Mahabharata Data Annotation 0006 Volume_1 Chapter_2 Subchapter_1 Verse 200.txt 1-8 / 8 sentences [doc 6 / 1000]

Annotations

Group by label

Filter

Sort by score

Suggestions first

1 upahArErvaYcayitVA varmanyeva suyoDanaH .

2 varadaM taM varaM vatre sAhAyyaM kriyatAM mama ..

3 *****

4 [--METADATA-DO NOT ANNOTATE:

5 Verse_ID: Volume_1.Chapter_2.Subchapter_1.Verse_200

6 Devanagari: उपहसिर्वज्रयित्वा वर्मण्येव सुयोधनः । वरदे तं वरं वज्रे साहाय्यं क्रियतां माम् ॥

7 Translation: Having deceived him on the way by present. Duryodhana induced him to grant him a boon and for that boon he asked his help in the war.

8 --]

No annotation selected

122.252.252.140:8080/p/admin Technische Universität Darmstadt -- Computer Science Department -- INCEpTION -- 36.2 (2025-04-29 21:05:57, build e6f18b79)

- In the **right-side panel**, under the search box labeled **identifier**, enter the **base form** (lemma) of the word (e.g., type *arjuna*, not *arjunena*).
- A list of matching entities will appear — select the correct one based on the context.

INCEpTION Dashboard

Mahabharata Data Annotation 0002 Volume_1 Chapter_1 Subchapter_0 Verse 123.txt 1-8 / 8 sentences [doc 2 / 1000]

Annotations

Group by label

Filter

Sort by score

Suggestions first

▼ [Named entity]

saMjaya .. ***** [--M]

▼ arjuna (802)

arjunena . stadA nASa

▼ drOpadI (3660)

drOpadI raNgamaDye

1 yadASrOzaM drOpadI raNgamaDye lakzyaM BitvA nirjitAmarjunena .

2 stadA nASaMse vijayAya saMjaya ..

3 *****

4 [--METADATA-DO NOT ANNOTATE:

5 Verse_ID: Volume_1.Chapter_1.Subchapter_0.Verse_123

6 Devanagari: यदश्रीषं द्रोपदी रङ्गमध्ये लक्ष्यं मित्वा निर्वितामर्जुनेन । सदा नाशंसे विजयाय संजय ॥

7 Translation: I had no hope of success, O Sanjaya, when I heard that Arjuna had obtained the hands of Draupadi by piercing the mark and the brave Panchalas had joined the Pandavas.

8 --]

Span

Layer

Named entity

Text

saMjaya

No links or relations connect to this annotation.

Identifier

value

PER

LOC

OTH

Idle Technische Universität Darmstadt -- Computer Science Department -- INCEpTION -- 36.2 (2025-04-29 21:05:57, build e6f18b79)

- To better understand the context, use the **Volume**, **Chapter**, **Subchapter**, and **Verse_ID** to locate the verse in the full *Mahābhārata* text.
- **mahabharata_corpus**

INCEPTION Dashboard

Mahabharata Data Annotation 0002_Volume_1_Chapter_1_Subchapter_0_Verse_123.txt 1-8 / 8 sentences [doc 2 / 1000]

Annotations

Group by label

Filter

Sort by score

Suggestions first

▼ [Named entity]

saMjaya . ***** [---M]

▼ arjuna (802)

arjunena . stadA nASa

▼ drOpadi (3660)

drOpadi raNgamaDye

1 yadASrOzaM drOpadi raNgamaDye lakzyaM BitvA nirjitAmarjunena .

2 stadA nASaMse vijayAya saMjaya ..

3 *****

4 [--METADATA-DO NOT ANNOTATE:

5 Verse_ID: Volume_1.Chapter_1.Subchapter_0.Verse_123

6 Devanagari: यदाश्रीषं द्रौपदीं रङ्गमये लक्ष्मं भिला निजितामर्जुनेन । सदा नाश्रिसे विजयाय संजय ॥

7 Translation: I had no hope of success, O Sanjaya, when I heard that Arjuna had obtained the hands of

8 --]

drOpadi (3659)

https://www.sanskrit-lexicon.uni-koeln.de/scans/INMScan/2020/web/webtr/indexcaller.php?input=slp1&output=deva&key=drOpadi&sub=3659&sub.entry=0

ALIAS OF: Dhṛvadyumna (3497_0) -- Draupadi ('son of Drupada') = Dhṛtadyumna: VIII, 288.

[2] drOpadi (3660)

https://www.sanskrit-lexicon.uni-koeln.de/scans/INMScan/2020/web/webtr/indexcaller.php?input=slp1&output=deva&key=drOpadi&sub=3660&sub.entry=0

Draupadi: the wife of the Pāṇḍavas (her proper name was Kṛṣṇā K.) as daughter of the Pāṇḍava king Drupada she was named Draupadi. Pāṇḍit (P), Pāṇḍit (Pā), and Vāṇḍit (Y). --4 4 (Anukram.); I. 1, [125 (K.), 148 (K.)], 152, 155, [167 (K.)], --4 10 (Parvasāṅg); I. 2, [314 (P)] --4 11 (do.); I. 2, [361 (P), 384 (K.)], 385, 388, 392, 394, 395, 402 (Yās tanayān), 413, [427 (P), 430 (K.)], 452, [456 (P)], 469, 473, [483 (P)], 573, 574, 578, 630, 632 --4 71 (Ādvamśāvataraṇa); I. 61, [2259 (K.)], 2260 (all: to § 231 (all)) --4 72 (do.); I. 62, 2269 (Kṛṣṇā) --4 83 (Ādvamśāvataraṇa); I. 63, [2438 (K., all: to § 218)], 2450 (Pāṇḍalāyāṁ, mother of five sons: Prativindhya from Yudhiṣṭhira.

[3] drOnadīa (3663)

- To **delete** a wrong annotation, click the **cross (x)** on the left-side panel next to that annotation.

Mahabharata Data Annotation 0002_Volume_1_Chapter_1_Subchapter_0_Verse_123.txt 1-8 / 8 sentences [doc 2 / 1000]

Toggle state (Ctrl+End)

1 yadASrOzaM drOpadi (3660) raNgamaDye lakzyaM BitvA nirjitAmarjunena

arjuna (802)

- Click the **Toggle State** button (or press **Ctrl + End**) to complete the annotation.
- Note: Even if you forget to click it, your annotation will still be saved.