

[Home](#)

Part 10: Step by Step Guide to Master NLP – Named Entity Recognition



CHIRAG GOYAL — June 23, 2021

[Advanced](#) [NLP](#) [Project](#) [Python](#) [Text](#) [Unsupervised](#)

This article was published as a part of the [Data Science Blogathon](#)

Introduction

This article is part of an ongoing blog series on Natural Language Processing (NLP). In the previous article, we discussed semantic analysis, which is a level of NLP tasks. In that article, we discussed the techniques of Semantic analysis in which we discussed one technique named entity extraction, which is very important to understand in NLP.

So, In this article, we will deep dive into the entity extraction technique named Named Entity Recognition, which is a very useful component in the pipeline of NLP.

This is part-10 of the blog series on the Step by Step Guide to Natural Language Processing.

Table of Contents

1. What is Named Entity Recognition (NER)?
2. Different blocks present in a Typical NER model
3. Deep understanding of Named Entity Recognition with an example
4. How does Named Entity Recognition work?
5. Use-cases of Named Entity Recognition
6. How can I use NER?

What is Named Entity Recognition (NER)?

Let's first discuss what entities mean?

Entities are the most important chunks of a particular sentence such as noun phrases, verb phrases, or both. Generally, Entity Detection algorithms are ensemble models of :

- Rule-based Parsing, python
- Dictionary lookups,
- POS Tagging,
- Dependency Parsing.

For Example,

In the above sentence, the entities are:

Date: Thursday, Time: night, Location: Chateau Marmont, Person: Cate Blanchett

Now, we can start our discussion on Named Entity Recognition (NER),

1. Named Entity Recognition is one of the key entity detection methods in NLP.
2. Named entity recognition is a natural language processing technique that can automatically scan entire articles and pull out some fundamental entities in a text and classify them into predefined categories. Entities may be,
 - Organizations,
 - Quantities,
 - Monetary values,
 - Percentages, and more.
 - People’s names
 - Company names
 - Geographic locations (Both physical and political)
 - Product names
 - Dates and times
 - Amounts of money
 - Names of events
3. In simple words, Named Entity Recognition is the process of detecting the named entities such as person names, location names, company names, etc from the text.
4. It is also known as entity identification or entity extraction or entity chunking.

For Example,

Ousted **WeWork** founder **Adam Neumann** lists his **Manhattan** penthouse for **\$37.5 million**
[organization] [person] [location] [monetary value]

5. With the help of named entity recognition, we can extract key information to understand the text, or merely use it to extract important information to store in a database.
6. The applicability of entity detection can be seen in many applications such as
 - Automated Chatbots,
 - Content Analyzers,
 - Consumer Insights, etc.

Commonly used types of named entity:

Named Entity Type	Example
ORGANIZATION	WHO
PERSON	President Obama
LOCATION	Mount Everest
DATE	2020-07-10
TIME	12:50 P.M.
MONEY	One Million Dollars
PERCENT	98.24%
FACILITY	Washington Monument

Image Source: Google Images

Different blocks present in a Typical Named Entity Recognition model

A typical NER model consists of the following three blocks:

Noun Phrase Identification

This step deals with extracting all the noun phrases from a text with the help of dependency parsing and part of speech tagging.

Phrase Classification

In this classification step, we classified all the extracted noun phrases from the above step into their respective categories. To disambiguate locations, `Google Maps API` can provide a very good path. and to identify person names or company names, the open databases from `DBpedia`, `Wikipedia` can be used. Apart from this, we can also make the lookup tables and dictionaries by combining information with the help of different sources.

Entity Disambiguation

Sometimes what happens is that entities are misclassified, hence creating a validation layer on top of the results becomes useful. The use of knowledge graphs can be exploited for this purpose. Some of the popular knowledge graphs are:

- `Google Knowledge Graph`,
- `IBM Watson`,
- `Wikipedia`, etc.

Deep understanding of NER with an Example

Consider the following sentence:

London is the **capital** and most populous **city** of **England** and the **United Kingdom**.

The blue cells represent the nouns. Some of these nouns describe real things present in the world.

For Example, From the above, the following nouns represent physical places on a map.

“London”, “England”, “United Kingdom”

It would be a great thing if we can detect that! With that amount of information, we could automatically extract a list of real-world places mentioned in a document with the help of NLP.

