



ZA Cloud SRL, str. Govora 16A, 400664
Cluj-Napoca, România, Tel: +40 723 175 635,
Email: office@zettacloud.ro

Zetta Cloud

IntelliDockers - Multilingual Named Entity Recognition (NER) Specifications

[May 2023]



ZA Cloud SRL, str. Govora 16A, 400664
Cluj-Napoca, România, Tel: +40 723 175 635,
Email: office@zettacloud.ro

TABLE OF CONTENTS

About Us	3
Zetta Cloud Products	3
IntelliDockers	4
Pricing Model	4
Entities Extractor	4
Supported Entity Types	5
Usage Example	6
XLU - Cross Language Understanding (40+ languages)	8
HW Requirements	9



ZA Cloud SRL, str. Govora 16A, 400664
Cluj-Napoca, România, Tel: +40 723 175 635,
Email: office@zettacloud.ro

About Us

[Zetta Cloud](#) Romania was established in 2013 in Cluj-Napoca, Romania's "Silicon Valley" and an ideal habitat for fostering technology talent and innovation. Our company's goal is to provide software solutions for deep understanding of online content using Artificial Intelligence.

The founders of Zetta Cloud have a strong background in the software industry, artificial intelligence and enterprise solutions. Our team comprises of **Big Data specialists** (Senior engineers skilled in Big Data technologies), Senior **AI researchers** (Top academic experts in Hybrid AI, Machine Learning, Game Theory, Graph Theory, Evolutionary Algorithms) and **Social & Political Sciences researchers** (OSINT, Social Media, Statistics, Political & Social Sciences).

Our solutions portfolio ranges from **Digital News Technologies** (AI-based software solutions for the Digital Media sector), through **Artificial Intelligence Research** (Development of AI algorithms using Deep Learning & Machine Learning approaches), to **Intelligent Data Analytics Services** (Data Analytics services for Open Data & Social Media).

Zetta Cloud Products

IntelliDockers (www.intelldockers.com) are self-contained text analytics engines for secure data processing. We provide state of the art Natural Language Processing technology based on Deep Learning and Neural Networks AI as self-contained engines. IntelliDockers are deployed on your isolated infrastructure making them suitable for secure data processing.

Factory (www.factory-nocode.ai) is a no-code AI training platform for secure data processing that offers complete safety for clients data and that can be used without any coding know-how or previous experience. With Factory you can train, evaluate and deploy your own AI models for document understanding without writing any line of code.

TrustServista (www.trustservista.com) is a unique software platform that can automatically determine the origin, content quality and trustworthiness of online news. It uses Artificial Intelligence to automatically score online news articles or news content from a quality and trustworthiness perspective.



ZA Cloud SRL, str. Govora 16A, 400664
Cluj-Napoca, România, Tel: +40 723 175 635,
Email: office@zettacloud.ro

IntelliDockers



IntelliDockers are Self-Contained Text Analytics Engines for Secure Data Processing.

- Work offline and on your hardware.
- Out-of-the-box pre-trained engines to understand a wide variety of content.
- Custom engines can be created for any specific needs.
- Multi-language processing capabilities.
- No data volume limitations are imposed per instance.

IntelliDockers are deployed on your isolated infrastructure making them suitable for secure data processing. You can deploy your IntelliDockers on your local servers or on any major cloud infrastructures out there, like: Google Cloud Platform, Microsoft Azure, Digital Ocean or Amazon ECS (Elastic Container Service).

The engines can be easily integrated with your existing software platforms through the exposed REST API interface, augmenting them with state of the art NLP technology.

For the full list of IntelliDockers AI Text Analytics engines please visit:

<https://hub.docker.com/u/intelldockers>

Pricing Model

- **Pricing is per instance.**
- 1-year term with included support or maintenance.
- No data volume limitations are imposed per instance.
- Support for any new language can be added in 3-5 months.
- Custom engines can be created for any specific needs.

Entities Extractor

Named entities, categorized by type (people, organizations, locations, products, etc.), constitute the core factual information of any content. These engines extract named entities out of given



text. All flavors described below are supporting the same list of entity types, see the following section.

Supported Entity Types

We are currently supporting the following complete list of Entity Types.

