

From BERT Fine-tuning to LLM Prompting

Geospatial Named Entity Recognition in Historical French Texts

Ludovic Moncla

ludovic.moncla@insa-lyon.fr



About me

Ludovic Moncla

- Associate professor in Computer Science at INSA Lyon & LIRIS
- NLP, GIS, Digital Humanities



Current PhD students

- **Alice Brenon** (Feb. 2021-): Methods and tools for geographic discourse analysis
- **Yassir Lairgi** (Dec. 2023-): Large Language Models, Knowledge Graphs and Dynamic Graph Neural Networks for competitive intelligence (*industrial CIFRE agreement*)
- **Antoine Taroni** (Oct. 2024-): Interpretability of Large Language Models: towards a hybrid neuro-symbolic approach for the analysis of geographical discourse

Summary of the talk

1. The GeoEDdA dataset: A Gold Standard Dataset for Geo-semantic Annotation
2. BERT fine-tuning for NER
3. Few-shot prompting for NER
4. Perspectives: Knowledge Graph construction using LLMs

Summary of the talk

1. The GeoEDdA dataset: A Gold Standard Dataset for Geo-semantic Annotation
 - collaborative work with D. Vigier & K. McDonough
2. BERT fine-tuning for NER
3. Few-shot prompting for NER
4. Perspectives: Knowledge Graph construction using LLMs

A Gold Standard Dataset for Geo-semantic Annotation

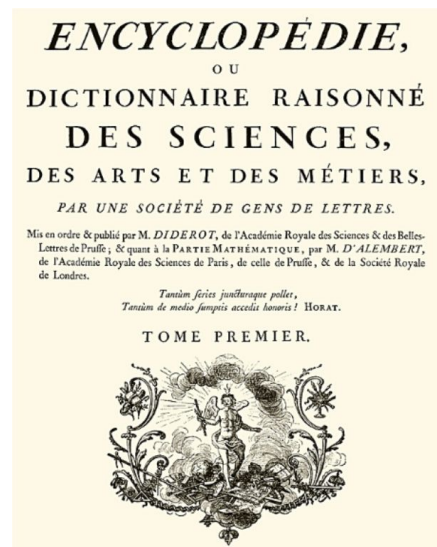
Introduction

Objectives

- Propose a methodology for building a gold standard dataset for geo-semantic annotation
 - Define an annotation schema that goes beyond standard NER
 - Experiment an active learning method for assisted corpus construction
- Train a first span categorization model

Dataset

- Diderot & d'Alembert, eds., Encyclopédie
 - ARTFL Project (<https://artfl-project.uchicago.edu/>)
 - 74,000 articles



A Gold Standard Dataset for Geo-semantic Annotation

GO TO HOME PAGE ARTFL PROJECT TEXTUAL OPTICS LAB UNIVERSITY OF CHICAGO ATILF-CNRS CONTACT US 

CITE US **ARTFL ENCYCLOPÉDIE** REPORT ERROR

CONCORDANCE KEYWORD IN CONTEXT AGGREGATION COLLOCATION

Search Terms ? Search Clear Show search options

Head word

GLASCOW • Géographie • Jaucourt

< Table of contents >

[page 7:704]

GLASCOW
GLASCOW, *Glarona*, (*Géog.*) ville d'Ecosse dans la province de Clydale, avec une célèbre université; elle étoit autrefois archi-épiscopale : la cathédrale subsiste encore, & c'est un beau morceau d'Architecture. On appelle Glasgow *le paradis d'Ecosse*. Il s'y fait un grand commerce, à cause de son port & de son havre ; elle est sur la Clyde, à quatre lieues S. O. de Dumbarton, quatorze O. d'Edimbourg, six de Sterling, cent-quatorze N. O. de Londres. *Longit.* 13. 36. *latit.* 56. 20. ²

Cette ville a produit plusieurs gens éminens dans les Sciences ; je n'en nommerai que deux qui se présentent à ma mémoire, Cameron & Spotswood. Le premier s'est distingué par ses *remarques sur le nouveau testament*, qui sont également savantes & judicieuses ; il mourut à Montauban vers l'an 1625 à quarante-six ans. Spotswood devint archevêque de Saint André, & primat de toute l'Ecosse : il couronna Charles I. en 1633, fut lord chancelier, & mourut en 1639, âgé de soixante-quatorze ans. On a de lui une *histoire ecclésiastique d'Ecosse* fort estimée ; elle s'étend depuis l'an 203 de J. C. jusqu'en 1624. (*D. J.*)

Powered by


de Zurich. *Longit.* 26. 48. *latit.* 47. 6. (*D. J.*)
GLASCOW, *Glarona*, (*Géog.*) ville d'Ecosse dans la province de Clydale, avec une célèbre université; elle étoit autrefois archi-épiscopale : la cathédrale subsiste encore, & c'est un beau morceau d'Architecture. On appelle Glasgow *le paradis d'Ecosse*. Il s'y fait un grand commerce, à cause de son port & de son havre ; elle est sur la Clyde, à quatre lieues S. O. de Dumbarton, quatorze O. d'Edimbourg, six de Sterling, cent-quatorze N. O. de Londres. *Longit.* 13. 36. *latit.* 56. 20.

GLASCOW, *Glarona*, (*Geog.*) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton. W of Dumbarton, fourteen W of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. *Longit.* 13. 36. *latit.* 56. 20. [...]

A Gold Standard Dataset for Geo-semantic Annotation

Structure: headword & domain-mark

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal:

Head

Domain-mark

the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year 1625 at the age of forty-six. Spotswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Spatial named entities (place names)

GLASCOW, Glarona, (Geog.) city of **Scotland** in the province of **Clydale**, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. **Glasgow** is called the paradise of **Scotland**. There is a great deal of trade there, because of its port and harbour; it is on the **Clyde**, four leagues south-west of **Dumbarton** fourteen west of **Edinburgh**, six of **Sterling**, one hundred and fourteen N. W. of **London**. Longit. 13. 36. latit. 56. 20. This town has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spootswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in **Montauban** around the year 1625 at the age of forty-six. Spootswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of **Scotland**, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Person named entities

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town has produced several eminent people in Science; I will name only two who stand out in my memory, **Cameron** & **Spootswood**. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year 1625 at the age of forty-six. **Spootswood** became archbishop of Saint Andrew, and primate of all Scotland: he crowned **Charles I.** in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Named entities not classified as spatial or person (e.g. dates, objects, events, ...)

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished himself by his remarks on **the New Testament**, which are equally learned and judicious; he died in Montauban around the year **1625** at the age of forty-six. Spotswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I. **1633**, was Lord Chancellor, and died in **1639**, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of Scotland, which extends from **203 A.D.** to **1624** (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Spatial nominal mentions / geographic features

GLASCOW, Glarona, (Geog.) **city** of Scotland in **the province** of Clydale, with a famous university; it was formerly **archi-episcopal**
NC-Spatial **NC-Spatial** **NC-Spatial**
: **the cathedral** still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great
NC-Spatial
deal of trade there, because of its **port** and **harbour**; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of
NC-Spatial **NC-Spatial**
Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. **This town** has produced several
NC-Spatial
eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished
himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year
1625 at the age of forty-six. Spotswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I.
in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of
Scotland, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Spatial relations

GLASCOW, Glarona, (Geog.) city of Scotland **in** the province of Clydale, with a famous university; it was formerly archi-episcopal: **Relation**

the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is **on** the Clyde, **four leagues south-west of** Dumbarton **fourteen west of** Edinburgh, **six of** Sterling, **one hundred and fourteen N. W. of** London. Longit. 13. 36. latit. 56. 20. This town has produced **Relation** **Relation**

several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year 1625 at the age of forty-six. Spotswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Nested spatial named entities

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal

ENE-Spatial

ENE-Spatial

ENE-Spatial

: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spootswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year 1625 at the age of forty-six. Spootswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Geographic coordinates

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. **Longit. 13. 36. latit. 56. 20.** This town has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year 1625 at the age of forty-six. Spotswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J).

A Gold Standard Dataset for Geo-semantic Annotation

Person nominal mentions

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town has produced several **eminent people** in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year 1625 at the age of forty-six. Spotswood became **archbishop** of Saint Andrew, and **primate** of all Scotland: he crowned Charles I. in 1633, was **Lord Chancellor**, and died in 1639, aged seventy-four. He is **the author** of a highly esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Nested person named entities

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around the year 1625 at the age of forty-six. Spotswood became **archbishop of Saint Andrew**, and **primate of all Scotland**: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

Nested named entities (not classified as spatial or person)

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called **the paradise of Scotland**. There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood. The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in Montauban around **the year 1625** at the age of forty-six. Spotswood became archbishop of Saint Andrew, and primate of all Scotland: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly esteemed **ecclesiastical history of Scotland**, which extends from 203 A.D. to 1624 (D. J.).

A Gold Standard Dataset for Geo-semantic Annotation

GLASCOW, Glarona, (Geog.) city of Scotland in the province of Clydale, with a famous university; it was formerly



archi-episcopal: the cathedral still survives, & it is a beautiful piece of Architecture. Glasgow is called the paradise of Scotland.



There is a great deal of trade there, because of its port and harbour; it is on the Clyde, four leagues south-west of Dumbarton



fourteen west of Edinburgh, six of Sterling, one hundred and fourteen N. W. of London. Longit. 13. 36. latit. 56. 20. This town



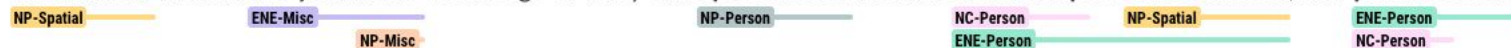
has produced several eminent people in Science; I will name only two who stand out in my memory, Cameron & Spotswood.



The former distinguished himself by his remarks on the New Testament, which are equally learned and judicious; he died in



Montauban around the year 1625 at the age of forty-six. Spotswood became archbishop of Saint Andrew, and primate of all



Scotland: he crowned Charles I. in 1633, was Lord Chancellor, and died in 1639, aged seventy-four. He is the author of a highly

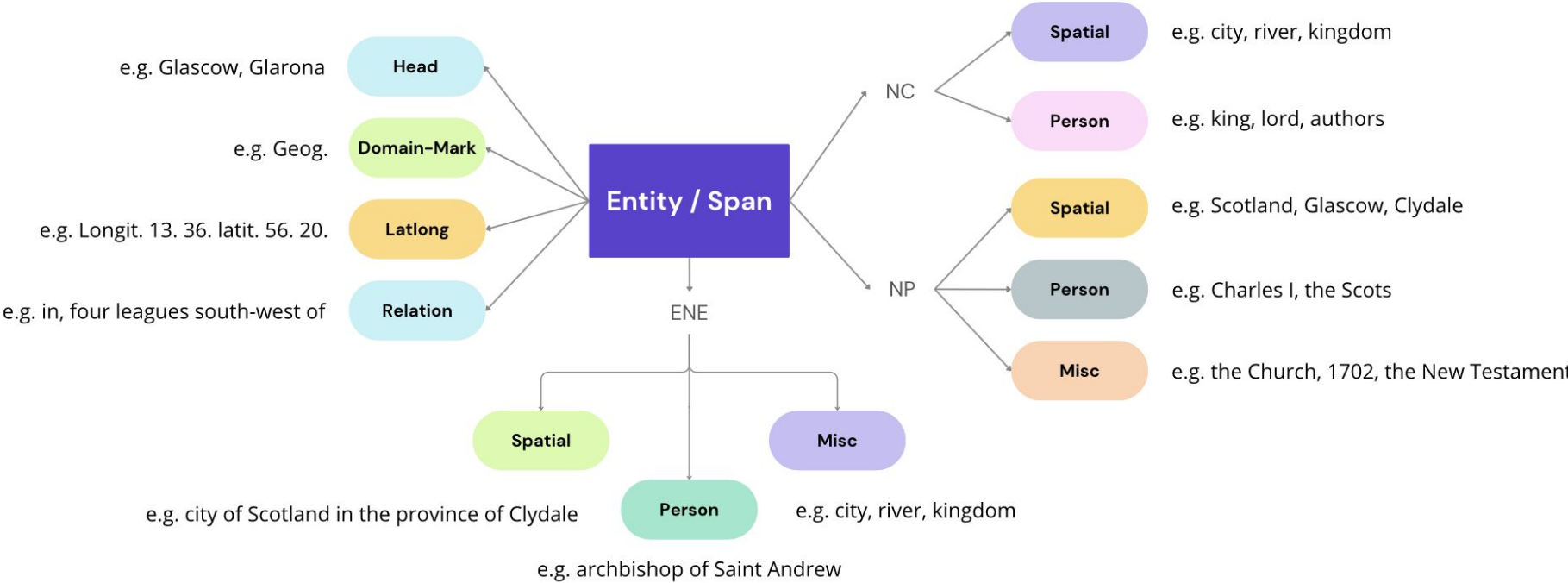


esteemed ecclesiastical history of Scotland, which extends from 203 A.D. to 1624 (D. J.).



A Gold Standard Dataset for Geo-semantic Annotation

Annotation schema



Active learning for dataset labeling

Semi-automatic annotation

- First, manual annotation of few samples (around 40 paragraphs):
 - Number of spans: NP-Spatial: 147, ENE-Spatial: 89, NC-Spatial: 72, Relation: 62, Latlong: 16, NP-Person: 16, NC-Person: 8, ENE-Person: 5, NP-Misc: 3, ENE-Misc: 2
 - Precision: **73.86**, Recall: **58.56**, F-score: **65.33** (on a very small test set)
- Then, the active learning loop:
 - Train spaCy spancat model (with Prodigy)
 - Evaluate model
 - Correct predicted annotations on new paragraphs

ILLESCAS, (Géog.) petite ville d'Espagne, dans la nouvelle Castille, à six lieues au sud de Madrid.

