# LInC Internships Summer 2020

## Project Descriptions and Requirements

---

| | |
|---|---|
| **Title** | Extract Keyword and/or Keyphrases from Sentences |
| **Project Description** | The purpose of this project is to familiarize with fundamental NLP techniques and apply them to extract keywords and important phrases from a text corpus.<br><br>The project can be split into two main phases. In the first phase the intern will familiarize with NLP techniques and work with Jupyter Notebooks in order to have a solid workflow. In the next phase, the intern will apply NLP methods for extracting keywords over a provided sentence dataset and analyze the results. |
| **Technologies** | • Python<br>• NLTK, SpaCy, AllenNLP<br>• Jupyter Notebooks |
| **Mentors** | Demetris Paschalides <dpasch01@cs.ucy.ac.cy> |

| | |
|---|---|
| **Title** | Infer Directly Affiliated Social Network Accounts of Named Entities |

| | |
|---|---|
| **Project Description** | The purpose of this project is use graph analysis in order to index and relate affiliated social network accounts e.g. Twitter with Named Entities. For example given the named entity “Elon Musk”, we should retrieve accounts like:<br><br>- Twitter: @elonmusk, @tesla<br>- Instagram: @elonmusk, @rareteslas<br>- Facebook: @tesla<br><br>The project can be split into two main phases. In the first phase the intern will familiarize graph analysis techniques and work with Jupyter Notebooks in order to have a solid workflow. In the next phase, the intern will apply these methods over e.g. Twitter network and try to identify the affiliated accounts of given entities. |
| **Technologies** | ● Python<br>● NLTK, Tweepy, NetworkX<br>● Jupyter Notebooks |
| **Mentors** | Demetris Paschalides <dpasch01@cs.ucy.ac.cy> |

| | |
|---|---|
| **Title** | Develop, Deploy and Manage a microservice application using Kubernetes |
| **Project Description** | The purpose of this project is to develop, deploy and manage a microservices application on top of the Kubernetes ecosystem. The intern will have the opportunity to gain useful skills required by modern |

| | |
|---|---|
| | cloud software applications, such as microservices concepts, containerization, CI/CD and orchestration.<br><br>The project can be split into two main phases. In the first phase the intern will develop a distributed web application and package it using docker containers. In the next phase, the intern will deploy the application on top of a Kubernetes cluster making use of necessary features such as, dynamic load balancing, service discovery, monitoring and auto-scaling. |
| **Technologies** | • Java/Python<br>• SpringBoot/Flask(Rest API)<br>• Html/Javascript/css (optional for developing a UI)<br>• Docker<br>• Kubernetes |
| **Mentors** | Zacharias Georgiou <zgeorg03@cs.ucy.ac.cy><br>Moysis Symeonides <msymeo03@cs.ucy.ac.cy><br>Thanasis Tryfonos <atryfo03@cs.ucy.ac.cy> |

| | |
|---|---|
| **Title** | Develop and Deploy a PV inverter monitoring agent for near real-time extraction of power production metrics. |
| **Project Description** | The purpose of this project is to extract in real-time power production measurements from the installed PV inverter from the roof of building FST01.<br><br>To measure production, the intern will need to develop a monitoring agent that, taking into consideration, the specifics of the device, will extract the current measurement and disseminate the measurement to the central monitoring platform that is responsible to store and manage it in the underlying timeseries database. |

| | |
|---|---|
| **Technical Skills** | ● Familiarity with Raspberry Pi and Arduino Programming(Desired)<br>● Docker containers<br>● Understanding of communication protocols HTTP Modbus(desired)<br>● (Time Series) Databases<br>● Front-end technologies (HTML, CSS, JS(Typescript) , Angular(desired)) and visualization libraries<br>● Go programming language |
| **Mentors** | Thanasis Tryfonos |

| | |
|---|---|
| **Title** | Apache Spark vs Apache Flink: Which is better? |
| **Project Description** | The purpose of this project is to compare two state-of-the-art big data processing engines - Apache Spark and Apache Flink. The intern will use containerization technologies (e.g., docker-compose) to automate the setup, bootstrap and deployment of the big data engines on a cloud infrastructure and then evaluate their performance under different configurations and scenarios.<br><br>The intern will have the opportunity to gain useful skills in data engineering, such as, deployment, maintenance and management of big data processing engines, and containerization technologies. |
| **Technologies** | ● Docker<br>● Docker-Compose |

| | |
|---|---|
| | • Apache Spark<br>• Apache Flink |
| **Mentors** | Zacharias Georgiou <zgeorg03@cs.ucy.ac.cy><br>Moysis Symeonides <msymeo03@cs.ucy.ac.cy> |

