C1: Research Computing

Max Talberg

December 17, 2023

Contents

0.1 Introduction

This short report presents the development process of a Sudoku solver project, developed in Python. The emphasis of this project is to implement robust software development practices in conjunction with the development of sophisticated coding techniques. Throughout the report, we will detail these processes and the challenges faced in their implementation. We will also discuss prototyping, unit testing, version control, optimisation, documentation and the use of containerisation tools, providing insight into the software development process.

0.2 Algorithm Design and Prototyping

The initial phase of the Sudoku project involved conceptualising the entire development process. This prototyping step outlined a high-level workflow, which we refer to as the larger software development cycle (LSDC). The LSDC encompasses all of the phases involved in good software development practices, from the initial planning to the final deployment. Within this high-level workflow, we focused on the prototyping logic specific to the algorithm design which is to be coded. Figure 1 illustrates the high-level LSDC and the detailed prototyping logic.

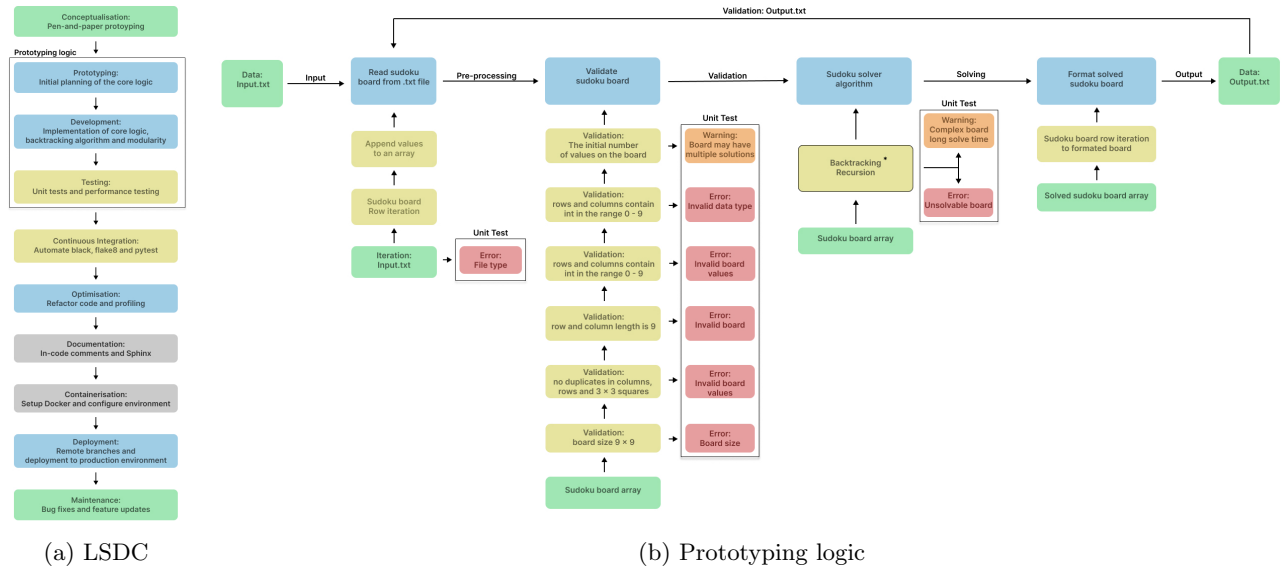(a) LSDC   (b) Prototyping logic

Figure 1: (a) High-level workflow of the entire Sudoku solver project, detailing the integration of appropriate software development practices with the core algorithmic logic. (b) A closer look at the detailed prototyping of the core algorithmic logic to be coded.

The first step when implementing the prototyping logic is to understand the constraints and laws of Sudoku. This project is built on the following classic rules of Sudoku[1]: The objective is to fill a 9 x 9 grid with digits so that each column, row and 3 x 3 square contains all the digits from 1 to 9, without repetition. A well-constructed Sudoku has a single solution, with a minimum number of 17 clues provided[2].