VOLUME: 8 NUMERO: 2637 PARAGRAPH_ID: 1 HEAD: ILLESCAS AUTHOR: unsigned DOMAIN_ARTICLE: Géographie DOMAIN_PARAGRAPH: Géographie

Active learning for dataset labeling

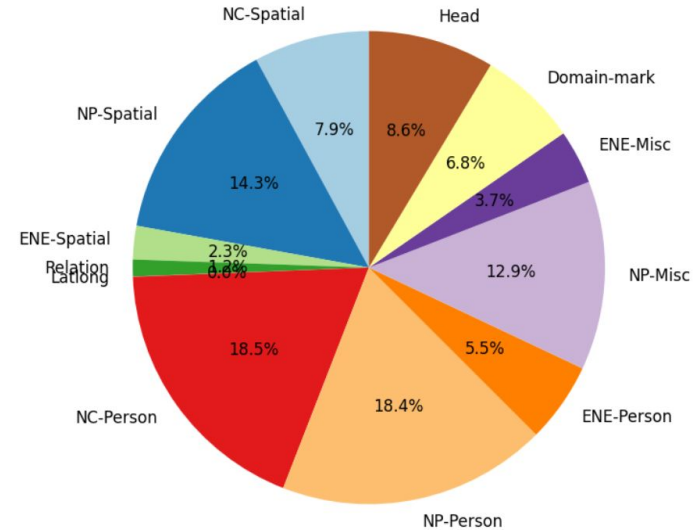
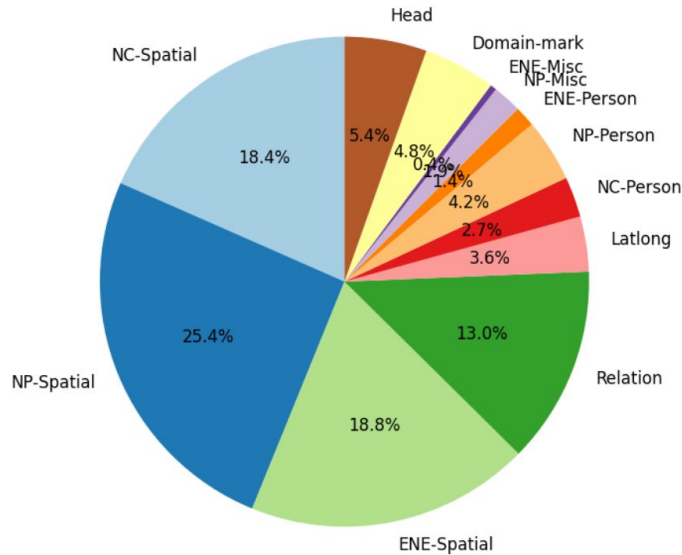
Output data format (JSONL)

```
1 {
2   "text": "KEITH, (Geog.) île de l'Ecosse méridionale, dans la riviere de Forth : elle est fertile en bons pâturages. pour les chevaux. Long. 14. 46. lat. 56. 20. (D. J.)",
3   "meta": {"volume": 9, "head": "KEITH", "author": "Jaucourt", "domain_article": "Géographie", "domain_paragraph": "Géographie", "article": 536, "paragraph": 1},
4   "tokens": [
5     {"text": "KEITH", "start": 0, "end": 5, "id": 0, "ws": false},
6     {"text": ",", "start": 5, "end": 6, "id": 1, "ws": true},
7     {"text": "(", "start": 7, "end": 8, "id": 2, "ws": false},
8     {"text": "Geog", "start": 8, "end": 12, "id": 3, "ws": false},
9     {"text": ".", "start": 12, "end": 13, "id": 4, "ws": false},
10    {"text": ")", "start": 13, "end": 14, "id": 5, "ws": true},
11    {"text": "île", "start": 15, "end": 18, "id": 6, "ws": true},
12    {"text": "de", "start": 19, "end": 21, "id": 7, "ws": true},
13    {"text": "l'", "start": 22, "end": 24, "id": 8, "ws": false},
14    {"text": "Ecosse", "start": 24, "end": 30, "id": 9, "ws": true},
15    {"text": "méridionale", "start": 31, "end": 42, "id": 10, "ws": false},
16    {...}
17  ],
18  "spans": [
19    {"start": 0, "end": 5, "token_start": 0, "token_end": 0, "label": "Head", "text": "KEITH"},
20    {"start": 8, "end": 13, "token_start": 3, "token_end": 4, "label": "Domain-mark", "text": "Geog."},
21    {"start": 15, "end": 18, "text": "île", "token_start": 6, "token_end": 6, "label": "NC-Spatial"},
22    {"start": 22, "end": 42, "text": "l'Ecosse méridionale", "token_start": 8, "token_end": 10, "label": "NP-Spatial"},
23    {"start": 15, "end": 42, "text": "île de l'Ecosse méridionale", "token_start": 6, "token_end": 10, "label": "ENE-Spatial"},
24    {...}
25  ]
26 }
```

Active learning for dataset labeling

Gold standard

- 2,200 randomly selected paragraphs (from 2,001 entries)
- 147,778 tokens
- 25,087 annotated spans



(a) in paragraphs classified under Géographie (b) in paragraphs not classified under Géographie

Spans distribution

Active learning for dataset labeling

Prepare the dataset for machine learning

- train / val / test split

Span	Train	Validation	Test	Span	Train	Validation	Test
NC-Spatial	3,252	358	355	NP-Person	1,599	170	150
NP-Spatial	4,707	464	519	ENE-Person	492	49	57
ENE-Spatial	3,043	326	334	NP-Misc	948	108	96
Relation	2,093	219	226	ENE-Misc	255	31	22
Latlong	553	66	72	Head	1,261	142	153
NC-Person	1,378	132	133	Domain-Mark	1,069	122	133
Paragraphs	1,800	200	200	Spans	20,650	2,187	2,250
Tokens	132,398	14,959	13,881				

Distribution of spans across the datasets

Moncla, L., Vigier, D., & McDonough, K. (2024). *GeoEDdA: A Gold Standard Dataset for Named Entity Recognition and Span Categorization Annotations of Diderot & d'Alembert's Encyclopédie*. Proceedings of the 2nd International Workshop on Geographic Information Extraction from Texts (GeoExT). Glasgow, Scotland.

<https://huggingface.co/datasets/GEODE/GeoEDdA>

Training a Span Categorization Model

spaCy custom spancat

- spaCy span categorizer pipeline: suggester function and labeler model
 - suggester: identifies potentially overlapping entities (i.e. spans)
 - classifier: predicts zero or more labels for each span

Evaluation

Tag	Precision	Recall	F-score	Tag	Precision	Recall	F-score
NC-Spatial	96.50	93.24	94.84	NP-Person	92.47	90.00	91.22
NP-Spatial	92.74	95.95	94.32	ENE-Person	92.16	82.46	87.04
ENE-Spatial	91.67	95.51	93.55	NP-Misc	93.24	71.88	81.18
Relation	97.33	64.60	77.66	ENE-Misc	0.00	0.00	0.00
Latlong	0.00	0.00	0.00	Head	97.37	24.18	38.74
NC-Person	93.07	70.68	80.34	Domain-mark	99.19	91.73	95.31

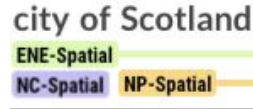
Model performance by span (test set)

Summary of the talk

1. The GeoEDdA dataset: A Gold Standard Dataset for Geo-semantic Annotation
2. BERT fine-tuning for NER
 - collaborative work with H. Zeghidi (Master's intern)
3. Few-shot prompting for NER
4. Perspectives: Knowledge Graph construction using LLMs

BERT fine-tuning for NER

- Nested entities: overlapping spans (one token can have several tags)



Multi-label span classification

- *Scotland* will be tagged as both NP-Spatial and ENE-Spatial

Single-label span classification (joint-label)

- *Scotland* will be tagged as NP-Spatial+ENE-Spatial

BERT fine-tuning for NER

Nested entities (level 1 only, e.g., *city of Scotland*)

	Precision	Recall	F-score
Micro Avg	89.8	92.7	91.2
Macro Avg	84.4	89.5	86.4
Weighted Avg	90.9	92.7	91.6

Multi-label

	Precision	Recall	F-score
Micro Avg	90.1	92.0	91.0
Macro Avg	84.6	87.9	85.9
Weighted Avg	91.0	92.0	91.3

Joint-label

BERT fine-tuning for NER

Nested entities (level 1 only, *e.g., city of Scotland*)

	Precision	Recall	F-score
Micro Avg	89.8	92.7	91.2
Macro Avg	84.4	89.5	86.4
Weighted Avg	90.9	92.7	91.6

Multi-label

	Precision	Recall	F-score
Micro Avg	90.1	92.0	91.0
Macro Avg	84.6	87.9	85.9
Weighted Avg	91.0	92.0	91.3

Joint-label

Nested entities (all levels, *e.g., city of Scotland in the province of Clydale*)

	Precision	Recall	F-score
Micro Avg	87.4	91.0	89.2
Macro Avg	56.3	60.1	58.0
Weighted Avg	87.3	91.0	89.0

Multi-label

	Precision	Recall	F-score
Micro Avg	86.8	90.7	88.7
Macro Avg	62.3	63.5	62.2
Weighted Avg	87.7	90.7	88.9

Joint-label

BERT fine-tuning for NER

Nested entities (level 1 only, *e.g., city of Scotland*)

	Precision	Recall	F-score
Micro Avg	89.8	92.7	91.2
Macro Avg	84.4	89.5	86.4
Weighted Avg	90.9	92.7	91.6

Multi-label

	Precision	Recall	F-score
Micro Avg	90.1	92.0	91.0
Macro Avg	84.6	87.9	85.9
Weighted Avg	91.0	92.0	91.3

Joint-label

Nested entities (all levels, *e.g., city of Scotland in the province of Clydale*)

	Precision	Recall	F-score
Micro Avg	87.4	91.0	89.2
Macro Avg	56.3	60.1	58.0
Weighted Avg	87.3	91.0	89.0

Multi-label

	Precision	Recall	F-score
Micro Avg	86.8	90.7	88.7
Macro Avg	62.3	63.5	62.2
Weighted Avg	87.7	90.7	88.9

Joint-label

BERT fine-tuning for NER

With Nested Entities (only level 1)

Multi-label

	Precision	Recall	F-score	Support
Domain-mark	99.2	99.0	99.1	392
Head	97.2	96.5	96.8	254
NC-Person	63.0	83.1	71.6	225
NC-Spatial	91.6	95.3	93.4	592
NP-Misc	69.6	76.0	72.7	175
NP-Person	86.1	88.7	87.4	203
NP-Spatial	95.6	93.7	94.7	718
Relation	90.8	93.6	92.2	452
Latlong	96.0	98.0	97.0	789
ENE-Misc	40.7	72.8	52.2	81
ENE-Person	88.7	86.4	87.5	199
ENE-Spatial	94.4	90.9	92.6	802

Joint-label

	Precision	Recall	F-score	Support
Domain-mark	99.7	99.0	99.4	392
Head	97.3	98.0	97.6	254
NC-Person	67.2	80.9	73.4	225
NC-Spatial	90.5	95.1	92.8	592
NP-Misc	68.7	76.6	72.4	175
NP-Person	87.9	86.2	87.1	203
NP-Spatial	95.2	95.7	95.4	718
Relation	87.5	93.1	90.2	452
Latlong	98.1	96.8	97.4	789
ENE-Misc	40.0	66.7	50.0	81
ENE-Person	89.5	77.4	83.0	199
ENE-Spatial	94.0	89.7	91.8	802

BERT fine-tuning for NER

With Nested Entities (only level 1)

Multi-label

	Precision	Recall	F-score	Support
Domain-mark	99.2	99.0	99.1	392
Head	97.2	96.5	96.8	254
NC-Person	63.0	83.1	71.6	225
NC-Spatial	91.6	95.3	93.4	592
NP-Misc	69.6	76.0	72.7	175
NP-Person	86.1	88.7	87.4	203
NP-Spatial	95.6	93.7	94.7	718
Relation	90.8	93.6	92.2	452
Latlong	96.0	98.0	97.0	789
ENE-Misc	40.7	72.8	52.2	81
ENE-Person	88.7	86.4	87.5	199
ENE-Spatial	94.4	90.9	92.6	802