| | |
|---|---|
| **Title** | How different is your Twitter network from you? |
| **Project Description** | ● **Data:** Retrieve twitter profiles, tweets and social connections (e.g. followings and followers) using a specific sample of individuals (e.g. entrepreneurs, etc.).<br>● **Sentiment Analytics and Topics:** Use pre-trained machine learning algorithms for identifying the sentiment and topics of tweets.<br>● **Characteristics of Users vs. their social network:** Find in which aspects (e.g. sentiment, number of followers, etc.) specific users differentiate from their social network users (e.g. followings and followers).<br>● **Machine Learning:** Build a simple prediction model in order to automatically predict Twitter accounts that belong to specific groups of people (e.g. entrepreneurs, etc.) based on their network properties. |
| **Key Technologies** | ● Python<br>● HTTP requests to APIs (e.g. Twitter API)<br>● Database (e.g. MongoDB)<br>● NLTK<br>● Scikit-learn |
| **Mentors** | Dimos Stefanidis |

| | |
|---|---|
| **Title** | Detect Twitter accounts belonging to companies |
| **Project Description** | ● **Data:** Collect tweets and Twitter profile info of the companies and individual users from the Twitter API using a specific sample of companies and individual users. |

| | |
|---|---|
| | ● **Sentiment Analytics and Topics:** Use pre-trained machine learning algorithms for recognizing the sentiment and the topic of each tweet.<br>● **Companies accounts vs. Individual users:** Analyze the companies and the individual users Twitter activity in order to identify significant features and differences between them.<br>● **Machine Learning:** Build a simple prediction model in order to automatically predict the Twitter accounts that belong to companies. |
| **Key Technologies** | ● Python<br>● HTTP requests to APIs (e.g. Twitter API)<br>● Database (e.g. MongoDB)<br>● NLTK<br>● Scikit-learn |
| **Mentors** | Dimos Stefanidis |

| | |
|---|---|
| **Title** | Fake News: Evolution Analysis |
| **Project Description** | The rise of online misinformation spread on news media and social network platforms is considered a major societal challenge. The identification of them is yet an open issue. A crucial step to support fake news detection is to analyze the evolution of them through the years.<br><br>**Responsibilities:**<br><br>● Crawl famous fact-checking sites such as PolitiFact and Snopes.<br>● Implement known similarity check algorithms like SimHash.<br>● Verify the trustworthiness of already collected articles.<br>● Analyze the outcomes and provide a visual representation of them. |
| **Key Technologies** | ● Python<br>● Scrapy<br>● NLTK<br>● Scikit-learn<br>● Elasticsearch<br>● Kibana |
| **Mentors** | Chrysovalantis Christodoulou <cchris47@cs.ucy.ac.cy> |

| | |
|---|---|
| **Title** | Annotation Platform |
| **Project Description** | Machine learning models have gained an increasing amount of interest due to the ability of them to learn from data and decide the best possible solutions. Supervised machine learning applications need annotated data in order to train themselves. For example, on fake news detection, we have to label if an article is fake or not. The goal of this project is to provide a user-friendly annotation platform which will let the users to manually label data.<br><br>**Responsibilities:**<br><br>● Implementation of a user-friendly platform.<br>● Implementation of a REST API using NodeJS.<br>● Store the data into a MongoDB database. |
| **Key Technologies** | ● HTML<br>● CSS<br>● JAVASCRIPT<br>● NODEJS<br>● MONGODB |
| **Mentors** | Chrysovalantis Christodoulou <cchris47@cs.ucy.ac.cy><br><br>Joanna Georgiou <jgeorg02@cs.ucy.ac.cy> |

| | |
|---|---|
| **Title** | Data visualization framework |
| **Project Description** | Data is everywhere and we process them in order to extract knowledge. The key to easily understand this knowledge and communicate it, is a well formed visualization. The goal of this project is to visualize time-series data and give the opportunity to the user to interact with this visualization. Part of the project is searching the most efficient ways of doing so.<br><br>**Responsibilities:**<br>● Implementation of a data visualization framework.<br>● Integration with existing data sources. |
| **Key Technologies** | ● Python<br>● Python data visualization libraries |
| **Mentors** | Georgios Argyriou <argyriou.georgios@cs.ucy.ac.cy> |

| | |
|---|---|
| **Title** | Data Storage on Blockchain |
| **Project Description** | The purpose of this project is the development of a data storage platform based on the blockchain. Initially, the intern will study the blockchain technology and do some research on the appropriate ways to store data on it. The intern will then develop a smart contract based on the research that has been done.<br><br>The intern will have the opportunity to become familiar with the blockchain technology and learn both its advantages and the limitations that exist in the field. She/he will also gain useful skills on cryptography and smart contract development. |
| **Technologies** | • Blockchain<br>• Solidity<br>• JavaScript<br>• NodeJS |
| **Mentors** | Ioannis Savvidis<br><savvidis.ioannis@ucy.ac.cy> |