

Institute of Technology
Nirma University, Ahmedabad



Subject Details:

2CSDE70

Natural Language Processing

Student Roll No. and Name:

20BCE057 Devasy Patel

Practical 1:

Task 1: Explore Libraries of NLP (NLTK, spaCy, Gensim) and compare them

Task 2: Explore Regular Expressions and validate Email using RegEx

Task 3: List Features and Limitations of ELIZA and have a conversation with ELIZA

Task 1: Explore Libraries of NLP (NLTK, spaCy, Gensim) and compare them

Feature	NLTK	spaCy	Gensim
USP	Most well known, and full NLP library	Fastest and most optimized NLP library	Known for topic and vector space modeling for document similarity analysis
Used for	NLP Research and Education	Production level NLP Applications	Large Scale NLP Applications
Scalability	Can be used for small to medium-scale applications	Can be used for large-scale applications	Good for large-scale applications
Fullness	Full NLP library, provides tools for almost all NLP tasks with plenty of approaches to each task	Almost a Full NLP, it has highly optimized tools for every NLP task	Not a Full NLP library, and does not have enough tools to do all NLP tasks, it is accompanied by some other library like previous 2.
Performance	Slow as compared to SpaCy and Gensim	Fast and efficient, as it contains the most efficient and optimized tools.	Fast and Scalable

Processing	NLTK returns lists of string as output	spaCy is object oriented, it returns objects as output	Gensim used word vector model
Word Vectors	No integrated Word Vectors	Provides word vectors	Provides word vectors
Tokenization	Yes	Yes	Yes
Sentence Segmentation	Yes	Yes	Yes
Part-of-Speech Tagging	Yes	Yes	No
Named Entity Recognition	Yes	Yes	No
Dependency Parsing	Yes	Yes	No
Semantic Reasoning	Yes	No	No

Task 2: Explore Regular Expressions and validate Email using RegEx

Nirma University Student email RegEx:

The screenshot shows the regex101 website interface. The regular expression being tested is `^/[0-9]{2}+[a-z]{3}+[0-9]+@nirmauni.ac.in`. The test string contains several email addresses, and the tool highlights four matches: `20bce090@nirmauni.ac.in`, `20bce090@gmail.com`, `19bce090@nirmauni.ac.in`, and `bce105@nirmauni.ac.in`. The explanation section on the right provides a detailed breakdown of the regex components, and the match information section lists the matches with their positions and the matched text.

Explanation of the RegEx:

- The string should start with any 2 digits, that would represent the year of admission of the student who holds that roll number.
- This would be followed by the 3 characters, which would represent 1. the institute, 2. The branch, for eg: bce represent b- > ITNU, ce- > computer engineering

- This would be followed by any number of numbers which would represent the branch roll number of the student.
- The end will we @nirmauni.ac.in

Task 3: List Features and Limitations of ELIZA and have a conversation with ELIZA



ELIZA is a fairly basic chatbot (basically a NLP program) for Rogerian psychotherapy.

ELIZA was created from 1964 to 1966 at MIT AI Lab by Joseph Weizenbaum.

ELIZA was a pioneering effort in the field of conversational AI and had a significant impact on subsequent developments in the field. It inspired the development of other chatbots and natural language processing tools that continue to be used today.

My conversation with ELIZA:

Talk to Eliza by typing your questions and answers in the input box.

```
> Hello, I am Eliza. I'll be your therapist today.  
* ELIZA I think I am depressed, I feel that people around are better than me  
and I can't cope up with them.  
> Do you often feel that people around am better than you and you can't cope  
up with them?  
* Yes I feel that  
> Tell me more about such feelings.  
* I always feel lonely and alone, all the time. I am mentally ill.  
> How long have you been mentally ill?  
* Since I was admitted into University for my higher education, it is really  
stressful.  
> Would you say that you have psychological problems?
```

TYPE HERE

Talk to Eliza by typing your questions and answers in the input box.

```
* Yes I am sure I have brain problems.  
> Did you come to me because you are sure you have brain problems?  
* Yes  
> I see. Let's explore that a bit more.  
* How can I be more happy?  
> I'm not convinced that you truly want to be more happy?  
* Every person's ultimate goal is to be happy and so is mine, I don't know  
that why are you not convinced about that.  
> Don't you really know that why am I not convinced about that?  
* No  
> You are being a bit negative and that won't help us move forward.
```

TYPE HERE

Talk to Eliza by typing your questions and answers in the input box.

```
* How can I be more happy?  
> I'm not convinced that you truly want to be more happy?  
* Every person's ultimate goal is to be happy and so is mine, I don't know  
that why are you not convinced about that.  
> Don't you really know that why am I not convinced about that?  
* No  
> You are being a bit negative and that won't help us move forward.  
* Ok I will be more positive from today  
> What would you like to discuss in this session?  
* Now I am cure, thanks to you, now I am a positive person.  
> Do you believe it is normal to be cure, thanks to me, now you are a  
positive person?
```

TYPE HERE Yes, thanks a lot

Features of ELIZA:

- **Natural Language Processing:** ELIZA was designed to simulate human-like conversation using natural language processing techniques. It used pattern matching and substitution to generate responses to user inputs.
- **Simple Interface:** ELIZA had a simple text-based interface that allowed users to type in responses to its questions.
- **Rule-Based System:** ELIZA's conversational flow was based on a set of rules that were predefined in the program. The rules were designed to mimic the conversational patterns of a Rogerian therapist, using open-ended questions and reflective statements to encourage users to talk about their feelings.

Limitations of ELIZA:

- **Minimal Contextual Understanding:** ELIZA did not have a deep understanding of the context of a conversation. It relied on superficial pattern matching to generate responses, which often led to non sequiturs or irrelevant responses.
- **Limited Vocabulary:** ELIZA had a limited vocabulary and was not capable of understanding complex sentences or nuances in meaning. It relied on simple keyword matching to generate responses.
- **Limited Domain:** ELIZA was designed to simulate a Rogerian therapist and was limited to conversational topics related to mental health and personal problems. It was not capable of discussing topics outside of this domain.