The coding process began in a singular Python file, containing functions for reading the puzzle, solving it and printing the output. This initial approach evolved into a more sophisticated structure with error handling and classes. To improve modality we transitioned to four separate classes: **SudokuReader**, **SudokuBoard**, **SudokuAlgorithm** and **SudokuFormat**. Generating these four classes in individual files greatly improved the readability and facilitated targeted optimisation.

0.2.0.1 SudokuReader Class

This class is responsible for the reading of a Sudoku board from a text file. It parses the file's contents, transforming the Sudoku grid into a structured list of lists for further processing. This initialisation is key as it ensures the board is in an appropriate format for the validation and solving algorithms.

0.2.0.2 SudokuBoard Class

This class is the central component of the puzzle processing and handles all things related to the Sudoku board. It serves two fundamental roles: first, it transforms the initial representation of the Sudoku board from a list of string values into a 2D list of integers, in preparation for validation. Second, it undertakes the validation process, ensuring the board adheres to the classical rules of Sudoku, as well as some additional edge cases that are discussed in more detail later. Upon confirmation that all the conditions are satisfied the class returns 'True' status which signals a valid Sudoku board and allows the solving process to begin.

0.2.0.3 SudokuAlgorithm Class

This class is the core of the Sudoku solver, executing a backtracking algorithm to solve puzzles. After receiving a validated 2D list of integers, it employs backtracking a method similar to a brute force that systematically traverses the Sudoku grid in search of a solution [3]. Backtracking is a recursive algorithm, in this context the algorithm cycles through possible values (1 to 9) for each cell, advancing when a valid number is found or backtracking when a dead end is reached. This iteration process continues until the algorithm finds a solution or exhausts all possible values for the grid and determines the puzzle to be unsolvable.

Although backtracking is straightforward and effective for simpler puzzles, it becomes less efficient with increasing difficulty. Recognising this, we considered implementing constraint propagation on top of the backtracking algorithm. Constraint propagation uses the rules of Sudoku to prune the Sudoku gird and guide the solver to areas where solutions are more likely to be found [4]. Much later on this led to the implementation of the Minimum Remaining Value (MRV) heuristic [5]. The MRV heuristic is an addition to the backtracking algorithm, it chooses squares with the fewest legal moves. By prioritising the most solvable squares the MRV heuristic improves the efficiency of the backtracking algorithm.

This was the only other algorithm we had time to implement. The ***SudokuAlgorithm*** was initially developed with the idea that it could contain multiple solving algorithms, to later launch an optimisation investigation comparing algorithm performance on puzzles of varying difficulty.

0.2.0.4 SudokuFormat Class

This class handles the formatting of the solved Sudoku board. It takes the solved 2D list and iteratively transforms the list into the specified output format to be printed on the console.

0.3 Development and Unit Testing

This section will take a closer look at the development process and unit test practices adopted during the project.

0.3.0.1 SudokuRead Class

In the early stages, error handling was integrated during the reading phase. Invalid characters and incorrect row lengths would be flagged as the data was iteratively read in. However to maintain modularity and due to the low computational cost, validation checks were isolated from the reading function. The final class only reads text files and returns a 2D list of strings, with error handling for non-existent files, incorrect file types or empty files set as runtime errors and captured in unit tests (`tests/test_sudoku_read.py`).

0.3.0.2 SudokuBoard Class

This class began by handling column-related errors but was refactored to perform all the validation tasks. The class checks the rules of Sudoku are being followed: It ensures board values are integers in the appropriate range, with no duplicates in any row, column or 3 x 3 square. The class also handles edge cases: Including invalid board sizes, invalid board format and text files containing normal text. These errors are set as value errors and captured in unit tests (`tests/test_sudoku_board.py`). We also implemented warnings for scenarios likely to lead to multiple solutions, such as a puzzle with fewer than 17 clues. This was informed by the characteristics of a well-constructed puzzle and an investigation into the solver's time complexity, concerning the number of initial empty squares. A more detailed analysis of these findings and their implications is explored in the Advanced Experimentation and Refinement section.

