

Informatics Institute of Technology  
In Collaboration with  
University Of Westminster



# **THE ARCHITECT**

## **Automated Software Architecture Diagram Generator using Natural Language Processing**

A Dissertation by  
**Mr Yasitha Lalanga Gamage**

Supervised by  
**Mr Iresh Bandara**

May 2023

Submitted in partial fulfilment of the requirements for the  
BSc (Hons) in Computer Science degree at the University of Westminster, UK

## ABSTRACT

In software engineering, software architecture serves as a key component for building complex software solutions. Software architecture diagrams provide a high-level view of the system's core structure, including its components, relationships and the necessary dependencies. Software architecture diagramming is considered a critical stage in the development process, as its layout is the foundation for the development lifecycle. However, the manual creation of diagrams can greatly be affected by human error, is time-consuming and lead to inconsistencies throughout the software development life cycle (SDLC).

Automating the process of generating architecture diagrams can help to overcome these challenges by leveraging complex pattern-matching algorithms and entity extraction techniques. The proposed automated application employs a novel pipeline for information extraction and diagram generation, incorporating a custom Named Entity Recognition Model and a Masked BERT Relation Classifier (Context + Mention) along with shallow heuristics and a knowledge graph for diagram relation mapping. The application also supports Diagram Structural language (DSL) for diagram editing.

The proposed application automates the process of software architecture diagramming by utilizing efficient methods, resulting in accurate and easily comprehensible diagrams with minimal errors. It utilizes the c4 modelling techniques to generate dependable architecture diagrams. The application evaluates the requirements based on the data extracted and generate abstract-level diagrams in system context-level and container-level.

**Key Words:** Software Architecture Diagram, Natural Language Processing, C4 Model, Transformers, Relation Extraction, Named Entity Recognition

### Subject Descriptors:

- Computing methodologies → Artificial intelligence → Natural language processing → Information extraction
- Software and its engineering → Software notations and tools → System description languages → Architecture description languages
- Computer systems organization → Architectures → Distributed architecture