Therefore, the goal of NER is to detect and label these nouns with the real-world concepts that they represent.

So, when we run each token present in the sentence through a NER tagging model, our sentence looks like as,

London is the capital and most populous city of **England** and the **United Kingdom**.
Geographic Entity Geographic Entity Geographic Entity

Let's discuss what exactly the NER system does?

NER systems aren't just doing a simple dictionary lookup. Instead, they are using the context of how a word appears in the sentence and used a statistical model to guess which type of noun that particular word represents.

Since NER makes it easy to grab structured data out of the text, therefore it has tons of uses. It's one of the easiest methods to quickly get insightful value out of an NLP pipeline.

If you want to try out NER yourself, then refer to the [link](#) .

How does Named Entity Recognition work?

As we can simply observe that after reading a particular text, naturally we can recognize named entities such as people, values, locations, and so on.

For Example, Consider the following sentence:

Sentence: Sundar Pichai, the CEO of Google Inc. is walking in the streets of California.

From the above sentence, we can identify three types of entities: (Named Entities)

- ("person": "Sundar Pichai"),
- ("org": "Google Inc."),
- ("location": "California").

But to do the same thing with the help of computers, we need to help them recognize entities first so that they can categorize them. So, to do so we can take the help of machine learning and Natural Language Processing (NLP).

Let's discuss the role of both these things while implementing NER using computers:

- **NLP:** It studies the structure and rules of language and forms intelligent systems that are capable of deriving meaning from text and speech.
- **Machine Learning:** It helps machines learn and improve over time.

To learn what an entity is, a NER model needs to be able to detect a word or string of words that form an entity (e.g. California) and decide which entity category it belongs to.

So, as a concluding step we can say that the heart of any NER model is a two-step process:

- Detect a named entity
- Categorize the entity

So first, we need to create entity categories, like Name, Location, Event, Organization, etc., and feed a NER model relevant training data.

Then, by tagging some samples of words and phrases with their corresponding entities, we'll eventually teach our NER model to detect the entities and categorize them.

Use-Cases of Named Entity Recognition

As we have discussed in the above section that the Named entity recognition (NER) will help us to easily identify the key components in a text, such as names of people, places, brands, monetary values, and more.

And extracting the main entities from a text helps us to sort the unstructured data and detect the important information, which is crucial if you have to deal with large datasets.

So, Let's discuss some of the interesting use cases of Named Entity Recognition:

Customer Support



Image Source: Google Images

Let's discuss the use case of customer support tickets where we deal with a rising number of tickets, there we can use named entity recognition techniques to handle the customer requests faster.

From a business perspective, if we automate the repetitive customer service tasks, such as categorizing customers' issues, and queries, then it saves you valuable time. As a result, it helps to improve your resolution rates and boost customer satisfaction.

Here, we can also use entity extraction to pull the relevant pieces of information, like product names or serial numbers, which makes it easier to route tickets to the most suitable agent or team for handling that issue.

Gain Insights from Customer Feedback



Image Source: Google Images

For almost all of the product-based companies, Online reviews are a great source of taking the customer feedback, as they can provide rich insights about what customers like and dislike about your products and the aspects of your business that need improvements for business increment.

For Example, we can use the NER system to detect locations that are mentioned most often in negative customer feedback, which might lead you to focus on a particular office branch.