0.3.0.3 SudokuAlgorithm Class

With the assumption of a validated board, the algorithmic part of the code has few error-handling instances. Unit tests confirm the solver's performance and solution accuracy (`tests/test_sudoku_algorithm.py`), as well as the edge case of an unsolvable puzzle. During this development, we utilised a combination of mocking and data file imports for testing. Importing files improves readability and replicates the end application use. Whereas the mocking board ensures testing isolates components, as we are not relying on reading and validating a puzzle before it is solved, we have just isolated the expected solution.

0.3.0.4 SudokuFormat Class

This class was initially part of the validation function, although was separated for modularity, following the initial prototype logic. It converts the solved board as a 2D list of integers into a formatted string for output, with unit tests to ensure correct conversion (`tests/test_sudoku_format.py`). This test used mocking to isolate the unit test to just the formatting class.

0.3.0.5 Continuous Integration

The use of Continuous Integration (CI) early on in the coding process proved to be incredibly useful. CI was integrated into the project using pre-commit hooks with Black, Flake8 and Pytest, which enforced the PEP8 styling and conducted continuous testing. CI meant that at each commit all of the unit tests would run and the code would be formatted and styled. This ensured early bug detection and consistent code quality. This was particularly useful when transitioning from a singular-class project into a more complex four-class system.

0.4 Advanced Experimentation and Refinement

0.4.0.1 Exploring Edge Cases and Complex Scenarios

The design of our unit tests initially focused on ensuring compliance with Sudoku rules, with code identifying and specifying the location of errors. Through research and brainstorming [6], we identified several edge cases including: Invalid board sizes, invalid board format, inputs that didn't represent a board at all and unsolvable puzzles. After developing unit tests and error codes for these edge cases it became clear there are scenarios where the solver should issue warnings without stopping execution. An example of this is when scenarios produce multiple solutions, even when following classical Sudoku rules. This can typically occur when there are less than 17 clues, this led to some experimentation to understand the solvers performance at various numbers of clues.
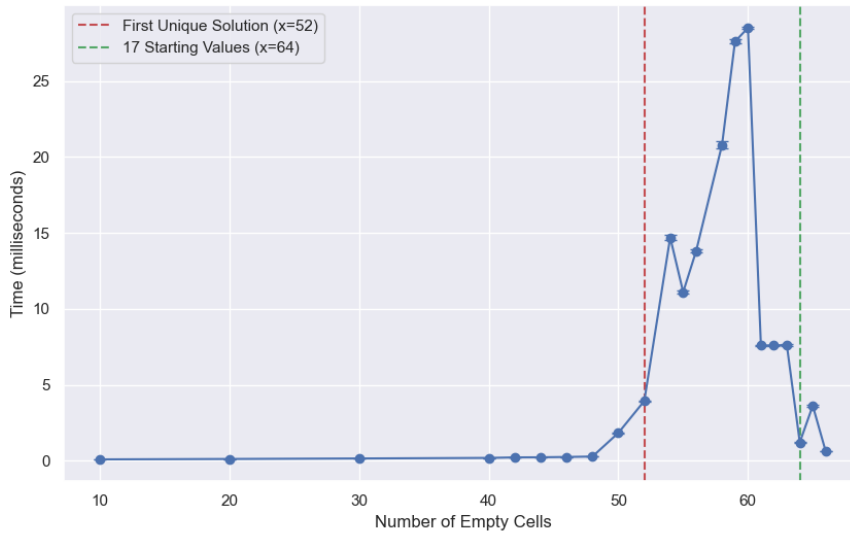
Figure 2: Time complexity of the backtracking algorithm

Experimentation of boards with less than 17 clues led to an investigation into the time complexity of the backtracking algorithm. This involved measuring how long our solver took to solve a Sudoku puzzle while systematically varying the number of initial empty squares. This study revealed a significant increase in solve time and non-uniqueness of solutions when the empty cells exceeded a certain threshold (52 empty cells). There is less variability in this puzzle complexity forcing the backtracking algorithm to deeper levels of recursion. After 61 empty cells, the time taken significantly drops, likely due to the increased accessibility of the numerous potential solutions.

