**Task 1:**

import org.jsoup.Jsoup;

import org.jsoup.nodes.Document;

import org.jsoup.nodes.Element;

import org.jsoup.select.Elements;

import java.io.IOException;

import java.util.Scanner;

public class QuoteScraper {

public static void main(String[] args) {

Scanner scanner = new Scanner(System.in);

System.out.print("Enter the page number (1-10): ");

int pageNumber = scanner.nextInt();

if (pageNumber < 1 || pageNumber > 10) {

System.out.println("Invalid page number. Please enter a number between 1 and 10.");

return;

}

String url = "https://quotes.toscrape.com/page/" + pageNumber;

try {

Document document = Jsoup.connect(url).get();

Elements quotes = document.select(".quote .text");

System.out.println("Quotes from Page " + pageNumber + ":");

for (Element quote : quotes) {

System.out.println("- " + quote.text());

}

} catch (IOException e) {

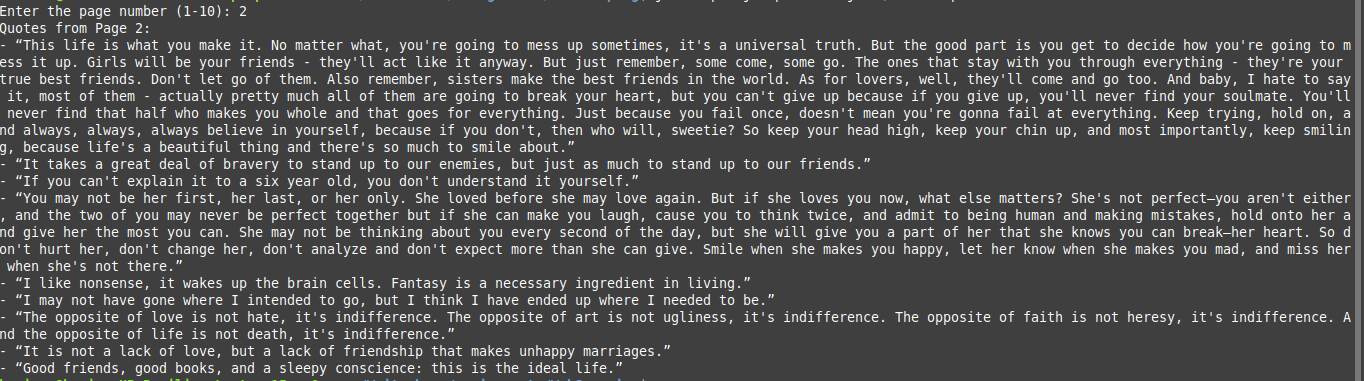
e.printStackTrace();

}

}

}

**Screenshot task 1:**



**Task 2:**

import org.apache.pdfbox.pdmodel.PDDocument;

import org.apache.pdfbox.text.PDFTextStripper;

import org.apache.pdfbox.Loader;

import java.io.File;

import java.io.FileWriter;

import java.io.IOException;

import java.io.Writer;

public class PDFReader {

public static void main(String[] args) {

try {

String file = "USCensus.pdf";

PDDocument document = Loader.loadPDF(new File(file));

PDFTextStripper pdfStripper = new PDFTextStripper();

String text = pdfStripper.getText(document);

String myTxt = "\_USCensus.txt";

Writer writer = new FileWriter(myTxt);

pdfStripper.writeText(document, writer);

System.out.println("Text extracted from PDF:\n" + text);

writer.close();

document.close();

} catch (IOException e) {

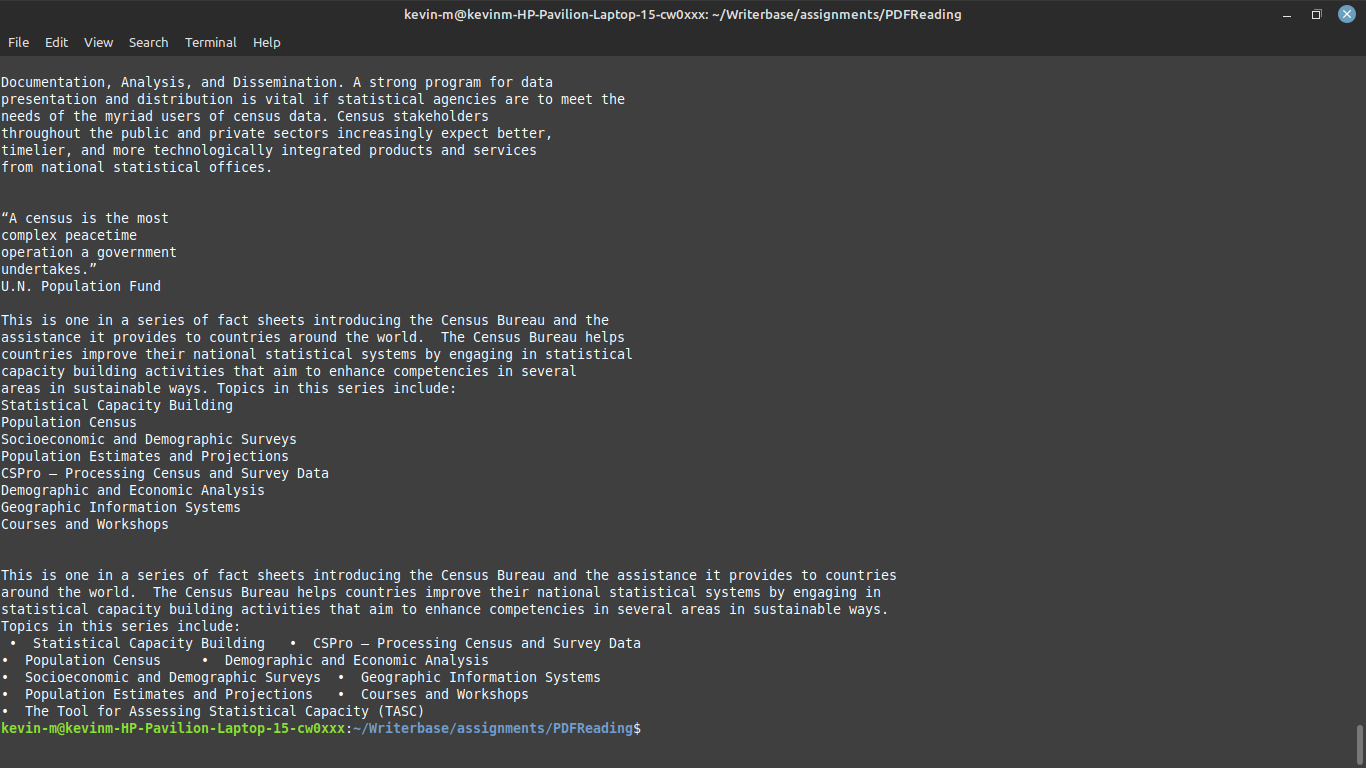
e.printStackTrace();

}

}

}

**Screenshot task 2:**



**Instructions**

### ensure you have Java installed, and download jsoup together with pdfbox.

Navigate to the location where the java files have been stored:

## To Run the WebScraping program:

Run on your terminal:

'javac -cp .:jsoup-1.17.1.jar QuoteScraper.java' then

'java -cp .:jsoup-1.17.1.jar QuoteScraper'

# This will run the java file and generate a some quote data

## To Run the PDFReader program:

Run on your terminal:

'javac -cp .:pdfbox-app-3.0.0.jar PDFReader.java'

'java -cp .:pdfbox-app-3.0.0.jar PDFReader'

# This will generate a new file