Recommendation System

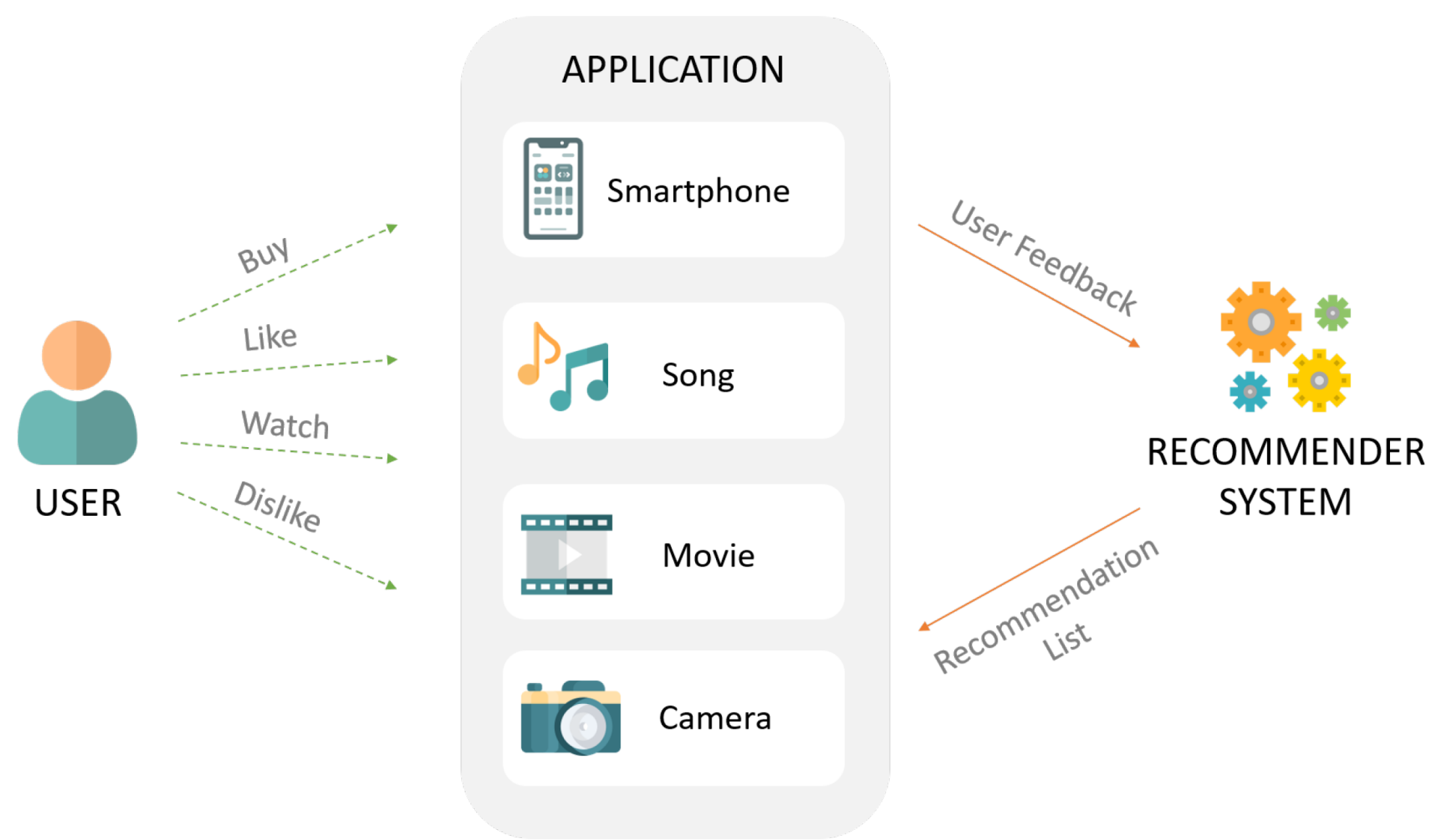


Image Source: Google Images

Many modern applications such as Netflix, YouTube, Facebook, etc. rely on recommendation systems to produces optimal customer experiences. A lot of these systems rely on named entity recognition, which can give suggestions based on the user search history.

For Example, if you watch a lot of educational videos on YouTube, then you'll get more recommendations that have been classified as entity education.

Summarizing Resumes

John Smith

134 Rightward Way, Portland, ME, 04019

Phone: (678) 757 9413 / E-mail: john@johnsmith.com

CAREER OBJECTIVE

I'm a professional Sales Associate with years of experience. I have achieved great results last year. I'm looking for new job opportunities to improve my skills.

PROFESSIONAL EXPERIENCE

Armani Menswear

2012 – Present

Sales Associate

Working at Armani, helping clients make good choices.

- Received honours and awards.
- Worked in sales department.
- Focused on delivering excellent customer service.
- Took part in many trainings and courses.
- Achieved great results in resolving customer complaints.

Nordstrom Menswear

2008 – 2012

Sales Associate

Worked at Nordstrom Menswear Sales Department.

- Worked closely with customers, took care of their complaints.
- Exceeded sales targets.
- Was responsible for up-selling products.
- Contributed to many different activities.