To address this problem, warnings were implemented for boards with fewer than 17 clues or empty boards. The code still runs, it will just produce a warning that there is a possibility of multiple solutions.

Although this was only performed on one puzzle , figure 2 gave practical insight into the performance characteristics of Sudoku boards with a high number of empty cells.

Iterative Development and Branch Management

Utilising Git alongside CI maintained a high quality of code throughout development. Branching allowed for safe experimentation and feature development without risking the functionality of existing work. The evolution of the project is evident in version control: the first release branch (V1.0.0) contained a single class in one file, while the third release branch (V3.0.0) had four separate files for source and testing. Version facilitated continuous development while preserving previous work done.

0.4.0.2 Flask

An exploratory extension was developing a user interface (UI) for the Sudoku solver using Flask [7]. Flask is a lightweight Python module for developing web applications. The layout of the interface was developed in HTML, this simple layout consisted of an option to choose or drag a file, an upload button and a solve button. The upload button is linked to the **SudokuRead** and **SudokuBoard** classes, once clicked the text file is read and the puzzle is validated. The read-in board will fill up the on-screen board, unless an error is thrown in which case the relevant error message will pop up. The solve button is linked to the **SodukuAlgorithm** class, once clicked the on-screen board becomes solved, unless an unsolvable board error occurs. Time permitting, future enhancements could include an animation of the backtracking process or other algorithms.
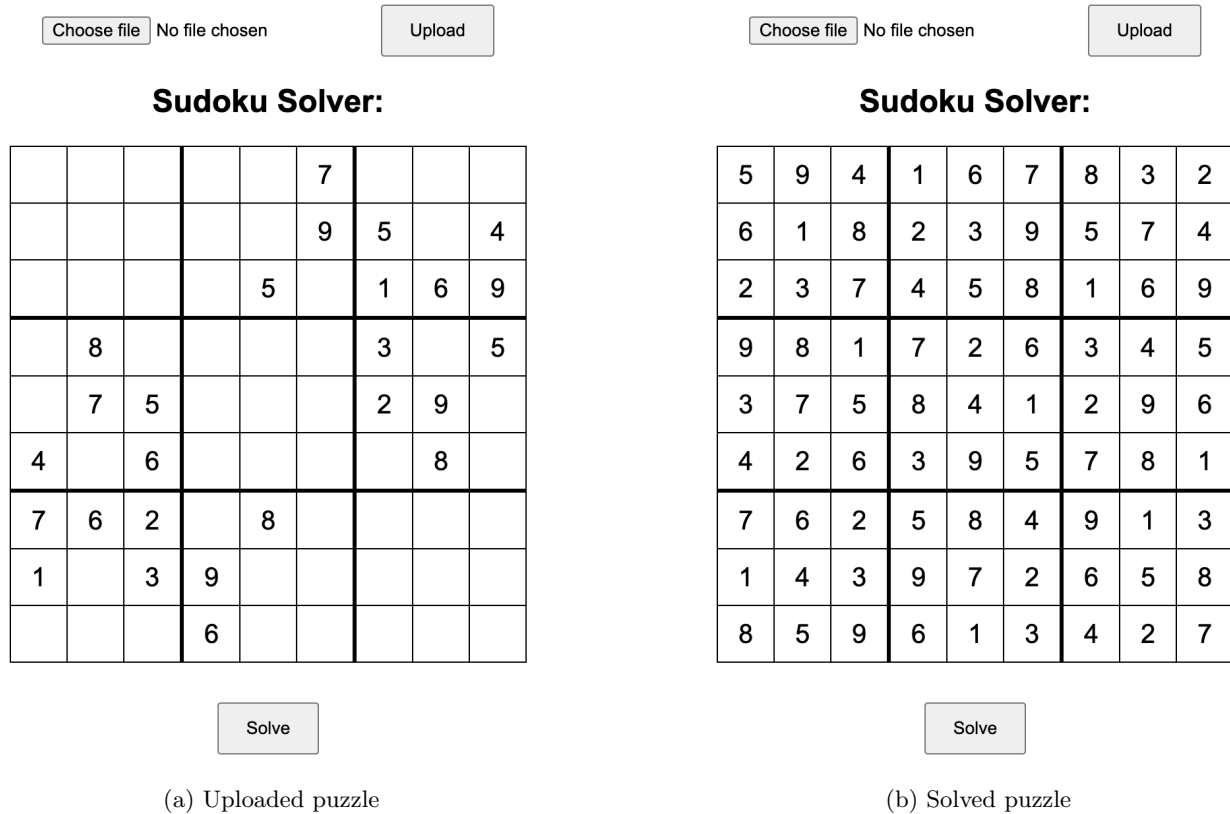
(a) Uploaded puzzle