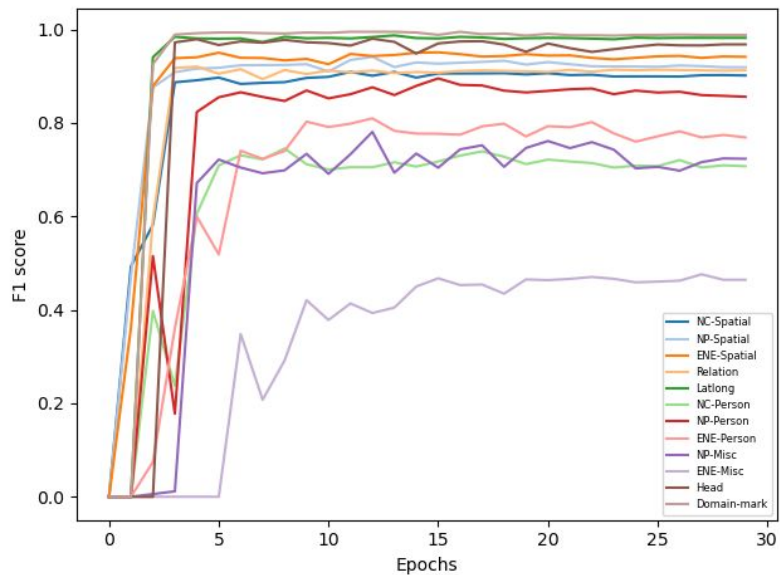
Joint-label

	Precision	Recall	F-score	Support
Domain-mark	99.7	99.0	99.4	392
Head	97.3	98.0	97.6	254
NC-Person	67.2	80.9	73.4	225
NC-Spatial	90.5	95.1	92.8	592
NP-Misc	68.7	76.6	72.4	175
NP-Person	87.9	86.2	87.1	203
NP-Spatial	95.2	95.7	95.4	718
Relation	87.5	93.1	90.2	452
Latlong	98.1	96.8	97.4	789
ENE-Misc	40.0	66.7	50.0	81
ENE-Person	89.5	77.4	83.0	199
ENE-Spatial	94.0	89.7	91.8	802

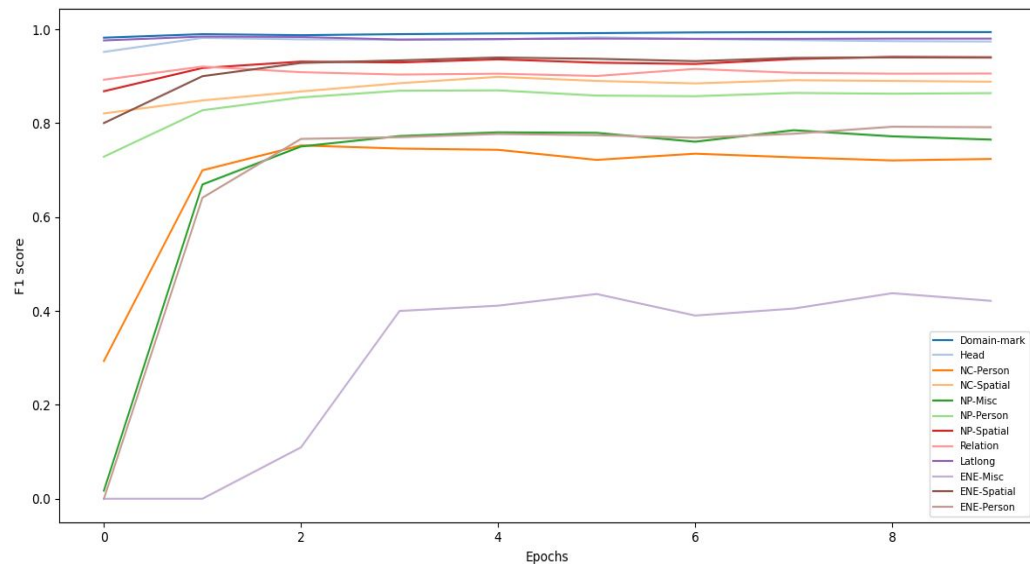
BERT fine-tuning for NER

With Nested Entities (only level 1)

Multi-label



Joint-label



- F1-score per class on validation set during training

BERT fine-tuning for NER

With Nested Entities (all levels)

Multi-label

	Precision	Recall	F-score	Support
ENE-Misc-0	35.1	49.4	41.0	81
ENE-Misc-1	0.00	0.00	0.00	5
ENE-Misc-2	0.00	0.00	0.00	0
ENE-Person-0	88.3	79.9	83.9	199
ENE-Person-1	0.00	0.00	0.00	21
ENE-Person-2	0.00	0.00	0.00	0
ENE-Spatial-0	93.2	92.5	92.9	802
ENE-Spatial-1	84.7	85.4	85.0	685
ENE-Spatial-2	76.1	94.6	84.4	425
ENE-Spatial-3	70.3	88.0	78.2	175
ENE-Spatial-4	0.00	0.00	0.00	40
ENE-Spatial-5	0.00	0.00	0.00	0
ENE-Spatial-6	0.00	0.00	0.00	0

Joint-label

	Precision	Recall	F-score	Support
ENE-Misc-0	42.1	63.0	50.5	81
ENE-Misc-1	0.00	0.00	0.00	5
ENE-Misc-2	0.00	0.00	0.00	0
ENE-Person-0	87.9	80.4	84.0	199
ENE-Person-1	75.0	57.1	64.9	21
ENE-Person-2	0.00	0.00	0.00	0
ENE-Spatial-0	92.5	92.1	92.3	802
ENE-Spatial-1	85.8	86.3	86.0	685
ENE-Spatial-2	72.1	88.9	79.7	425
ENE-Spatial-3	60.6	76.6	67.7	175
ENE-Spatial-4	66.7	25.0	36.4	40
ENE-Spatial-5	0.00	0.00	0.00	0
ENE-Spatial-6	0.00	0.00	0.00	0

BERT fine-tuning for NER

With Nested Entities (all levels)

Multi-label

	Precision	Recall	F-score	Support
ENE-Misc-0	35.1	49.4	41.0	81
ENE-Misc-1	0.00	0.00	0.00	5
ENE-Misc-2	0.00	0.00	0.00	0
ENE-Person-0	88.3	79.9	83.9	199
ENE-Person-1	0.00	0.00	0.00	21
ENE-Person-2	0.00	0.00	0.00	0
ENE-Spatial-0	93.2	92.5	92.9	802
ENE-Spatial-1	84.7	85.4	85.0	685
ENE-Spatial-2	76.1	94.6	84.4	425
ENE-Spatial-3	70.3	88.0	78.2	175
ENE-Spatial-4	0.00	0.00	0.00	40
ENE-Spatial-5	0.00	0.00	0.00	0
ENE-Spatial-6	0.00	0.00	0.00	0

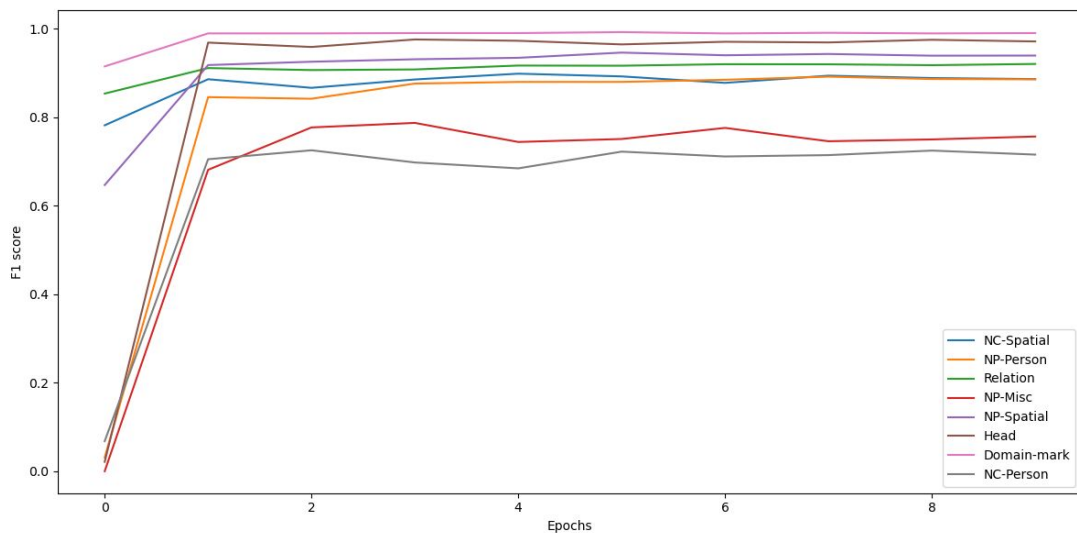
Joint-label

	Precision	Recall	F-score	Support
ENE-Misc-0	42.1	63.0	50.5	81
ENE-Misc-1	0.00	0.00	0.00	5
ENE-Misc-2	0.00	0.00	0.00	0
ENE-Person-0	87.9	80.4	84.0	199
ENE-Person-1	75.0	57.1	64.9	21
ENE-Person-2	0.00	0.00	0.00	0
ENE-Spatial-0	92.5	92.1	92.3	802
ENE-Spatial-1	85.8	86.3	86.0	685
ENE-Spatial-2	72.1	88.9	79.7	425
ENE-Spatial-3	60.6	76.6	67.7	175
ENE-Spatial-4	66.7	25.0	36.4	40
ENE-Spatial-5	0.00	0.00	0.00	0
ENE-Spatial-6	0.00	0.00	0.00	0

BERT fine-tuning for NER

Without Nested Entities

Validation (during training)



Test

	Precision	Recall	F-score	Support
Domain-mark	99.2	100.0	99.6	392
Head	97.6	97.2	97.4	254
NC-Person	67.5	84.0	74.9	225
NC-Spatial	91.6	95.3	93.4	592
NP-Misc	72.4	76.6	74.4	175
NP-Person	87.4	89.2	88.3	203
NP-Spatial	95.9	95.5	95.7	718
Relation	89.4	94.7	91.9	452

	Precision	Recall	F-score
Micro Avg	90.1	93.7	91.9
Macro Avg	87.6	91.6	89.5
Weighted Avg	90.6	93.7	92.1



Comparison with other architectures

	<i>BASE+POS+DEP FEATURES</i>	<i>BASE BERT</i>	<i>SPAN BERT</i>	<i>SPACY SPANCAT</i>	<i>BI-LSTM MODEL</i>	<i>CNN Model</i>
	<i>CRF MODEL NER</i>	<i>WITH ALL TAGS</i>	<i>WITH ALL TAGS</i>		<i>WITH ALL TAGS</i>	<i>WITH ALL TAGS</i>
<i>Domain-mark</i>	99.0	99.9	99.4	95.8	99.1	94.1
<i>Head</i>	87.7	97.5	97.6	45.1	95.4	88.4
<i>Relation</i>	91.0	91.4	91.0	52.5	93.2	79.4
<i>Latlong</i>	-	97.4	97.2	0.00	98.4	95.3
<i>NC-Person</i>	66.0	73.5	76.3	78.0	74.1	61.7
<i>NC-Spatial</i>	89.4	92.6	91.8	95.3	92.9	87.1
<i>NP-Misc</i>	63.8	71.5	74.1	71.9	74.5	52.4
<i>NP-Person</i>	77.4	87.3	87.1	93.0	88.5	73.4
<i>NP-Spatial</i>	91.1	95.4	95.7	95.4	95.5	83.0
<i>ENE-Misc-0</i>	-	50.5	41.0	0.00	39.3	27.7
<i>ENE-Misc-1</i>	-	0.00	0.00	0.00	0.00	0.00
<i>ENE-Person-0</i>	-	84.0	83.9	88.2	79.3	64.4
<i>ENE-Person-1</i>	-	64.9	0.00	41.0	32.0	8.7
<i>ENE-Spatial-0</i>	-	92.3	92.9	94.1	91.6	85.1
<i>ENE-Spatial-1</i>	-	86.0	85.0	89.6	87.0	68.7
<i>ENE-Spatial-2</i>	-	79.7	84.4	87.8	83.4	49.4
<i>ENE-Spatial-3</i>	-	67.7	78.2	78.0	72.5	13.3
<i>ENE-Spatial-4</i>	-	36.4	0.00	51.6	2.1	15.9

GliNER

- Generalist Model for NER using Bidirectional Transformer
- Zero-shot NER
- Not limited to a predefined set of entities
- Lighter than LLMs

	Precision	Recall	F-score
Micro Avg	52.7	45.0	48.6
Macro Avg	40.1	40.6	37.2
Weighted Avg	55.0	45.0	46.4

	Precision	Recall	F-score	Support
NP-Spatial	61.8	72.8	66.9	718
NP-Person	31.4	60.1	41.2	203
NP-Misc	11.3	24.0	15.3	175
Latlong	96.1	46.3	62.4	789
Relation	0.00	0.00	0.00	452

Summary of the talk

1. The GeoEDdA dataset: A Gold Standard Dataset for Geo-semantic Annotation
2. BERT fine-tuning for NER
3. Few-shot prompting for NER
 - collaborative work with H. Zeghidi (Master's intern)
4. Perspectives: Knowledge Graph construction using LLMs

Few-shot prompting for NER

- How effective are few-shot learning techniques with LLMs for performing NER tasks?

Our proposal

- Evaluate few-shot prompting
- With predefined JSON-format output
- Experiment both span and token-based annotations

Few-shot prompting for NER

- A preliminary experiment reveals that LLMs struggle to accurately retrieve or calculate token or span positions from raw input text.
- Although the output format was correct, the generated position values were inaccurate, resembling hallucinations or random numbers rather than referring to actual token positions.
- We propose a solution that includes tokenization information, specifically, token position details for span detection, within the input data.

```
[
  {
    "entities": [
      {
        "text": "COMPIEGNE",
        "start": 0,
        "end": 0,
        "label": "Head"
      },
      {
        "text": "Géog. mod.",
        "start": 3,
        "end": 6,
        "label": "Domain-mark"
      },
      {
        "text": "ville",
        "start": 8,
        "end": 8,
        "label": "NC-Spatial"
      },
      {
        "text": "France",
        "start": 10,
        "end": 10,
        "label": "NP-Spatial"
      },
      {
        "text": "dans",
        "start": 12,
        "end": 12,
        "label": "Relation"
      },
    ]
  }
]
```


Few-shot prompting for NER

Prompt part 1: description of the task and tagset

You are an expert in Natural Language Processing. Your task is to identify common Named Entities (NER) in a given text.

