

Interim Demo Report

Student ID: 130012271

3rd Feb 2017

Progress

Currently the project has a functional predictor that can make predictions given what past users have drawn with the tool. The architecture of the program is described in figure 1.

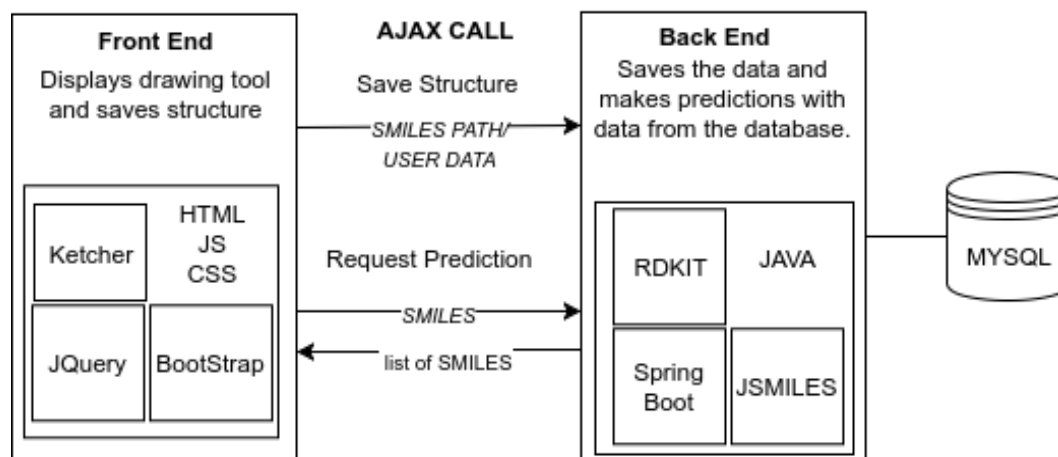


Figure 1: System Architecture

Front-End

The front end is a simple component that allows the user to draw structures and then save them. As it draws the front end makes requests to the back end and it returns a list of smiles strings that might be where the diagram is going next.

A SMILE string is a string representation of a molecular structure and along with a molfile it can be inputted into the Ketcher tool. Then this can be edited by the users. The Ketcher tool is an open source drawing tool for molecular structures that I am using for the project.

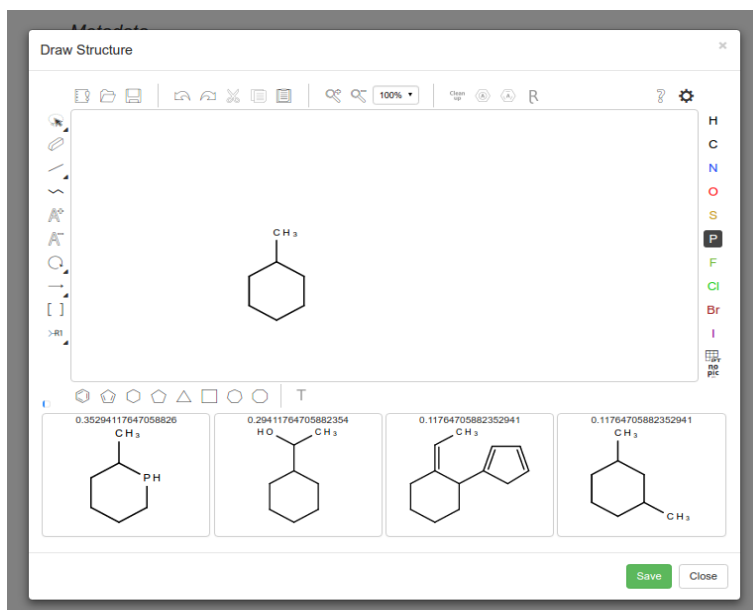


Figure 2: Front End Screenshot

Back-End

The back-end saves the structures into a database along with data that describes the number of times the user has drawn that structure. When a prediction is requested it uses the JSMILES library to create a Bayesian network of what the next structure or step of the drawing the user might take. It uses Bayes theorem to calculate the probabilities for each next step that the user could draw, given the past information.

Database

The database is best described by this script that creates the tables:

```
CREATE TABLE structure (smiles VARCHAR(200), mol VARCHAR(20000),
end TINYINT(1) DEFAULT 0, PRIMARY KEY (smiles));
```

```
CREATE TABLE edge(smiles_to VARCHAR(200), smiles_from VARCHAR(200),
PRIMARY KEY (smiles_to, smiles_from),
FOREIGN KEY (smiles_to) REFERENCES structure(smiles),
FOREIGN KEY (smiles_from) REFERENCES structure(smiles));
```

```
CREATE TABLE edge_metadata(user_id INTEGER, group_id INTEGER,
times INTEGER not null default 1, last_date TIMESTAMP,
```

```
smiles_to VARCHAR(200), smiles_from VARCHAR(200),  
FOREIGN KEY (smiles_to) REFERENCES edge(smiles_to),  
FOREIGN KEY (smiles_from) REFERENCES edge(smiles_from),  
PRIMARY KEY (user_id, smiles_to, smiles_from));
```

Future Work

- Design a study to test the program and see if the predictions make any difference to user experience. This would be in terms of speed of drawing.
- Make the overall predictions better by increasing the number of tables used by the Bayesian network
- look into comparison of molecular structures and create this type of prediction also. This could allow me to compare this type of prediction and my current version.