(b) Solved puzzle

Figure 3: Figures show the UI before and after puzzle resolution, illustrating the solver's user interaction flow.

0.4.0.3 Performance Profiling and Optimisation

Profiling and optimisation are essential processes in enhancing the efficiency of code. After releasing the V1.0.0, we applied the `%timeit` to measure the time complexities of the code. This is expanded upon in the Benchmark Time Complexity Investigation below.

In the initial development of the validation code, checks for duplicates in the rows, columns and 3 x 3 squares were performed in three separate iterations. Optimising the code led to these checks being consolidated into a single iteration. Additionally, more extensive use of list comprehensions was implemented. Although solving a Sudoku isn't computationally intensive, such optimisations are crucial in more complex projects that require scaling up.

0.4.0.4 Benchmark Time Complexity Investigation

puzzles_kaggle	puzzles/sec
rc_mt942v3.0.0	31.3
rc_mt942v3.1.0	10.27
fsss [8]	1,477,910.9
jczsolve [9]	597,074.2
sk_bforce2 [10]	1,234,240.0

Table 1: Performance comparison of different Sudoku solvers on a Kaggle dataset of 1,000,000 puzzles.

Table 1 presents a performance benchmark using Kaggle dataset of 1,000,000 Sudoku puzzles [11]. It compares puzzles solved per second between our solver, rc_mt942v3.0.0 and several leading solvers. While rc_mt942v3.0.0 successfully solved all puzzles, its rate (0.17 puzzles/sec) is significantly lower than others like fsss2, jczsolve3, and sk_bforce2, which solve tens of thousands per second.

puzzles_magic_tourtop1465	puzzles/sec
rc_mt942v3.0.0	0.17
rc_mt942v3.1.0	1.13
fsss2	69,752.5
jczsolve3	77,888.4
sk_bforce2	86,535.4

Table 2: Performance comparison of different Sudoku solvers.

In Table 2, we assess rc_mt942v3.0.0 against the Magic Tour Top 1465 dataset, a challenging set of puzzles [12]. Our solver's performance, although successful in solving all puzzles, remains slower in comparison.

Analysis revealed the competitor solvers use an advanced and optimised brute-force approach with bitwise operations [13]. Notably, these solvers are implemented in C. The bitwise operations allow the solver to handle multiple candidates for multiple cells simultaneously, when executed in C it is clear why these solvers have such a performance advantage.

Utilising the `%timeit` function to pinpoint performance, revealed that the clear bottleneck of our code is the backtracking algorithm. The process of reading, validating and formatting all take nanoseconds, while the algorithm class requires microseconds.

Based on this bottleneck we implemented the minimum related value (MRV) heuristic into our Sudoku solver (rc_mt942v3.1.0) to enhance the backtracking algorithm. The MRV heuristic chooses the square that has the fewest legal values remaining and guides the backtracking algorithm to this square. To investigate the performance we ran the solver through the benchmark tests. As expected the solver was much faster solving the Magic Tour Top dataset, almost an order of magnitude faster, demonstrating MRV's effectiveness on difficult puzzles. However, on the simpler Kaggle dataset, the solver showed a slowdown, suggesting an overhead in MRV calculations for easier puzzles. Despite the decrease in speed, the MRV is fast for hard Sudoku puzzles and unless the user is iterating through a million puzzles the additional MRV heuristic has a better net performance than just the backtracking algorithm.