The possible common Named Entities (NER) types are exclusively: (Domain-mark, Head, NC-Person, NC-Spatial, NP-Misc, NP-Person, NP-Spatial, Relation)

:

1.Domain-mark: words indicating the knowledge domain (usually after the head and between parenthesis) such as 'Géog., Géog. mod., Géog. anc., Géogr., ist. nat., Gram., Géogr. anc., Jurisprud., Géog. anc. & mod., Gramm., Geog.'

2.Head: entry name at the beginning of the sentence and is almost always in uppercase such as 'Aire, Afrique, Aigle, ILLESCAS, MULHAUSEN, ADDA, SINTRU AKSTEDE, KEITH, CAÇERES, CARMAGNOLE, AGRIGNON, INSPRUCK'

3.NC-Person: a common noun that identifies a person such as 'M., roi, S., peuples, l'empereur, son fils, les habitants, prince, peuple, le roi, fils,

4.NC-Spatial: a common noun that identifies a spatial entity including natural features such as 'ville, petite ville, la riviere, la mer, royaume, la ville, l'île, cette ville, pays, la côte, riviere'

5.NP-Misc: a proper noun identifying entities not classified as spatial or person such as 'l'Eglise, grec, 1707, russe, Glaciale, Noire, romain, la nt, 1693, Sud, 1614'

6.NP-Person: a proper noun identifying the name of a person (person named entities) such as 'Ptolomée, Plin, Strabon, Euripide, les Romains, Pierre, urcs, Dieu, César, Antonin, les Espagnols'

7.NP-Spatial: a proper noun identifying the name of a place (spatial named entities) such as 'France, Allemagne, Italie, Espagne, Afrique, Asie, Parome, Russie, la Chine, l'Amérique méridionale'

8.Relation: spatial relation such as 'dans, sur, au, en, entre, près de, se jette dans, proche, par, vers, près du, jusqu'à, à l'orient'.

9.Latlong: geographic coordinates such as 'Long. 31. 58. lat. 40. 55, Long. 10. 27. lat. 43. 30, Long. 28. 14. lat. 51. 13, Long. 14. 46. lat. 56. 20 15, Long. 25. 20. lat. 44. 43, Lat. 19. 40, Long. selon Harris, 29. 16. 15. lat. 47. 15, Long. 14. 28. lat : 53. 50, Long. 57. lat. 38. 35, Long. 22.

11. 18. lat. 40. 41, Long. 27. 40. lat. 51. 50'.

Few-shot prompting for NER

Prompt part 2: example (input and output)

Here are some examples:

EXAMPLE 1:

```
INPUT:('PIRANO' ,0) (' ,' ,1) ('(' ,2) ('Géog' ,3) ('.' ,4) ('mod' ,5) ('.' ,6) (')' ,7) ('ville' ,8) ('d'' ,9) ('Italie' ,10) ('dans' ,11) ('l'' ,12) ('Istrie' ,13) (' ,' ,14) ('environ' ,15) ('à' ,16) ('14' ,17) ('milles' ,18) ('de' ,19) ('Capo' ,20) ('d'' ,21) ('Istria' ,22) (' ,' ,23) ('en' ,24) ('tirant' ,25) ('vers' ,26) ('le' ,27) ('midi' ,28) ('occidental' ,29) ('.' ,30) ('Elle' ,31) ('est' ,32) ('sur' ,33) ('une' ,34) ('petite' ,35) ('presqu'' ,36) ('île' ,37) ('formée' ,38) ('par' ,39) ('le' ,40) ('golfe' ,41) ('Largone' ,42) (' ,' ,43) ('&' ,44) ('celui' ,45) ('de' ,46) ('Trieste' ,47) ('.' ,48) ('Les' ,49) ('Vénitiens' ,50) ('en' ,51) ('sont' ,52) ('les' ,53) ('maîtres' ,54) ('depuis' ,55) ('1583' ,56) ('.' ,57) ('Long' ,58) ('.' ,59) ('31' ,60) ('.' ,61) ('46' ,62) ('.' ,63) ('lat' ,64) ('.' ,65) ('45' ,66) ('.' ,67) ('48' ,68) ('.' ,69)
```

```
OUTPUT:[{'label': 'Head', 'text': 'PIRANO', 'start': 0, 'end': 0}, {'label': 'Domain-mark', 'text': 'Géog. mod.', 'start': 3, 'end': 6}, {'label': 'NC-Spatial', 'text': 'ville', 'start': 8, 'end': 8}, {'label': 'NP-Spatial', 'text': 'Italie', 'start': 10, 'end': 10}, {'label': 'Relation', 'text': 'dans', 'start': 11, 'end': 11}, {'label': 'NP-Spatial', 'text': "l'Istrie", 'start': 12, 'end': 13}, {'label': 'Relation', 'text': 'environ à 14 milles de', 'start': 15, 'end': 19}, {'label': 'NP-Spatial', 'text': "Capo d'Istria", 'start': 20, 'end': 22}, {'label': 'Relation', 'text': 'vers le midi occidental', 'start': 26, 'end': 29}, {'label': 'NC-Spatial', 'text': "une petite presqu'île", 'start': 34, 'end': 37}, {'label': 'Relation', 'text': 'formée par', 'start': 38, 'end': 39}, {'label': 'NC-Spatial', 'text': 'le golfe', 'start': 40, 'end': 41}, {'label': 'NP-Spatial', 'text': 'Largone', 'start': 42, 'end': 42}, {'label': 'NP-Spatial', 'text': 'Trieste', 'start': 47, 'end': 47}, {'label': 'NP-Person', 'text': 'Les Vénitiens', 'start': 49, 'end': 50}, {'label': 'NC-Person', 'text': 'les maîtres', 'start': 53, 'end': 54}, {'label': 'NP-Misc', 'text': '1583', 'start': 56, 'end': 56}, {'label': 'Latlong', 'text': 'Long. 31. 46. lat. 45. 48', 'start': 58, 'end': 68}]
```

Few-shot prompting for NER

Prompt part 2: example (input and output)

Here are some examples:

EXAMPLE 1:
INPUT:('PIRANO',0) ('',1) ('',2) ('Géog',3) ('.',4) ('mod',5) ('.',6) ('',7) ('ville',8) ('d',9) ('Italie',10) ('dans',11) ('l',12) ('Istrie',13) ('',14) ('environ',15) ('à',16) ('14',17) ('milles',18) ('de',19) ('Capo',20) ('d',21) ('Istria',22) ('',23) ('en',24) ('tirant',25) ('vers',26) ('le',27) ('midi',28) ('occidental',29) ('.',30) ('Elle',31) ('est',32) ('sur',33) ('une',34) ('petite',35) ('presqu',36) ('île',37) ('formée',38) ('par',39) ('le',40) ('golfe',41) ('Largone',42) ('',43) ('&',44) ('celui',45) ('de',46) ('Trieste',47) ('.',48) ('Les',49) ('Vénitiens',50) ('en',51) ('sont',52) ('les',53) ('maîtres',54) ('depuis',55) ('1583',56) ('.',57) ('Long',58) ('.',59) ('31',60) ('.',61) ('46',62) ('.',63) ('lat',64) ('.',65) ('45',66) ('.',67) ('48',68) ('.',69)

OUTPUT:[{'label': 'Head', 'text': 'PIRANO', 'start': 0, 'end': 0}, {'label': 'Domain-mark', 'text': 'Géog. mod.', 'start': 3, 'end': 6}, {'label': 'NC-Spatial', 'text': 'ville', 'start': 8, 'end': 8}, {'label': 'NP-Spatial', 'text': 'Italie', 'start': 10, 'end': 10}, {'label': 'Relation', 'text': 'dans', 'start': 11, 'end': 11}, {'label': 'NP-Spatial', 'text': "l'Istrie", 'start': 12, 'end': 13}, {'label': 'Relation', 'text': 'environ à 14 milles de', 'start': 15, 'end': 19}, {'label': 'NP-Spatial', 'text': "Capo d'Istria", 'start': 20, 'end': 22}, {'label': 'Relation', 'text': 'vers le midi occidental', 'start': 26, 'end': 29}, {'label': 'NC-Spatial', 'text': "une petite presqu'île", 'start': 34, 'end': 37}, {'label': 'Relation', 'text': 'formée par', 'start': 38, 'end': 39}, {'label': 'NC-Spatial', 'text': 'le golfe', 'start': 40, 'end': 41}, {'label': 'NP-Spatial', 'text': 'Largone', 'start': 42, 'end': 42}, {'label': 'NP-Spatial', 'text': 'Trieste', 'start': 47, 'end': 47}, {'label': 'NP-Person', 'text': 'Les Vénitiens', 'start': 49, 'end': 50}, {'label': 'NC-Person', 'text': 'les maîtres', 'start': 53, 'end': 54}, {'label': 'NP-Misc', 'text': '1583', 'start': 56, 'end': 56}, {'label': 'Latlong', 'text': 'Long. 31. 46. lat. 45. 48', 'start': 58, 'end': 68}]

Few-shot prompting for NER

Different types of errors

GPT input

```
('HILPERHAUSEN',0) ('',1) ((' ',2) ('Géog.',3) ('.',4) (' ',5) ('ville',6) ('d',7) ('Allemagne',8) ('en',9) ('Franconie',10) ('',11) ('sur',12) ('la',13) ('Werra',14) ('',15) ('au',16) ('comté',17) ('de',18) ('Henneberg',19) ('',20) ('entre',21) ('Cobourg',22) ('&',23) ('Smalcalde',24) (';',25) ('elle',26) ('appartient',27) ('à',28) ('une',29) ('branche',30) ('de',31) ('la',32) ('maison',33) ('de',34) ('Saxe',35) ('-',36) ('Gotha',37) ('.',38) ('Long',39) ('.',40) ('28',41) ('.',42) ('15',43) ('.',44) ('lat',45) ('.',46) ('50',47) ('.',48) ('35',49) ('.',50) ('(',51) ('D.',52) ('J.',53) (' ',54)
```

```
{'entities': [{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'}, {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'}, {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'}, {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'}, {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'}, {'text': 'Werra', 'start': 14, 'end': 14, 'label': 'NC-Spatial'}, {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'}, {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'}, {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'}, {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'}, {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'}, {'text': 'Long.', 'start': 39, 'end': 40, 'label': 'NP-Misc'}, {'text': '28.15', 'start': 41, 'end': 43, 'label': 'NP-Misc'}, {'text': 'lat.', 'start': 45, 'end': 46, 'label': 'Latlong'}, {'text': '50.35', 'start': 47, 'end': 49, 'label': 'Latlong'}, {'text': 'D.J.', 'start': 52, 'end': 53, 'label': 'Latlong'}]}
```

GPT output

```
[{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'}, {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'}, {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'}, {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'}, {'text': 'en', 'start': 9, 'end': 9, 'label': 'Relation'}, {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'}, {'text': 'sur', 'start': 12, 'end': 12, 'label': 'Relation'}, {'text': 'la Werra', 'start': 13, 'end': 14, 'label': 'NP-Spatial'}, {'text': 'au', 'start': 16, 'end': 16, 'label': 'Relation'}, {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'}, {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'}, {'text': 'entre', 'start': 21, 'end': 21, 'label': 'Relation'}, {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'}, {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'}, {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'}, {'text': 'Long. 28. 15. lat. 50. 35', 'start': 39, 'end': 49, 'label': 'Latlong'}]
```

Ground truth

Few-shot prompting for NER

Different types of errors

GPT input

```
('HILPERHAUSEN',0) ('',1) ((' ',2) ('Géog.',3) ('.',4) (' ',5) ('ville',6) ('d',7) ('Allemagne',8) ('en',9) ('Franconie',10) ('',11) ('sur',12) ('la',13) ('Werra',14) ('',15) ('au',16) ('comté',17) ('de',18) ('Henneberg',19) ('',20) ('entre',21) ('Cobourg',22) ('&',23) ('Smalcalde',24) (';',25) ('elle',26) ('appartient',27) ('à',28) ('une',29) ('branche',30) ('de',31) ('la',32) ('maison',33) ('de',34) ('Saxe',35) ('-',36) ('Gotha',37) ('.',38) ('Long',39) ('.',40) ('28',41) ('.',42) ('15',43) ('.',44) ('lat',45) ('.',46) ('50',47) ('.',48) ('35',49) ('.',50) ('(',51) ('D.',52) ('J.',53) ('',54)
```

```
{'entities': [{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'},  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'},  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'},  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'},  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'},  
  {'text': 'Werra', 'start': 14, 'end': 14, 'label': 'NC-Spatial'},  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'},  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'},  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'},  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'},  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'},  
  {'text': 'Long.', 'start': 39, 'end': 40, 'label': 'NP-Misc'},  
  {'text': '28.15', 'start': 41, 'end': 43, 'label': 'NP-Misc'},  
  {'text': 'lat.', 'start': 45, 'end': 46, 'label': 'Latlong'},  
  {'text': '50.35', 'start': 47, 'end': 49, 'label': 'Latlong'},  
  {'text': 'D.J.', 'start': 52, 'end': 53, 'label': 'Latlong'}]}
```

GPT output

```
[{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'},  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'},  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'},  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'},  
  {'text': 'en', 'start': 9, 'end': 9, 'label': 'Relation'},  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'},  
  {'text': 'sur', 'start': 12, 'end': 12, 'label': 'Relation'},  
  {'text': 'la Werra', 'start': 13, 'end': 14, 'label': 'NP-Spatial'},  
  {'text': 'au', 'start': 16, 'end': 16, 'label': 'Relation'},  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'},  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'},  
  {'text': 'entre', 'start': 21, 'end': 21, 'label': 'Relation'},  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'},  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'},  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'},  
  {'text': 'Longe. 28. 15. lat. 50. 35', 'start': 39, 'end': 49, 'label': 'Latlong'}]
```