LOCATION	A city, state, country, region, building, monument, body of water, park, or address.
ORGANIZATION	A corporation, institution, agency, or other group defined by an organizational structure.
PERSON	A human identified by name, nickname, or alias.
TITLE	Appellation associated with an occupation, office, or status.
NATIONALITY	Reference to a country or region of origin.
RELIGION	Reference to an organized religion or theology, as well as its followers.
IDENTIFIER_CREDIT_CARD_NUM	Credit card numbers.
IDENTIFIER_EMAIL	Email addresses.
IDENTIFIER_MONEY	Currencies.
IDENTIFIER_PERSONAL_ID_NUM	Personal identification numbers.
IDENTIFIER_PHONE_NUMBER	Phone numbers.
IDENTIFIER_URL	Web addresses.
TEMPORAL_DATE	Date.
TEMPORAL_TIME	Time.
IDENTIFIER_DISTANCE	Distance.
IDENTIFIER_LATITUDE_LONGITUDE	Geographic locations in latitude and longitude coordinates.
HASHTAG	Hashtags found inside an article. Any keyword that is preceded by #. Example: The article http://money.cnn.com/2018/01/24/media/releasethememo-twitter-accounts/index.html contains the entity of type HASHTAG #ReleaseTheMemo



Usage Example

This is an example of calling the Entities Extractor on text in Romanian using simple *curl*:

```
curl -X 'POST' \  
  'http://192.168.56.25:8989/rest/process' \  
  -H 'accept: application/json' \  
  -H 'Content-Type: application/json' \  
  -d '  
  {  
    "content": "Candidatul independent la Primăria Capitalei susținut de USR, Nicușor Dan,  
    afirmă că bugetul Primăriei Municipiului București pe 2020 \nu are nicio legătură cu realitatea",  
    în condițiile în care veniturile estimate – 7,2 miliarde de lei – sunt, ca și în anii precedenți,  
    supraevaluate cu aproximativ 75%, transmite Agerpres.",  
    "language": "ron"  
  }  
}
```

Calling the Entities Extractor as above will generate the simple JSON response below:

```
{  
  "entities": [  
    {  
      "entity": "Primăria Capitalei",  
      "type": "ORGANIZATION",  
      "count": 1,  
      "details": [  
        {  
          "score": 0.9728477001190186,  
          "start": 26,  
          "end": 44  
        }  
      ]  
    },  
    {  
      "entity": "USR",  
      "type": "ORGANIZATION",  
      "count": 1,  
      "details": [  
        {  
          "score": 0.9728477001190186,  
          "start": 26,  
          "end": 44  
        }  
      ]  
    }  
  ]  
}
```



ZA Cloud SRL, str. Govora 16A, 400664
Cluj-Napoca, România, Tel: +40 723 175 635,
Email: office@zettacloud.ro

```
{
  "score": 0.9944616556167603,
  "start": 57,
  "end": 60
}
],
},
{
  "entity": "Nicușor Dan",
  "type": "PERSON",
  "count": 1,
  "details": [
    {
      "score": 0.9825072884559631,
      "start": 62,
      "end": 73
    }
  ]
},
},
{
  "entity": "Primăriei Municipiului București",
  "type": "ORGANIZATION",
  "count": 1,
  "details": [
    {
      "score": 0.9880717992782593,
      "start": 93,
      "end": 125
    }
  ]
},
},
{
  "entity": "2020",
  "type": "TEMPORAL_DATE",
  "count": 1,
  "details": [
    {
      "score": 0.8982105255126953,
```



ZA Cloud SRL, str. Govora 16A, 400664
Cluj-Napoca, România, Tel: +40 723 175 635,
Email: office@zettacloud.ro

```
"start": 129,  
  "end": 133  
}  
]  
},  
{  
  "entity": "7,2 miliarde de lei",  
  "type": "IDENTIFIER_MONEY",  
  "count": 1,  
  "details": [  
    {  
      "score": 0.9840089082717896,  
      "start": 217,  
      "end": 236  
    }  
  ]  
},  
{  
  "entity": "Agerpres",  
  "type": "ORGANIZATION",  
  "count": 1,  
  "details": [  
    {  
      "score": 0.9890756011009216,  
      "start": 315,  
      "end": 323  
    }  
  ]  
}  
]  
}  
}
```

XLU - Cross Language Understanding (40+ languages)

The Entity Extractor XLU is a versatile engine that **generalizes for a lot of languages**. That means that you can throw any language on it, even if the language is not officially listed. So, you could even try **Farsi**, if needed.



ZA Cloud SRL, str. Govora 16A, 400664
Cluj-Napoca, România, Tel: +40 723 175 635,
Email: office@zettacloud.ro

HW Requirements

Due to the large size of the model, we recommend running this engine in an environment equipped with an NVIDIA Graphical Processing Unit (GPU).

Minimum recommended requirements for the GPU:

- Type: NVIDIA Quadro
- RAM: 4041MB

Other minimum recommended requirements:

- 4xCPU
- RAM: 4GB
- HDD: 10GB