A potential optimisation, given more time, would be integrating the Degree Heuristic alongside the MRV heuristic during backtracking [14]. The Degree Heuristic assigns a value to the variable involved in the largest number of constraints. Paired with MRV this provides a balance between searching for the most constrained variables (MRV) and those that have the greatest impact on the rest of the puzzle (Degree).

Furthermore, the adoption of Cython or a complete transition to C could prove beneficial. Cython enables the use of the speed benefits of C while maintaining the readability and structure of our Python code. Alternatively, rewriting the code in C would provide the performance benefits of a compiled language compared to an interpreted language, although sacrifices readability and ease of maintenance.

Further research could focus on a comparative analysis of various algorithms against these benchmark datasets .

0.5 Packaging and Usability

The project commenced with the creation of a dedicated Conda environment. This isolates the current project allowing control over the dependencies used.

0.5.0.1 Dockerfile and Environment Setup

Deploying a project via Docker ensures consistent performance across different hardware configurations. Our setup includes a Dockerfile, which contains a series of commands to build an image of the project. This includes specifications for working directories, an exposed port for the Flask application and the project's environment. The Dockerfile references an `environment.yml` file that lists all the necessary environmental dependencies. These dependencies are installed during the image build, ensuring the project runs reliably in the Docker container.

0.5.0.2 Documentation

Sphinx was chosen for documentation due to its aesthetic appeal compared to Doxygen. Sphinx is integrated with the code base, automatically updating after each run. Implementing a Continuous Development (CD) pipeline would streamline this process automating documentation updates and making them available on GitLab. This is however only possible in a public directory. Our project follows NumPy style docstrings, by implementing `sphinx.ext.napoleon` in the `conf.py` file ensures Sphinx correctly interprets this notation. The Dockerfile was setup so the Sphinx documentation would build when the Dockerfile is built.

Documenting code is key for simplifying code maintenance and debugging. Using Sphinx, we can centralise the documentation for each release branch or commit. Facilitating easier access and comprehension of the code throughout the project.

0.6 Summary and Future Work

In this project, we developed a functional Sudoku solver using the backtracking algorithm compatible with the classical rules of Sudoku.

The project began by prototyping the high-level workflow with the LDSC and the core algorithmic logic, which was then turned into code. This initial version of the code was later refactored into four distinct classes: ***SudokuReader***, ***SudokuBoard***, ***SudokuAlgorithm*** and ***SudokuFormat***. Collaboratively these classes solve Sudoku puzzles.

Further complexity was introduced with advanced error handling within these classes. Addressing issues such as breaches of the classical rules, handling of incorrect inputs and warnings for less than 17 starting values. These errors were integrated into our unit tests to ensure the robustness of the code.

The incorporation of CI tools like Black, Flake8 and Pytest ensured automatic testing with every commit. This greatly aided the development process by ensuring early bug detection and a consistently clean codebase.

Once the core logic was established and functioning alongside unit tests, the project entered an advanced stage of experimentation and refinement. This involved optimising the backtracking algorithm by implementing the MRV heuristic and assessing the model's performance on a variety of benchmark datasets, alongside renowned Sudoku solvers. Additionally, this experimentation stage resulted in the development of a simple UI using Flask.

The final stages of the LSDC consisted of refining the project docstrings for documentation via Sphinx. We also developed a Docker image, ensuring a self-contained environment for easy deployment across different devices.

Potential Areas for Future Improvement

The most significant area for improvement in Sudoku performance lies in optimising the MRV backtracking algorithm, which currently is the largest bottleneck. Future work should focus on implementing the Degree heuristic to enhance the algorithm's efficiency. Additionally, integrating Cython into the project will offer faster solutions due to the performance advantage of compiled languages. Alternatively, implementing an advanced and optimised brute-force approach in C, similar to the leading solver, might be the fastest approach.