Ground truth

Entities not found

Few-shot prompting for NER

Different types of errors

GPT input

```
('HILPERHAUSEN',0) ('',1) ((' ',2) ('Géog.',3) ('.',4) (' ',5) ('ville',6) ('d',7) ('Allemagne',8) ('en',9) ('Franconie',10) ('',11) ('sur',12) ('la',13) ('Werra',14) ('',15) ('au',16) ('comté',17) ('de',18) ('Henneberg',19) ('',20) ('entre',21) ('Cobourg',22) ('&',23) ('Smalcalde',24) (';',25) ('elle',26) ('appartient',27) ('à',28) ('une',29) ('branche',30) ('de',31) ('la',32) ('maison',33) ('de',34) ('Saxe',35) ('-',36) ('Gotha',37) ('.',38) ('Long',39) ('.',40) ('28',41) ('.',42) ('15',43) ('.',44) ('lat',45) ('.',46) ('50',47) ('.',48) ('35',49) ('.',50) ('(',51) ('D.',52) ('J.',53) ('',54)
```

```
{'entities': [{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'},  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'},  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'},  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'},  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'},  
  {'text': 'Werra', 'start': 14, 'end': 14, 'label': 'NC-Spatial'},  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'},  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'},  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'},  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'},  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'},  
  {'text': 'Long.', 'start': 39, 'end': 40, 'label': 'NP-Misc'},  
  {'text': '28.15', 'start': 41, 'end': 43, 'label': 'NP-Misc'},  
  {'text': 'lat.', 'start': 45, 'end': 46, 'label': 'Latlong'},  
  {'text': '50.35', 'start': 47, 'end': 49, 'label': 'Latlong'},  
  {'text': 'D.J.', 'start': 52, 'end': 53, 'label': 'Latlong'}]}
```

GPT output

```
[{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'},  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'},  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'},  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'},  
  {'text': 'en', 'start': 9, 'end': 9, 'label': 'Relation'},  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'},  
  {'text': 'sur', 'start': 12, 'end': 12, 'label': 'Relation'},  
  {'text': 'la Werra', 'start': 13, 'end': 14, 'label': 'NP-Spatial'},  
  {'text': 'au', 'start': 16, 'end': 16, 'label': 'Relation'},  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'},  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'},  
  {'text': 'entre', 'start': 21, 'end': 21, 'label': 'Relation'},  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'},  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'},  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'},  
  {'text': 'Long. 28. 15. lat. 50. 35', 'start': 39, 'end': 49, 'label': 'Latlong'}]
```

Ground truth

Wrong boundaries

Few-shot prompting for NER

Different types of errors

GPT input

```
('HILPERHAUSEN',0) ('',1) ('(',2) ('Géog.',3) ('.',4) (')',5) ('ville',6) ('d',7) ('Allemagne',8) ('en',9) ('Franconie',10) ('',11) ('sur',12) ('la',13) ('Werra',14) ('',15) ('au',16) ('comté',17) ('de',18) ('Henneberg',19) ('',20) ('entre',21) ('Cobourg',22) ('&',23) ('Smalcalde',24) (';',25) ('elle',26) ('appartient',27) ('à',28) ('une',29) ('branche',30) ('de',31) ('la',32) ('maison',33) ('de',34) ('Saxe',35) ('-',36) ('Gotha',37) ('.',38) ('Long',39) ('.',40) ('28',41) ('.',42) ('15',43) ('.',44) ('lat',45) ('.',46) ('50',47) ('.',48) ('35',49) ('.',50) ('(',51) ('D.',52) ('J.',53) (')',54)
```

```
{'entities': [{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'},  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'},  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'},  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'},  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'},  
  {'text': 'Werra', 'start': 14, 'end': 14, 'label': 'NC-Spatial'},  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'},  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'},  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'},  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'},  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'},  
  {'text': 'Long.', 'start': 39, 'end': 40, 'label': 'NP-Misc'},  
  {'text': '28.15', 'start': 41, 'end': 43, 'label': 'NP-Misc'},  
  {'text': 'lat.', 'start': 45, 'end': 46, 'label': 'Latlong'},  
  {'text': '50.35', 'start': 47, 'end': 49, 'label': 'Latlong'},  
  {'text': 'D.J.', 'start': 52, 'end': 53, 'label': 'Latlong'}]}
```

GPT output

```
[{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'},  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'},  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'},  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'},  
  {'text': 'en', 'start': 9, 'end': 9, 'label': 'Relation'},  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'},  
  {'text': 'sur', 'start': 12, 'end': 12, 'label': 'Relation'},  
  {'text': 'la Werra', 'start': 13, 'end': 14, 'label': 'NP-Spatial'},  
  {'text': 'au', 'start': 16, 'end': 16, 'label': 'Relation'},  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'},  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'},  
  {'text': 'entre', 'start': 21, 'end': 21, 'label': 'Relation'},  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'},  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'},  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'},  
  {'text': 'Long. 15. lat. 50. 35', 'start': 39, 'end': 49, 'label': 'Latlong'}]
```

Ground truth

Wrong classification

Few-shot prompting for NER

Different types of errors

GPT input

```
('HILPERHAUSEN',0) ('.',1) ((' ',2) ('Géog.',3) ('.',4) (' ',5) ('ville',6) ('d',7) ('Allemagne',8) ('en',9) ('Franconie',10) ('.',11) ('sur',12) ('la',13) ('Werra',14) ('.',15) ('au',16) ('comté',17) ('de',18) ('Henneberg',19) ('.',20) ('entre',21) ('Cobourg',22) ('&',23) ('Smalcalde',24) (';',25) ('elle',26) ('appartient',27) ('à',28) ('une',29) ('branche',30) ('de',31) ('la',32) ('maison',33) ('de',34) ('Saxe',35) ('-',36) ('Gotha',37) ('.',38) ('Long',39) ('.',40) ('28',41) ('.',42) ('15',43) ('.',44) ('lat',45) ('.',46) ('50',47) ('.',48) ('35',49) ('.',50) ('(',51) ('D.',52) ('J.',53) (' ',54)
```

```
{'entities': [{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'}, ✓  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'}, ✓  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'}, ✓  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'}, ✓  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'}, ✓  
  {'text': 'Werra', 'start': 14, 'end': 14, 'label': 'NC-Spatial'}, ✗  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'}, ✓  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'}, ✓  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'}, ✓  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'}, ✓  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'}, ✓  
  {'text': 'Long.', 'start': 39, 'end': 40, 'label': 'NP-Misc'}, ✗  
  {'text': '28.15', 'start': 41, 'end': 43, 'label': 'NP-Misc'}, ✗  
  {'text': 'lat.', 'start': 45, 'end': 46, 'label': 'Latlong'}, ✗  
  {'text': '50.35', 'start': 47, 'end': 49, 'label': 'Latlong'}, ✗  
  {'text': 'D.J.', 'start': 52, 'end': 53, 'label': 'Latlong'}] ✗
```

GPT output

```
[{'text': 'HILPERHAUSEN', 'start': 0, 'end': 0, 'label': 'Head'},  
  {'text': 'Géog.', 'start': 3, 'end': 4, 'label': 'Domain-mark'},  
  {'text': 'ville', 'start': 6, 'end': 6, 'label': 'NC-Spatial'},  
  {'text': 'Allemagne', 'start': 8, 'end': 8, 'label': 'NP-Spatial'},  
  {'text': 'en', 'start': 9, 'end': 9, 'label': 'Relation'}, ✗  
  {'text': 'Franconie', 'start': 10, 'end': 10, 'label': 'NP-Spatial'},  
  {'text': 'sur', 'start': 12, 'end': 12, 'label': 'Relation'}, ✗  
  {'text': 'la Werra', 'start': 13, 'end': 14, 'label': 'NP-Spatial'},  
  {'text': 'au', 'start': 16, 'end': 16, 'label': 'Relation'}, ✗  
  {'text': 'comté', 'start': 17, 'end': 17, 'label': 'NC-Spatial'},  
  {'text': 'Henneberg', 'start': 19, 'end': 19, 'label': 'NP-Spatial'},  
  {'text': 'entre', 'start': 21, 'end': 21, 'label': 'Relation'},  
  {'text': 'Cobourg', 'start': 22, 'end': 22, 'label': 'NP-Spatial'},  
  {'text': 'Smalcalde', 'start': 24, 'end': 24, 'label': 'NP-Spatial'},  
  {'text': 'Saxe-Gotha', 'start': 35, 'end': 37, 'label': 'NP-Spatial'},  
  {'text': 'Long. 28. 15. lat. 50. 35', 'start': 39, 'end': 49, 'label': 'Latlong'}]
```

Ground truth

Entities not found
Wrong classification
Wrong boundaries

Few-shot prompting for NER

Impact of the output format



Span-based

```
[ {  
  {  
    "entities": [  
      {  
        "text": "COMPIEGNE",  
        "start": 0,  
        "end": 0,  
        "label": "Head"  
      },  
      {  
        "text": "Géog. mod.",  
        "start": 3,  
        "end": 6,  
        "label": "Domain-mark"  
      },  
      {  
        "text": "ville",  
        "start": 8,  
        "end": 8,  
        "label": "NC-Spatial"  
      },  
      {  
        "text": "France",  
        "start": 10,  
        "end": 10,  
        "label": "NP-Spatial"  
      },  
      {  
        "text": "dans",  
        "start": 12,  
        "end": 12,  
        "label": "Relation"  
      },  
    ]  
  }  
}
```

Token-based

```
[ {  
  {  
    "entities": [  
      {  
        "text": "COMPIEGNE",  
        "label": "Head"  
      },  
      {  
        "text": ",",  
        "label": "0"  
      },  
      {  
        "text": "(",  
        "label": "0"  
      },  
      {  
        "text": "Géog",  
        "label": "Domain-mark"  
      },  
      {  
        "text": ".",  
        "label": "Domain-mark"  
      },  
      {  
        "text": "mod",  
        "label": "Domain-mark"  
      },  
      {  
        "text": ".",  
        "label": "Domain-mark"  
      },  
      {  
        "text": ")",  
        "label": "0"  
      },  
      {  
        "text": "ville",  
        "label": "NC-Spatial"  
      },  
    ]  
  }  
}
```

Few-shot prompting for NER

Impact of the output format

Span-based evaluation

- Unusual issues compared to traditional NER arise as the output format is more complex than at the token level.
- The experiment confirmed the performance improvement between GPT versions:

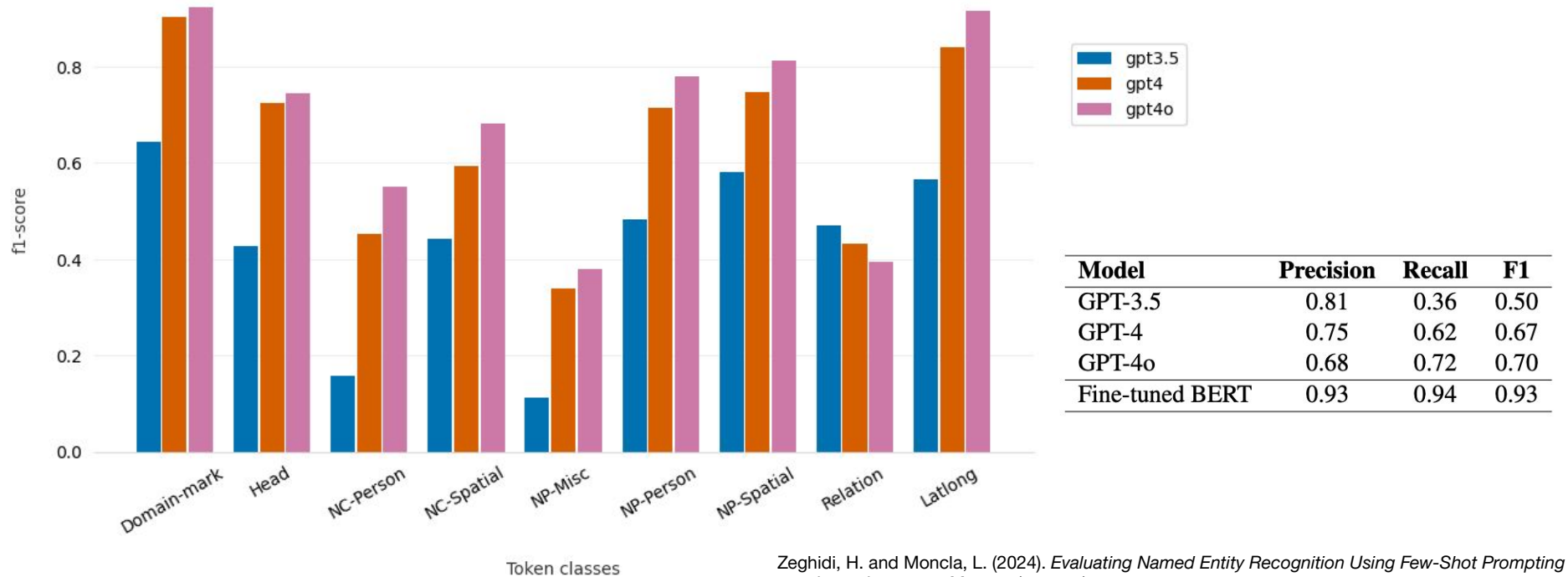
	GPT-3.5	GPT-4o
Correct predicted spans (both boundaries and labels)	28%	49%
Partially correct (partial boundaries and correct labels)	13%	9%
Total	41%	58%

- Using GPT-3.5, some answers do not refer to the input document but rather to the example from the prompt.
- In very few cases, some answers do not strictly follow the predefined JSON output format and some token attributes may be missing (such as start, end, text, or label).

