Ex. 9 Information extraction from text data

Information extraction from text data involves extracting structured information from unstructured text. This can be done using various techniques, including regular expressions, natural language processing libraries like spaCy or NLTK, and machine learning models. Here, I'll provide a Python program using spaCy for named entity recognition (NER) and extracting specific information from text.

First, you'll need to install spaCy and download the appropriate language model. In this program, I'll use the English model:

pip install spacy

python -m spacy download en_core_web_sm

Now, let's create a Python script for information extraction:

In this program, we load the spaCy English language model and process a text document. We extract named entities and specific information related to the founding of an organization (in this case, "Apple Inc.").

You can adapt and extend this code to extract other specific information or use more advanced NLP techniques depending on your specific requirements and the complexity of your text data.

```
import spacy

# Load the English NLP model
nlp = spacy.load("en_core_web_sm")

# Text data to extract information from
text = """
Apple Inc. is an American multinational technology company headquartered in Cupertino, California.
It was founded by Steve Jobs, Steve Wozniak, and Ronald Wayne in 1976.
Apple designs, manufactures, and markets consumer electronics, computer software, and online services.
"""

# Process the text with spaCy
doc = nlp(text)

# Extract named entities (e.g., organizations, persons, locations)
for entity in doc.ents:
    print(f"Entity: {entity.text}, Label: {entity.label }")

# Extract specific information
found keywords = []
for token in doc:
    if token.text.lower() == "founded":
        found_keywords.append(token)

if found_keywords:
    # Assuming we want to extract the organization's founding details
    organization = None
    for ent in found keywords[0].head.children:
        if ent.ent_type_ == "ORG":
            organization = ent.text
    if organization:
        print(f"{organization} was founded by:")
        for child in found_keywords[0].children:
            if child.ent type  == "PERSON":
                print(child.text)
```