Another area for development is the sophistication of the error-handling system, particularly for puzzles with multiple solutions. Given more time, we aim to identify puzzles with multiple solutions and present all of the solutions.

The current front end of the solver is quite basic. Plans would aim to visualise the solving process by integrating an animation of the backtracking process to provide a visualisation of the algorithm.

Bibliography

[1] Sudoku - Wikipedia. 2nd December, 2023. Available from: `https://en.wikipedia.org/wiki/Sudoku`

[2] Gary McGuire, Bastian Tugemann, Gilles Civario, There is no 16-Clue Sudoku: Solving the Sudoku Minimum Number of Clues Problem, arxiv:1201.0749.

[3] Backtracking - Wikipedia. 2nd December, 2023. Available from: `https://en.wikipedia.org/wiki/Backtracking`

[4] Sudoku Constraint Propagation - Washington University in St. Louis. 2nd December, 2023. Available from: `https://classes.engineering.wustl.edu/cse231/core/index.php?title=Sudoku_Constraint_Propagation`

[5] Constraint Satisfaction with Sudoku - Steven.codes. 14th December, 2023. Available from: `https://steven.codes/blog/constraint-satisfaction-with-sudoku/`

[6] Invalid Test Cases - Sudopedia. 4th December, 2023 Available from: `http://sudopedia.enjoysudoku.com/Invalid_Test_Cases.html`

[7] Flask Documentation. 10th December, 2023. Available from: `https://flask.palletsprojects.com/en/3.0.x/`

[8] dobrićhev, fsss2 - GitHub. 6th December, 2023. Available from: `https://github.com/dobrichev/fsss2`

[9] 3.77us Solver (2.8G CPU) Testcase 17Sudoku - Forum.EnjoySudoku. 6th December, 2023. Available from: `http://forum.enjoysudoku.com/3-77us-solver-2-8g-cpu-testcase-17sodoku-t30470-210.html#p249309`

[10] GPenet, SK/ BFORCE2 - GitHub. 6th December, 2023. Available from: `https://github.com/GPenet/SK_BFORCE2`

[11] Bryan Park. Sudoku Dataset - Kaggle. 6th December. Available from: `https://www.kaggle.com/datasets/bryanpark/sudoku`

[12] Magic Tour Sudoku. 6th December. Available from: http://magictour.free.fr/sudoku.htm

[13] C Arithmetic Operators - Cppreference. 8th December, 2023. Available from: `https://en.cppreference.com/w/c/language/operator_arithmetic`

[14] Java Sudoku Solver-AC3 Algorithm. Codepal. 8th December, 2023. Available from: `https://codepal.ai/code-generator/query/fB2PHcbt/java-sudoku-solver-ac3-algorithm%23:~:text=The%20MRV%20heuristic%20prioritizes%20variables,efficiently%20search%20for%20a%20solution.`

0.7   Appendix

0.7.1   README

Sudoku Solver

This is a Suoku solver written in Python. The algorithm uses MRV backtracking and comes with a user interface written in Flask.

by **Max Talberg**

Using Docker to Run the Sudoku Solver

This application can be run in two modes using Docker: as a command-line tool to process a puzzle file or as a web application. Below are the instructions for both.

Building the Docker Image

Clone the Repository:

- First, clone the repository from GitLab:

```
git clone git@gitlab.developers.cam.ac.uk:phy/data-intensive-science
    -mphil/c1_assessment/mt942.git
```

- Navigate to the project directory:

```
cd mt942
```

Build the Docker Image:

- Build the Docker image:

```
docker build -t sudoku-solver .
```

Running the Command-Line Application with Docker

Internal files

Run the Docker Container:

- Execute the following command in your terminal, to access internal data files contained in Docker. This example uses the `input.txt` file:

```
docker run sudoku-solver data/input.txt
```