Few-shot prompting for NER

OpenAI GPTs

Token-based evaluation



Zeghidi, H. and Moncla, L. (2024). *Evaluating Named Entity Recognition Using Few-Shot Prompting with Large Language Models*. (preprint)

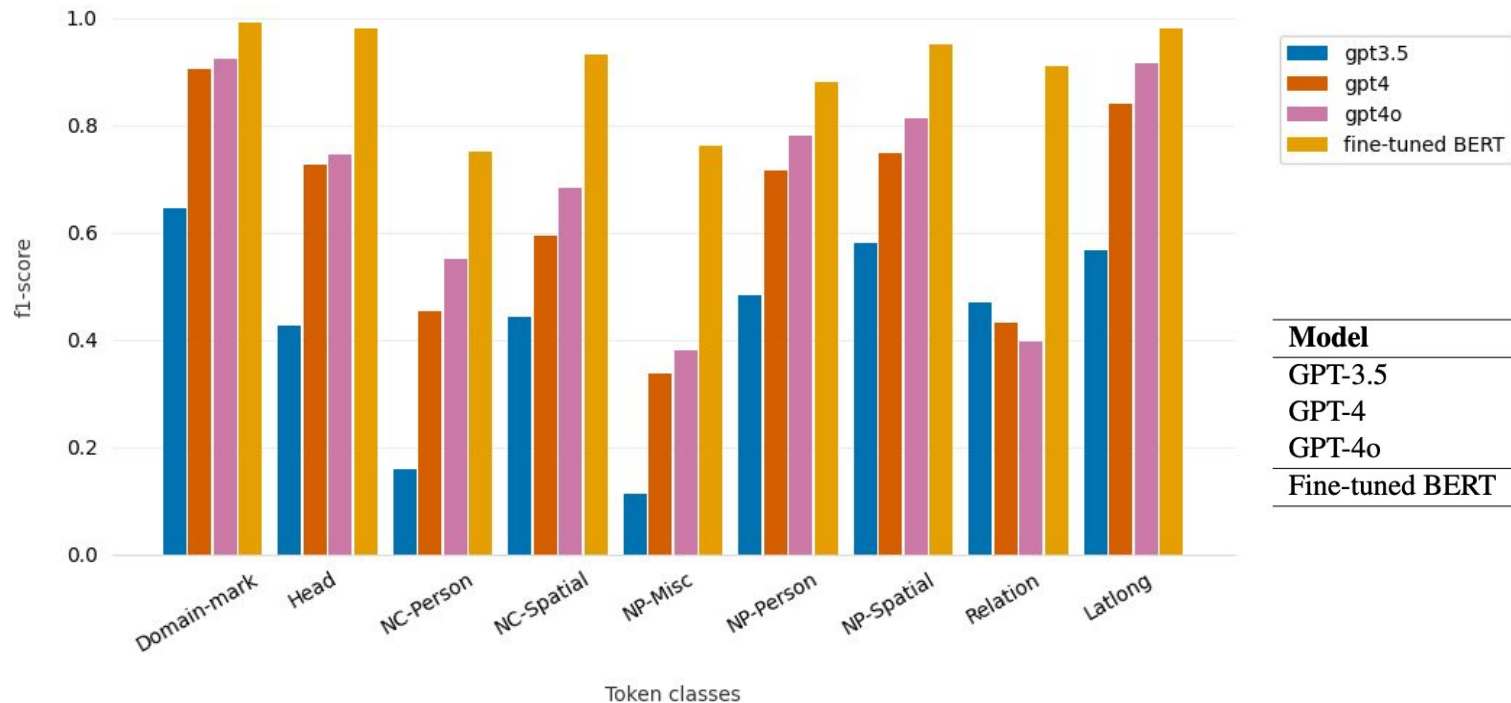
<https://arxiv.org/abs/2408.15796>

GitHub: <https://github.com/GEODE-project/ner-llm>

Few-shot prompting for NER

OpenAI GPTs

Token-based evaluation



Model	Precision	Recall	F1
GPT-3.5	0.81	0.36	0.50
GPT-4	0.75	0.62	0.67
GPT-4o	0.68	0.72	0.70
Fine-tuned BERT	0.93	0.94	0.93

Few-shot prompting for NER

What is the performance of “small” local LLMs

- Llama 3.1 - 8B (Meta),
- Phi3 mini - 4k (Microsoft),
- Gemma 2 - 9B (Google),
- Mistral v0.2 - 7B (MistralAI),
- Qwen 2 - 7B (Alibaba Group)

Experiments were conducted using LM Studio on two configurations: an Nvidia RTX 3500 ADA GPU and an Apple M1 Max with 32GB of RAM.

- Some LLMs provide the correct JSON output syntax but don't really understand the defined set of labels.
- Others completely don't understand the task and do anything or simply repeat the input sentence

Model RAM Usage 4.51 GB
RAM Usage 4.51 GB CPU: 0.00 %
lmstudio-community - Meta Llama 3 1 Instruct 7B Q4_K_M gguf

New Chat (2/10) Export Llama 7B Q4_K_M GGUF plaintext markdown monospace

Settings Reset to Default Settings

Preset 1 change 0
Llama 3 v2

System Prompt
You are an expert in Natural Language Processing. Your task is to identify common Named Entities (NER) in a given text. The possible common Named Entities (NER) types are exclusively: Domain-mark, Head, NC-Person, NC-Spatial, NP-Misc, NP-Person, NP-Spatial, Relation) and can be described as: 1.Domain-marks words indicating the knowledge domain (usually after the head and between parenthesis) such as 'Géog., Géog. mod., Géog. anc., Géogr., Géogr. mod., Marine, Hist nat., Gram., Géogr. anc., Jurisprud., Géog. anc. & mod., Gramm., Geog.' 2.Head: entry name at the beginning of the sentence and is almost always in uppercase such as Amr, Afrique, Algèr, LLESZAS, MULHAUSEN, ADDA, SINTRA ou CINTRA, ACHSTEDE, ou AKSTEDE, KEITH, CACERES, CARMAGNOLE, AGRIGNON, INSPRUCK) 3.NC-Person: a common noun that identifies a person such as "M. rol S. peoples, l'empereur, son fils, les habitants, prince, peuple, la roi, fils, le P. habitants" 4.NC-Spatial: a common noun that identifies a

Here is the parsed output with named entities:

```
[{"label": "Geography", "text": "Gènes", "start": 0, "end": 8}, {"label": "Geography", "text": "Genoa", "start": 1, "end": 5}, {"label": "City", "text": "Genoa", "start": 6, "end": 11}, {"label": "Country", "text": "Italy", "start": 12, "end": 16}, {"label": "Building", "text": "palais", "start": 17, "end": 26}, {"label": "Organization", "text": "république", "start": 21, "end": 29}, {"label": "Location", "text": "Méditerranée", "start": 38, "end": 41}, {"label": "City", "text": "Milan", "start": 42, "end": 47}, {"label": "Country", "text": "Italy", "start": 48, "end": 52}, {"label": "Location", "text": "Turin", "start": 53, "end": 58}, {"label": "City", "text": "Parme", "start": 59, "end": 64}, {"label": "Country", "text": "Italy", "start": 65, "end": 69}, {"label": "Location", "text": "Florence", "start": 70, "end": 78}, {"label": "City", "text": "Rome", "start": 79, "end": 83}, {"label": "Country", "text": "Italy", "start": 84, "end": 88}, {"label": "Location", "text": "Salvego", "start": 89, "end": 96}, {"label": "Person", "text": "Cassini", "start": 97, "end": 102}, {"label": "Person", "text": "Grimaldi", "start": 103, "end": 110}, {"label": "Location", "text": "Milan", "start": 111, "end": 116}, {"label": "City", "text": "Turin", "start": 117, "end": 122}, {"label": "City", "text": "Parme", "start": 123, "end": 128}, {"label": "Location", "text": "Florence", "start": 129, "end": 136}, {"label": "City", "text": "Rome", "start": 137, "end": 142}
```

Note that the entities are not necessarily in the correct order. You may need to sort them based on their start and end indices to get the desired output.

0.231 time to first token: 19.84s gen t: 45.49s speed: 14.39 tok/s stop reason: eosFound gpu layers: 10 cpu threads: 4 mlock: true token count: 2845/2048

Model Downloads 0 downloading - 5 completed

Meta Llama 3.1 - 7B

```
[
  {
    'label': 'Geography', 'text': 'Gènes', 'start': 0, 'end': 0},
    'label': 'Geography', 'text': 'Genua', 'start': 1, 'end': 5},
    'label': 'City', 'text': 'Genoa', 'start': 6, 'end': 11},
    'label': 'Country', 'text': 'Italy', 'start': 12, 'end': 16},
    'label': 'Building', 'text': 'palais', 'start': 17, 'end': 20},
    'label': 'Organization', 'text': 'république', 'start': 21, 'end': 29},
    'label': 'Location', 'text': 'Méditerranée', 'start': 30, 'end': 41},
    'label': 'City', 'text': 'Milan', 'start': 42, 'end': 47},
    'label': 'Country', 'text': 'Italy', 'start': 48, 'end': 52},
    'label': 'Location', 'text': 'Turin', 'start': 53, 'end': 58},
    'label': 'City', 'text': 'Parme', 'start': 59, 'end': 64},
    'label': 'Country', 'text': 'Italy', 'start': 65, 'end': 69},
    'label': 'Location', 'text': 'Florence', 'start': 70, 'end': 78},
    'label': 'City', 'text': 'Rome', 'start': 79, 'end': 83},
    'label': 'Country', 'text': 'Italy', 'start': 84, 'end': 88},
    'label': 'Location', 'text': 'Salvego', 'start': 89, 'end': 96},
    'label': 'Person', 'text': 'Cassini', 'start': 97, 'end': 102},
    'label': 'Person', 'text': 'Grimaldi', 'start': 103, 'end': 110},
    'label': 'Location', 'text': 'Milan', 'start': 111, 'end': 116},
    'label': 'City', 'text': 'Turin', 'start': 117, 'end': 122},
    'label': 'City', 'text': 'Parme', 'start': 123, 'end': 128},
    'label': 'Location', 'text': 'Florence', 'start': 129, 'end': 136},
    'label': 'City', 'text': 'Rome', 'start': 137, 'end': 142}
]
```

- Correct JSON output format
- Does not consider the tagset given in the prompt

Meta Llama 3.1 - 7B

```
[
  { 'label': 'Geography', 'text': 'Gènes', 'start': 0, 'end': 0},
  { 'label': 'Geography', 'text': 'Genua', 'start': 1, 'end': 5},
  { 'label': 'City', 'text': 'Genoa', 'start': 6, 'end': 11},
  { 'label': 'Country', 'text': 'Italy', 'start': 12, 'end': 16},
  { 'label': 'Building', 'text': 'palais', 'start': 17, 'end': 20},
  { 'label': 'Organization', 'text': 'république', 'start': 21, 'end': 29},
  { 'label': 'Location', 'text': 'Méditerranée', 'start': 30, 'end': 41},
  { 'label': 'City', 'text': 'Milan', 'start': 42, 'end': 47},
  { 'label': 'Country', 'text': 'Italy', 'start': 48, 'end': 52},
  { 'label': 'Location', 'text': 'Turin', 'start': 53, 'end': 58},
  { 'label': 'City', 'text': 'Parme', 'start': 59, 'end': 64},
  { 'label': 'Country', 'text': 'Italy', 'start': 65, 'end': 69},
  { 'label': 'Location', 'text': 'Florence', 'start': 70, 'end': 78},
  { 'label': 'City', 'text': 'Rome', 'start': 79, 'end': 83},
  { 'label': 'Country', 'text': 'Italy', 'start': 84, 'end': 88},
  { 'label': 'Location', 'text': 'Salvego', 'start': 89, 'end': 96},
  { 'label': 'Person', 'text': 'Cassini', 'start': 97, 'end': 102},
  { 'label': 'Person', 'text': 'Grimaldi', 'start': 103, 'end': 110},
  { 'label': 'Location', 'text': 'Milan', 'start': 111, 'end': 116},
  { 'label': 'City', 'text': 'Turin', 'start': 117, 'end': 122},
  { 'label': 'City', 'text': 'Parme', 'start': 123, 'end': 128},
  { 'label': 'Location', 'text': 'Florence', 'start': 129, 'end': 136},
  { 'label': 'City', 'text': 'Rome', 'start': 137, 'end': 142}
]
```

- Correct JSON output format
- Does not consider the tagset given in the prompt
- Hallucinations: find non-existing entities

GÈNES

Gènes, (*Géog.*) *Genua* ; & dans les siècles ignorans du moyen âge, *Janua*, comme si Janus en étoit le fondateur ; ancienne, forte, riche ville, & l'une des principales d'Italie, capitale de la république de Gènes, avec un archevêché & un bon port. Les églises, les édifices publics & les palais y sont magnifiques : les palais se suivent sans être joints avec des maisons ordinaires ; ce qui fait le plus bel effet qu'on puisse desirer. Cette ville commerçante est presque au milieu de l'état de Gènes, en partie dans la plaine, & en partie sur une colline près de la Méditerranée, dans une heureuse & riante situation, à 28 lieues sud-ouest de Milan, 25 sud-est de Turin, 26 sud-ouest de Parme, 45 nord-ouest de Florence, 90 nord-ouest de Rome. *Long.* suivant Salvego, Cassini & le pere Grimaldi, 26^{d.} 7'. 15". *latit.* 44^{d.} 25'. 0". (D. J.)