Chicago Public Lending Library

2008 – 2012

Librarian

Responsible for dealing with general public and primary school children. Catalogued books with the library's online systems.

EDUCATION

COLUMBIA UNIVERSITY, New York – BA Hons in History, 2006

CERTIFICATES & COURSES

Institute of Sales Management - Associate, 2008

Institute of Sales Management - Executive, 2014

History of Textile Manufacturing – University of Chicago, 2007

OTHER SKILLS & INTERESTS

1. Inventory

2. Good Communication

3. MS Office

4. Confident

5. I like to play piano and hike

LANGUAGE SKILLS

Spanish (Intermediate)

French (Basic)

Italian (Basic)

REFERENCES ARE AVAILABLE UPON REQUEST

John Smith

Sales Associate

Personal Info

Address

134 Rightward Way
Portland, ME, 04019

Phone

(678) 757 9413

E-mail

John@johnsmith.com

www

www.johnsmithswebsite.com

Linkedin

www.linkedin.com/in/john-smith/

Languages

Spanish

Intermediate

French

Basic

Italian

Basic

Skills

Tracking Frequent Customers

Inventory

Communication Skills

Anticipating Customer Needs

Confident and Friendly Manner

Proficient in MS Office

Product Knowledge

Brand Awareness

Motivated and highly professional Sales Associate with 6+ years of experience. Achieved a 42% increase in sales over the past year. A conscientious and hard working team player seeking to leverage my skills in the Senior Sales Associate position at Prada Menswear.

Experience

2012-06

Sales Associate

Armani Menswear

- Voted Sales Person of the Year in 2016.
- Exceeded sales target for the year by 25%.
- Attained 78% customer satisfaction this year, and 72% the previous year.
- Thorough knowledge of the department's stock and forthcoming ranges.
- Resolved 92% of customer complaints without delegating to a senior manager.

2008-06 - 2012-05

Sales Associate

Nordstrom Menswear

- Resolved all customer inquiries and complaints.
- Exceeded sales target by 5% in my first year.
- Consistently exceeded sales targets by at least 7% for every subsequent year.
- Managed to up-sell products by focusing on customers' past purchases.
- Contributed to the company's newsletter and Facebook account.

2002-09 - 2006-05

Librarian

Chicago Public Lending Library

- Dealt with the general public.
- Ran a reading group for primary school children.
- Helped parents select appropriate books for their children.
- Learned how to catalogue books with the library's online systems.

Education

2002-09 - 2006-05

Columbia University, New York

- BA Hons in History.
- Raised over \$20,000 for Cancer Research by organizing and promoting a university Fun Run.
- Attended night classes in Spanish for a year and am now reasonably fluent.
- Chaired the University Debating Society in 2003.
- Took part in two University theatre productions.

Certificates

2008-07

Institute of Sales Management Associate Certification

2014-06

Institute of Sales Management Executive Certification

Courses

2017-01

History of Textile Manufacturing, The University of Chicago

Interests

Music - Grade 8 piano. Occasionally play at retirement homes.

Amateur Dramatics - Member of amateur theater group for 3 years.

Creative writing group member and regular blogger.

Hiking

Image Source: Google Images

While recruiting new peoples, Recruiters spend many hours of their day going through resumes and finds for the right candidate. Each resume contains almost the same type of information, but their organized manner and formatting are different, so this becomes a classic example of unstructured data.

So, here with the help of an entity extractor, the recruitment teams can instantly extract the most relevant information about candidates, from personal information such as name, address, phone number, date of birth, and email, etc, to information related to their training and experience like certifications, degree, company names, skills, etc.

Some more use-cases of NER are:

- Optimizing search engine algorithms,
- Content classification for news channels, etc.

How can I use NER?

If you work on a business problem statement, and if you think that your business could benefit from NER, then you can use it pretty easily with the help of the following excellent open-source libraries:

- NLTK ,
- SpaCy ,
- Stanford NER .

Each has its own pros and cons, which you can explore by referring to the above-mentioned links.

This ends our Part-10 of the Blog Series on Natural Language Processing!

Other Blog Posts by Me

You can also check my previous blog posts.

We use cookies on Analytics Vidhya websites to deliver our services, analyze web traffic, and improve your experience on the site. By using Analytics Vidhya, you

agree to our [Privacy Policy](#) and [Terms of Use](#).