- Other files in the data directory include:
    - `empty_board.txt` - a sample empty board
    - `empty_file.txt` - a sample empty file
    - `input.txt` - a sample input board
    - `invalid_board.txt` - a sample invalid board
    - `invalid_character_board.txt` - a sample invalid character board
    - `invalid_column_board.txt` - a sample invalid column board
    - `invalid_row_board.txt` - a sample invalid row board
    - `invalid_size_board.txt` - a sample invalid size board
    - `invalid_square_board.txt` - a sample invalid square board
    - `less_than_17_board.txt` - a sample board with less than 17 clues
    - `slow_unsolvable_board.txt` - a sample unsolvable board
    - `solved_board.txt` - a sample solved board
    - `not_a_text_file.pdf` - a sample non-text file

External files

Prepare Your Input File:

- Place the input file (e.g., `puzzle.txt`) in a known directory on your machine.

  Run the Docker Container:

- To test out the solver on your own puzzles replace `/path/to/local/directory` with the path to the directory containing your puzzle file.

- Replace `puzzle.txt` with the name of your puzzle file and execute the following:

```
docker run -v /path/to/local/directory:/usr/src/app/data sudoku-
    solver data/puzzle.txt
```

Running the Web Application with Docker

1. **Run the Docker Container for the Web App:**

   - To run the web application, use the following command:

```
docker run -p 8888:80 sudoku-solver --web
```

2. **Access the Web Application:**

   - Open a web browser and navigate to `http://localhost:8888`.
   - The user interface is intuitive, uploading a file is performed by drag and drop or by clicking the 'Choose file' button.
   - Click 'Upload' to upload the puzzle to the board.
   - Click 'Solve' to solve the puzzle.

Running the Sudoku Solver Locally

This section provides instructions on how to run the Sudoku Solver project directly on your local machine.

Setting Up the Project

1. **Clone the Repository:**

   - Clone the repository from GitLab:

```
git clone git@gitlab.developers.cam.ac.uk:phy/data-intensive-
    science-mphil/c1_assessment/mt942.git
```

2. **Set Up a Virtual Environment:**

   - Navigate to the project directory:

```
cd mt942
```

   - Create a virtual environment:

```
conda env create -f environment.yml
```

- Activate the virtual environment:

```
conda activate sudoku-env
```

Running the Application

1. **Run the Command-Line Application:**

   - To solve a puzzle from the data file:

```
python src/solve_sudoku.py data/input.txt
```

2. **Run the Web Application Locally:**

   - If your project includes a web application:

```
python src/solve_sudoku.py --web
```

   - Access the web application by navigating to `http://127.0.0.1:80` in your web browser.

Notes

- This setup assumes the use of `conda` over `pip`.

Documentation

To build the documentation for the Sudoku Solver project, follow these steps:

Docker

- When the Docker image for the Sudoku Solver is built, the project's Sphinx documentation is automatically built.

Local

1. **Navigate to the `docs` directory:**

   - Navigate to the `docs` directory:

```
cd docs
```

2. **Build the Documentation:**

   - Build the documentation:

```
sphinx-build -b html source build
```

   - The documentation can be viewed by opening the `index.html` file in the `build` directory.

Testing

To run tests for the Sudoku Solver project, follow these steps:

Running Unit Tests

- Before running tests, ensure that the Python path is set up correctly to include the project's source code.

1. **Navigate to the project directory:**

```
cd mt942
```

2. **Modify PYTHONPATH:**

```
export PYTHONPATH=$PYTHONPATH:/src
```

3. **Run Unit Tests:**

```
pytest
```

License

This project is licensed under the MIT License - see the LICENSE file for details.

Use of Generative Tools

This project has utilised GitHub Copilot for code completion and assistance with text prompts.

- Although these tools aid the development process, its use was limited and only effective in scenarios where the developer had a clear idea of the code's structure and implementation.

- Other uses of these tools were used to aid the development of the project's documentation.

For instance:

- Plotting of functions

- Implementation of Latex and README

- Auto-complete of docstrings