Meta Llama 3.1 - 7B

```
[
  {'label': 'Geography', 'text': 'Gènes', 'start': 0, 'end': 0},
  {'label': 'Geography', 'text': 'Genua', 'start': 1, 'end': 5},
  {'label': 'City', 'text': 'Genoa', 'start': 6, 'end': 11},
  {'label': 'Country', 'text': 'Italy', 'start': 12, 'end': 16},
  {'label': 'Building', 'text': 'palais', 'start': 17, 'end': 20},
  {'label': 'Organization', 'text': 'république', 'start': 21, 'end': 29},
  {'label': 'Location', 'text': 'Méditerranée', 'start': 30, 'end': 41},
  {'label': 'City', 'text': 'Milan', 'start': 42, 'end': 47},
  {'label': 'Country', 'text': 'Italy', 'start': 48, 'end': 52},
  {'label': 'Location', 'text': 'Turin', 'start': 53, 'end': 58},
  {'label': 'City', 'text': 'Parme', 'start': 59, 'end': 64},
  {'label': 'Country', 'text': 'Italy', 'start': 65, 'end': 69},
  {'label': 'Location', 'text': 'Florence', 'start': 70, 'end': 78},
  {'label': 'City', 'text': 'Rome', 'start': 79, 'end': 83},
  {'label': 'Country', 'text': 'Italy', 'start': 84, 'end': 88},
  {'label': 'Location', 'text': 'Salvego', 'start': 89, 'end': 96},
  {'label': 'Person', 'text': 'Cassini', 'start': 97, 'end': 102},
  {'label': 'Person', 'text': 'Grimaldi', 'start': 103, 'end': 110},
  {'label': 'Location', 'text': 'Milan', 'start': 111, 'end': 116},
  {'label': 'City', 'text': 'Turin', 'start': 117, 'end': 122},
  {'label': 'City', 'text': 'Parme', 'start': 123, 'end': 128},
  {'label': 'Location', 'text': 'Florence', 'start': 129, 'end': 136},
  {'label': 'City', 'text': 'Rome', 'start': 137, 'end': 142}
]
```

- Correct JSON output format
- Does not consider the tagset given in the prompt
- Hallucinations: find non-existing entities
- Translate from French to English
- Find several occurrences of the same entity

GÈNES

Gènes, (*Géog.*) *Genua* ; & dans les siècles ignorans du moyen âge, *Janua*, comme si Janus en étoit le fondateur ; ancienne, forte, riche ville, & l'une des principales d'Italie, capitale de la république de Gènes, avec un archevêché & un bon port. Les églises, les édifices publics & les palais y sont magnifiques : les palais se suivent sans être joints avec des maisons ordinaires ; ce qui fait le plus bel effet qu'on puisse desirer. Cette ville commerçante est presque au milieu de l'état de Gènes, en partie dans la plaine, & en partie sur une colline près de la Méditerranée, dans une heureuse & riante situation, à 28 lieues sud-ouest de Milan, 25 sud-est de Turin, 26 sud-ouest de Parme, 45 nord-ouest de Florence, 90 nord-ouest de Rome. *Long.* suivant Salvego, Cassini & le pere Grimaldi, 26^{d.} 7'. 15". *latit.* 44^{d.} 25' 0". (D. J.)

Meta Llama 3.1 - 7B

```
[
  {'label': 'Geography', 'text': 'Gènes', 'start': 0, 'end': 0},
  {'label': 'Geography', 'text': 'Genua', 'start': 1, 'end': 5},
  {'label': 'City', 'text': 'Genoa', 'start': 6, 'end': 11},
  {'label': 'Country', 'text': 'Italy', 'start': 12, 'end': 16},
  {'label': 'Building', 'text': 'palais', 'start': 17, 'end': 20},
  {'label': 'Organization', 'text': 'république', 'start': 21, 'end': 29},
  {'label': 'Location', 'text': 'Méditerranée', 'start': 30, 'end': 41},
  {'label': 'City', 'text': 'Milan', 'start': 42, 'end': 47},
  {'label': 'Country', 'text': 'Italy', 'start': 48, 'end': 52},
  {'label': 'Location', 'text': 'Turin', 'start': 53, 'end': 58},
  {'label': 'City', 'text': 'Parme', 'start': 59, 'end': 64},
  {'label': 'Country', 'text': 'Italy', 'start': 65, 'end': 69},
  {'label': 'Location', 'text': 'Florence', 'start': 70, 'end': 78},
  {'label': 'City', 'text': 'Rome', 'start': 79, 'end': 83},
  {'label': 'Country', 'text': 'Italy', 'start': 84, 'end': 88},
  {'label': 'Location', 'text': 'Salvego', 'start': 89, 'end': 96},
  {'label': 'Person', 'text': 'Cassini', 'start': 97, 'end': 102},
  {'label': 'Person', 'text': 'Grimaldi', 'start': 103, 'end': 110},
  {'label': 'Location', 'text': 'Milan', 'start': 111, 'end': 116},
  {'label': 'City', 'text': 'Turin', 'start': 117, 'end': 122},
  {'label': 'City', 'text': 'Parme', 'start': 123, 'end': 128}
```

- Correct JSON output format
- Does not consider the tagset given in the prompt
- Hallucinations: find non-existing entities
- Translate from French to English
- Find several occurrences of the same entity
- Random numbers for span positions

```
('Gènes', 0) [...] ('Genua', 6) [...] ('Janua', 17) [...] ('Italie', 40) [...] ('république', 45)
('de', 46) ('Gènes', 47) [...] ('les', 64) ('palais', 65) [...] ('la', 121) ('Méditerranée', 122)
(',', 123) ('dans', 124) ('une', 125) ('heureuse', 126) ('&', 127) ('riante', 128) ('situation', 129)
(',', 130) ('à', 131) ('28', 132) ('lieues', 133) ('sud-ouest', 134) ('de', 135) ('Milan', 136)
(',', 137) ('25', 138) ('sud-est', 139) ('de', 140) ('Turin', 141) ('', 142) ('26', 143)
('sud-ouest', 144) ('de', 145) ('Parme', 146) ('', 147) ('45', 148) ('nord-ouest', 149) ('de', 150)
('Florence', 151) ('', 152) ('90', 153) ('nord-ouest', 154) ('de', 155) ('Rome', 156) ('.', 157)
('Long', 158) ('.', 159) ('suivant', 160) ('Salvego', 161) ('', 162) ('Cassini', 163) [...]
```

Meta Llama 3.1 - 7B

```
[
  {'label': 'Geography', 'text': 'Gènes', 'start': 0, 'end': 0},
  {'label': 'Geography', 'text': 'Genua', 'start': 1, 'end': 5},
  {'label': 'City', 'text': 'Genoa', 'start': 6, 'end': 11},
  {'label': 'Country', 'text': 'Italy', 'start': 12, 'end': 16},
  {'label': 'Building', 'text': 'palais', 'start': 17, 'end': 20},
  {'label': 'Organization', 'text': 'république', 'start': 21, 'end': 29},
  {'label': 'Location', 'text': 'Méditerranée', 'start': 30, 'end': 41},
  {'label': 'City', 'text': 'Milan', 'start': 42, 'end': 47},
  {'label': 'Country', 'text': 'Italy', 'start': 48, 'end': 52},
  {'label': 'Location', 'text': 'Turin', 'start': 53, 'end': 58},
  {'label': 'City', 'text': 'Parme', 'start': 59, 'end': 64},
  {'label': 'Country', 'text': 'Italy', 'start': 65, 'end': 69},
  {'label': 'Location', 'text': 'Florence', 'start': 70, 'end': 78},
  {'label': 'City', 'text': 'Rome', 'start': 79, 'end': 83},
  {'label': 'Country', 'text': 'Italy', 'start': 84, 'end': 88},
  {'label': 'Location', 'text': 'Salvego', 'start': 89, 'end': 96},
  {'label': 'Person', 'text': 'Cassini', 'start': 97, 'end': 102},
  {'label': 'Person', 'text': 'Grimaldi', 'start': 103, 'end': 110},
  {'label': 'Location', 'text': 'Milan', 'start': 111, 'end': 116},
  {'label': 'City', 'text': 'Turin', 'start': 117, 'end': 122},
  {'label': 'City', 'text': 'Parme', 'start': 123, 'end': 128}
```

- Correct JSON output format
- Does not consider the tagset given in the prompt
- Hallucinations: find non-existing entities
- Translate from French to English
- Find several occurrences of the same entity
- Random numbers for span positions

('Gènes', 0) [...] ('Genua', 6) [...] ('Janua', 17) [...] ('Italie', 40) [...] ('république', 45) ('de', 46) ('Gènes', 47) [...] ('les', 64) ('palais', 65) [...] ('la', 121) ('Méditerranée', 122) ('', 123) ('dans', 124) ('une', 125) ('heureuse', 126) ('&', 127) ('riante', 128) ('situation', 129) ('', 130) ('à', 131) ('28', 132) ('lieues', 133) ('sud-ouest', 134) ('de', 135) ('Milan', 136) ('', 137) ('25', 138) ('sud-est', 139) ('de', 140) ('Turin', 141) ('', 142) ('26', 143) ('sud-ouest', 144) ('de', 145) ('Parme', 146) ('', 147) ('45', 148) ('nord-ouest', 149) ('de', 150) ('Florence', 151) ('', 152) ('90', 153) ('nord-ouest', 154) ('de', 155) ('Rome', 156) ('.', 157) ('Long', 158) ('.', 159) ('suivant', 160) ('Salvego', 161) ('', 162) ('Cassini', 163) [...]

Meta Llama 3.1 - 7B

```
[
  {'label': 'Geography', 'text': 'Gènes', 'start': 0, 'end': 0},
  {'label': 'Geography', 'text': 'Genua', 'start': 1, 'end': 5},
  {'label': 'City', 'text': 'Genoa', 'start': 6, 'end': 11},
  {'label': 'Country', 'text': 'Italy', 'start': 12, 'end': 16},
  {'label': 'Building', 'text': 'palais', 'start': 17, 'end': 20},
  {'label': 'Organization', 'text': 'république', 'start': 21, 'end': 29},
  {'label': 'Location', 'text': 'Méditerranée', 'start': 30, 'end': 41},
  {'label': 'City', 'text': 'Milan', 'start': 42, 'end': 47},
  {'label': 'Country', 'text': 'Italy', 'start': 48, 'end': 52},
  {'label': 'Location', 'text': 'Turin', 'start': 53, 'end': 58},
  {'label': 'City', 'text': 'Parme', 'start': 59, 'end': 64},
  {'label': 'Country', 'text': 'Italy', 'start': 65, 'end': 69},
  {'label': 'Location', 'text': 'Florence', 'start': 70, 'end': 78},
  {'label': 'City', 'text': 'Rome', 'start': 79, 'end': 83},
  {'label': 'Country', 'text': 'Italy', 'start': 84, 'end': 88},
  {'label': 'Location', 'text': 'Salvego', 'start': 89, 'end': 96},
  {'label': 'Person', 'text': 'Cassini', 'start': 97, 'end': 102},
  {'label': 'Person', 'text': 'Grimaldi', 'start': 103, 'end': 110},
  {'label': 'Location', 'text': 'Milan', 'start': 111, 'end': 116},
  {'label': 'City', 'text': 'Turin', 'start': 117, 'end': 122},
  {'label': 'City', 'text': 'Parma', 'start': 123, 'end': 128}
]
```

- Correct JSON output format
- Does not consider the tagset given in the prompt
- Hallucinations: find non-existing entities
- Translate from French to English
- Find several occurrences of the same entity
- Random numbers for span positions

```
('Gènes', 0) [...] ('Genua', 6) [...] ('Janua', 17) [...] ('Italie', 40) [...] ('république', 45)
('de', 46) ('Gènes', 47) [...] ('les', 64) ('palais', 65) [...] ('la', 121) ('Méditerranée', 122)
(',', 123) ('dans', 124) ('une', 125) ('heureuse', 126) ('&', 127) ('riante', 128) ('situation', 129)
(',', 130) ('à', 131) ('28', 132) ('lieues', 133) ('sud-ouest', 134) ('de', 135) ('Milan', 136)
(',', 137) ('25', 138) ('sud-est', 139) ('de', 140) ('Turin', 141) ('', 142) ('26', 143)
('sud-ouest', 144) ('de', 145) ('Parme', 146) ('', 147) ('45', 148) ('nord-ouest', 149) ('de', 150)
('Florence', 151) ('', 152) ('90', 153) ('nord-ouest', 154) ('de', 155) ('Rome', 156) ('.', 157)
('Long', 158) ('.', 159) ('suivant', 160) ('Salvego', 161) ('', 162) ('Cassini', 163) [...]
```

Qwen - 7B

```
answer = [
  {'label': 'NP-Location', 'text': 'Gènes', 'start': 0, 'end': 4},
  {'label': 'NP-Time', 'text': 'milieu de Gènes', 'start': 432, 'end': 517},
  {'label': 'NP-Time', 'text': 'long suiv Salvego Cassini le pere Grimaldi', 'start': 190, 'end': 208},
  {'label': 'NP-Location', 'text': 'Méditerranée', 'start': 436, 'end': 453},
  {'label': 'NP-Time', 'text': 'latit suiv 44d 25"', 'start': 177, 'end': 198},
  {'label': 'NP-Location', 'text': 'Italy', 'start': 406, 'end': 413},
  {'label': 'NP-Time', 'text': 'long suiv', 'start': 0, 'end': 157},
  {'label': 'NP-Pronot', 'text': '的', 'start': 48, 'end': 52},
  {'label': 'NP-Pronot', 'text': '的', 'start': 134, 'end': 139},
  {'label': 'NP-Pronot', 'text': '的', 'start': 101, 'end': 103},
  {'label': 'NP-Pronot', 'text': '的', 'start': 56, 'end': 62},
  {'label': 'NP-Location', 'text': 'Milan', 'start': 98, 'end': 104},
  {'label': 'NP-Location', 'text': 'Turin', 'start': 137, 'end': 142},
  {'label': 'NP-Location', 'text': 'Parme', 'start': 146, 'end': 152},
  {'label': 'NP-Location', 'text': 'Rome', 'start': 154, 'end': 158},
  {'label': 'NP-Pronot', 'text': '的', 'start': 307, 'end': 316},
  {'label': 'NP-Location', 'text': 'North West', 'start': 159, 'end': 221},
  {'label': 'NP-Pronot', 'text': '的', 'start': 48, 'end': 50},
  {'label': 'NP-Pronot', 'text': '的', 'start': 63, 'end': 74},
  {'label': 'NP-Location', 'text': 'South West', 'start': 132, 'end': 185},
  {'label': 'NP-Location', 'text': 'South East', 'start': 139, 'end': 146},
  {'label': 'NP-Location', 'text': 'North West', 'start': 140, 'end': 154},
  {'label': 'NP-Location', 'text': 'North East', 'start': 150, 'end': 217},
  {'label': 'NP-Pronot', 'text': '的", 'start': 69, 'end': 73},
  {'label': 'NP-Location', 'text': 'Milan', 'start': 128, 'end': 140},
  {'label': 'NP-Location', 'text': 'Turin', 'start': 141, 'end': 150},
  {'label': 'NP-Location', 'text': 'Parme', 'start': 151, 'end': 163},
  {'label': 'NP-Location', 'text': 'Florence', 'start': 163, 'end': 172},
  {'label': 'NP-Pronot', 'text': '的', 'start': 90, 'end': 94}
```

- Correct JSON output format
- Does not consider the tagset given in the prompt
- Hallucinations: find non-existing entities
- Translate from French to English
- ~~Find several occurrences of the same entity~~
- Random numbers for span positions

Microsoft Phi 3 mini 4k

RAM Usage 3.38 GB
CPU: 0.00 %

lmstudio-community • Phi 3 1 mini 4k instruct 3B Q4_K_M gguf

+ New Chat (%N) Export

Phi-3 3B Q4_K_M GGUF

plaintext markdown

AI

Les informations données semblent décrits dans un format pré-agencé ou des annotations de texte, où les nombres sont associés à certains mots. Voici la traduction en français avec une interprétation possible basée sur le contexte historique et géographique :

Gènes (0) est une ville romaine située dans l'Italie actuelle (1). Elle était célèbre pour ses monuments architecturaux anciens, notamment les Gènes du Moyen Âge (2). Situées au centre de la région méditerranéenne d'Épire en Grèce antique, elles étaient un centre économique prospère avec une activité commerciale intense. La ville était également connue pour sa position stratégique entre le nord-ouest et le sud du pays (3).

Pendant les époques médiévale et moderne de Gènes, la cathédrale Sainte Marie a été l'un des monuments religieux emblématiques de la ville. Un autre monument notable était "la Maison d'Auguste" située sur le site antique du Palais Impérial (4).

En ce qui concerne les époques anciennes, Gènes avait une position stratégique significative en raison de sa situation géographique dans la plaine méditerranéenne. L'empereur Auguste a probablement établi son palais sur un site antique près du centre-ville (5).

Gènes était également célèbre pour ses édifices publics magnifiques et plusieurs palais importants existaient à cette époque, notamment ceux situés dans

Google Gemma 2 - 9B

RAM Usage 6.09 GB
CPU: 0.00 %

lmstudio-community · gemma 2 it 9B Q4_K_M gguf

+ New Chat (%N) Export gemma2 9B Q4_K_M GGUF plaintext markdown

It looks like we have a bunch of coordinates, measurements, names, and some strange code snippets!

Let's break down what might be going on:

Possible Scenarios:

- **Mapping Project:** The coordinates could represent locations, landmarks, or boundaries on a map. The measurements might indicate distances between points, elevations, or sizes of geographical features.
- **Historical Research:** The names and dates might refer to historical figures, events, or places. The code snippets could be fragments of old maps, documents, or architectural plans.
- **Data Analysis:** The numbers and letters might represent data points being analyzed for patterns or trends.

Let's Decode Further:

To understand the context better, we need more information:

1. **What is the Source?** Where did this information come from? A book, a website, an archive?
2. **What is the Overall Topic?** Is it related to geography, history, art, science?
3. **Can you Provide More Context?** Any additional details about the surrounding text or the purpose of gathering this information would be helpful.

Let me know if you have any more clues, and I'll do my best to help decipher this intriguing puzzle!

AI

Regenerate Continue

Mistral 0.2 - 7B

RAM Usage 4.49 GB
CPU: 0.00 %

TheBloke • mistral instruct v0 2 7B Q4_K_S gguf ▾

+ New Chat (36N) Export Llama 7B Q4_K_S GGUF plaintext markdown

,105) ('; ,106) ('en' ,107) ('partie' ,108) ('dans' ,109) ('la' ,110) ('plaine' ,111) ('; ,112) ('&' ,113) ('en' ,114) ('partie' ,115) ('sur' ,116) ('une' ,117) ('colline' ,118) ('près' ,119) ('de' ,120) ('la' ,121) ('Méditerranée' ,122) ('; ,123) ('dans' ,124) ('une' ,125) ('heureuse' ,126) ('&' ,127) ('riante' ,128) ('situation' ,129) ('; ,130) ('à' ,131) ('28' ,132) ('lieues' ,133) ('sud-ouest' ,134) ('de' ,135) ('Milan' ,136) ('; ,137) ('25' ,138) ('sud-est' ,139) ('de' ,140) ('Turin' ,141) ('; ,142) ('26' ,143) ('sud-ouest' ,144) ('de' ,145) ('Parme' ,146) ('; ,147) ('45' ,148) ('nord-ouest' ,149) ('de' ,150) ('Florence' ,151) ('; ,152) ('90' ,153) ('nord-ouest' ,154) ('de' ,155) ('Rome' ,156) ('; ,157) ('Long' ,158) ('; ,159) ('suivant' ,160) ('Salvego' ,161) ('; ,162) ('Cassini' ,163) ('&' ,164) ('le' ,165) ('pere' ,166) ('Grimaldi' ,167) ('; ,168) ('26d' ,169) ('; ,170) ('7' ,171) ('" ,172) ('; ,173) ('15' ,174) ('" ,175) ('; ,176) ('latit' ,177) ('; ,178) ('44d' ,179) ('; ,180) ('25' ,181) ('" ,182) ('; ,183) ('0' ,184) ('" ,185) ('; ,186) ('; ,187) ('D.' ,188) ('J.' ,189) ('; ,190)

This text appears to be describing the location and history of Genoa, a city in Italy. Here is a summary in English:

Genoa, located on the Italian Riviera, was an ancient wealthy city and the capital of the Republic of Genoa. Its exact origins are unknown but it is believed to have been founded by Janus, the two-faced Roman god. The city has been influential in Italy and Europe throughout history. It was a major commercial center, with many beautiful churches, public buildings, and palaces. Its location on the Mediterranean Sea made it an important trading port, and it was strategically situated between Milan, Turin, Parma, Florence, and Rome. The text mentions several famous historical figures associated with Genoa, including Salvego Cassini and the Grimaldi family.

The text also provides the latitude and longitude of Genoa as follows: 44°25' N, 8°31' E. This information could be useful for someone trying to locate Genoa on a map or using GPS coordinates. The text ends with a quote from Dante Alighieri's Divine Comedy, referencing Janus and the idea that Genoa has been an influential city throughout history.

AI

Regenerate Continue

Summary of the talk

1. The GeoEDdA dataset: A Gold Standard Dataset for Geo-semantic Annotation
2. BERT fine-tuning for NER
3. Few-shot prompting for NER
4. Perspectives: Knowledge Graph construction using LLMs
 - collaborative work with Y. Lairgi, R. Cazabet & K. Benabdeslem

Knowledge Graph construction

Grenoble est sur l'Isere, à onze lieues S O. de Chambéri, quarante-deux N. O. de Turin, seize S. E. de Vienne, cent vingt-quatre S. O. de Paris. Long. suivant Harris, 23^d. 31'. 15". suivant Cassini, 23^d. 14'. 15". latit 45^d. 11'.

DAUPHINÉ, (Géog. mod.) province de France bornée à l'occident par le Rhône, au septentrion par le Rhône & la Savoie, au midi par la Provence, & à l'orient par les Alpes. Elle est arrosée par le Rhône, la Durance, l'Isere, & le Drame. [...]. Elle se divise en haut & bas. Le haut comprend le Gresivaudan, le Briançonnois, l'Embrunois, le Gapançois, le Royannez, & les Baronies. Le bas contient le Valentinois, le Diois, & le Tricassinois. [...]. Grenoble en est la capitale. Long. 20-29. lat. 43-46.

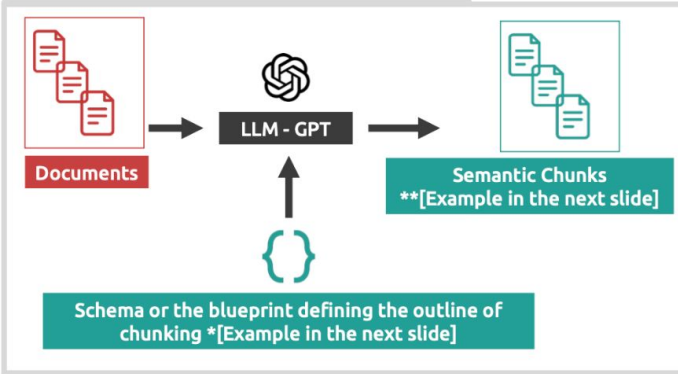
Vienne, (Géog. mod.) ville de France, dans le Dauphiné, sur le bord oriental du Rhône, à 5 lieues au midi & au dessous de Lyon, à 15 au nord-ouest de Grenoble, & à 108 au sud-est de Paris. [...] Long. 22. 30. latit. 45. 33.

Lyon est à six lieues N. O. de Vienne, vingt N. O. de Grenoble, vingt-huit S. O. de Genève, trente-six N. d'Avignon, quarante S. O. de Dijon, soixante N. O. de Turin, cent S. E. de Paris. Long. suivant Cassini, 22^d. 16'. 30". lat. 45^d. 45'. 20".

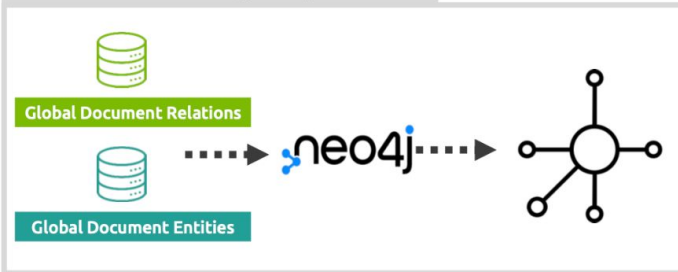
Lyon is located six leagues N. W. of Vienne, twenty N. W. of Grenoble, twenty-eight S. W. of Geneva, thirty-six N. of Avignon, forty S. W. of Dijon, sixty N. W. of Turin, one hundred S. E. of Paris. Long. following Cassini, 22d. 16'. 30". lat. 45d. 45'. 20".

Knowledge Graph construction using LLMs

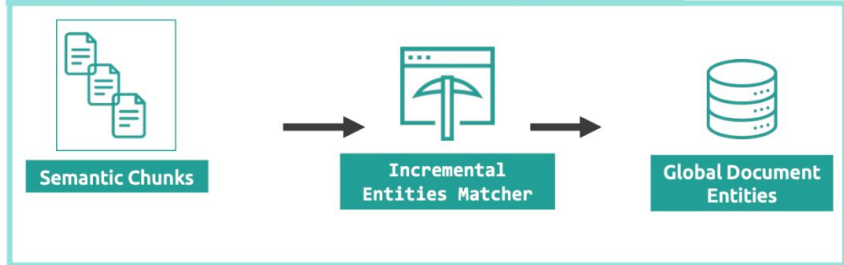
Module 1 – Document Distiller*



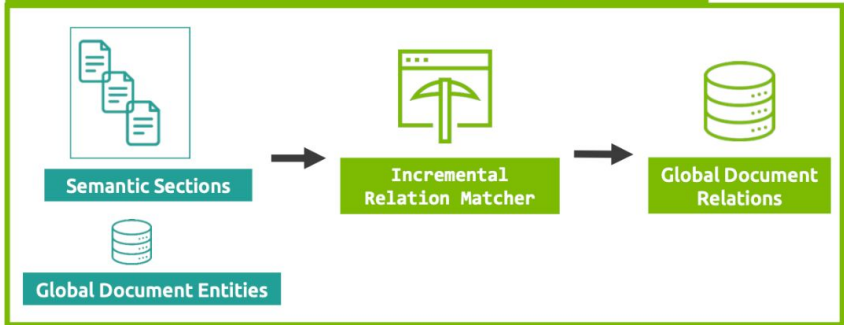
Module 4 – Neo4j Graph



Module 2 – Incremental Entities Extractor



Module 3 – Incremental Relations Extractor



Thank you !

Ludovic Moncla

ludovic.moncla@insa-lyon.fr

-

<https://ludovicmoncla.github.io>

