OSSMatch: A MATCHMAKING TOOL FOR DEVELOPERS TO GET CONNECTED TO OSS PROJECTS

Jevin Dement, Joseph Danciu, Ugo Dike, Stavros Triantis, Adriana Aguilera

Mentored by Fabio Santos | Sponsored by Dr. Igor Steinmacher

School of Informatics, Computing & Cyber Systems, Northern Arizona University, Flagstaff, AZ

College of Engineering, Informatics, and Applied Sciences

MOTIVATION

Open source software is software that is developed in a public collaborative environment and is free to the public to inspect, modify, enhance, and distribute[1]. Often these projects are developed on platforms such as GitHub and by volunteer developers. However, contribution is often difficult for newcomers. GitHub can be a tough place to navigate, on top of sorting through thousands of projects and determining whether or not you have the skills necessary to contribute. Contributors are necessary for OSS to exist. The more people contribute, the faster these software projects are developed, and the better the improvements. Our client Dr. Igor Steinmacher has been working on ways to facilitate entry for these newcomers. The solution he has come up with is a simple web application to ease the information overload, and pair developers with OSS projects tailored to each individual's skill set.

KEY FEATURES
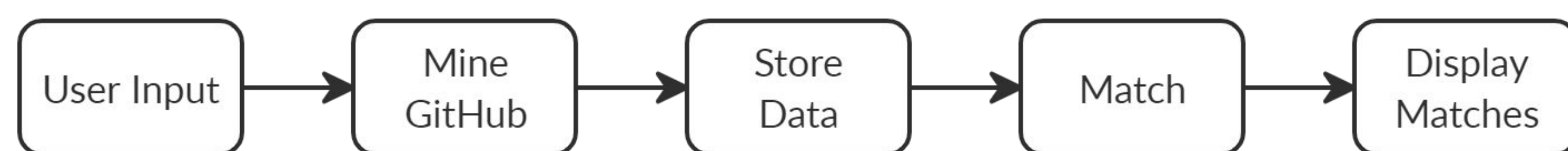
- **User Input:** Users enter skills on a questionnaire, as well as their GitHub username.
- **Mine GitHub:** User profiles and projects are mined for skills and requirements.
- **Store Data:** Projects requirements and user skills will be stored in their respective databases.
- **Match:** Matching algorithm will compare user skills and and project requirements, then proceed to make matches.
- **Display Matches:** Matches will be displayed within the UI.

CHALLENGES

CHALLENGE	RESOLUTION
Accuracy of Matches	Weighted system
Identifying what a Skill is from a Java Import	Classified skills by general category
Database Structure	Normalize Database
User/Account Security	Github Oauth

SOLUTION OVERVIEW

1. Users create an account or login into their existing account.
2. Users then have the option to fill out a questionnaire with skills not demonstrated on their GitHub account. The questionnaire is required if they do not have an existing account.
3. If the user has an existing account, they will be prompted to authenticate via GitHub. Our mining process will initiate and extract skills from repositories they have contributed to.
4. These skills will be used to match users to projects in our database based on a weight system.
5. User will be able to view their matches via the user interface where they will be stored on their account.
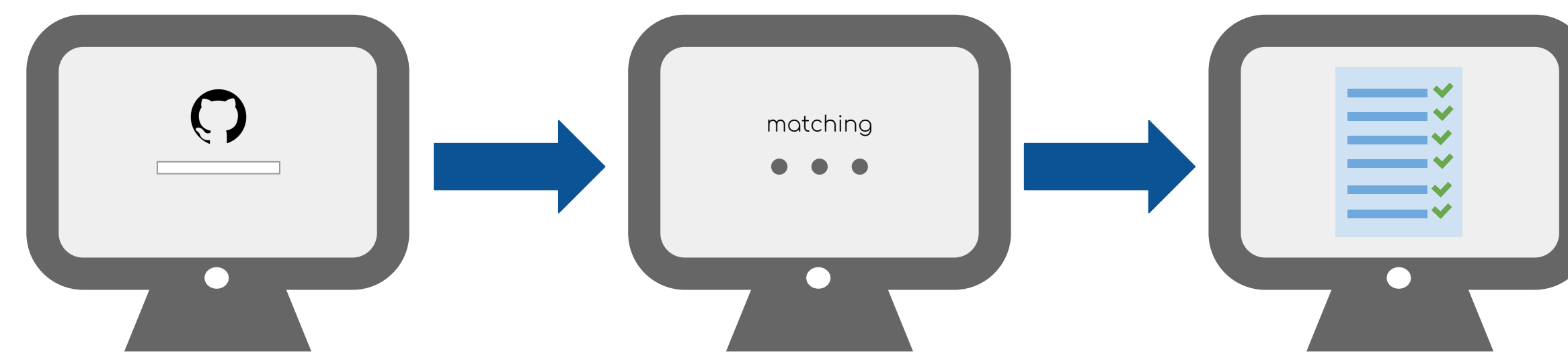
TECHNOLOGIES

Python (Language): Python has many 3rd party libraries that support GitHub mining and JSON parsing. After reviewing our options, we decided that Python offered what we needed and would allow us to get more done in less time in comparison to other languages.

PostgreSQL (Database): PostgreSQL supports both SQL for relational and JSON for non-relational queries[2]. After reviewing our options, PostgreSQL worked best for the type of information we would be storing such as user account information, project requirements, and user skills.

Digital Ocean (Web Server): This web hosting service is highly scalable and dependable, with an uptime average of 99.99%[3]. Our client already had an account with Digital Ocean, so there would be no worry handing off the final product.

Django (Web Framework): Django is a highly dependable and scalable Python based web framework. Django also officially supports our database, PostgreSQL This framework offered the best features that would allow us to work on both the front and back end. The added bonus was that it integrated well with other technologies we were using.

GitHub API V3: The GitHub API V3 is essential to our product, as it is how we determine user skills and acquire required skills from available OSS projects. The V3 API was selected in place of the of V4 as it it is much more user friendly and pairs well with our selected language, Python.

ARCHITECTURE
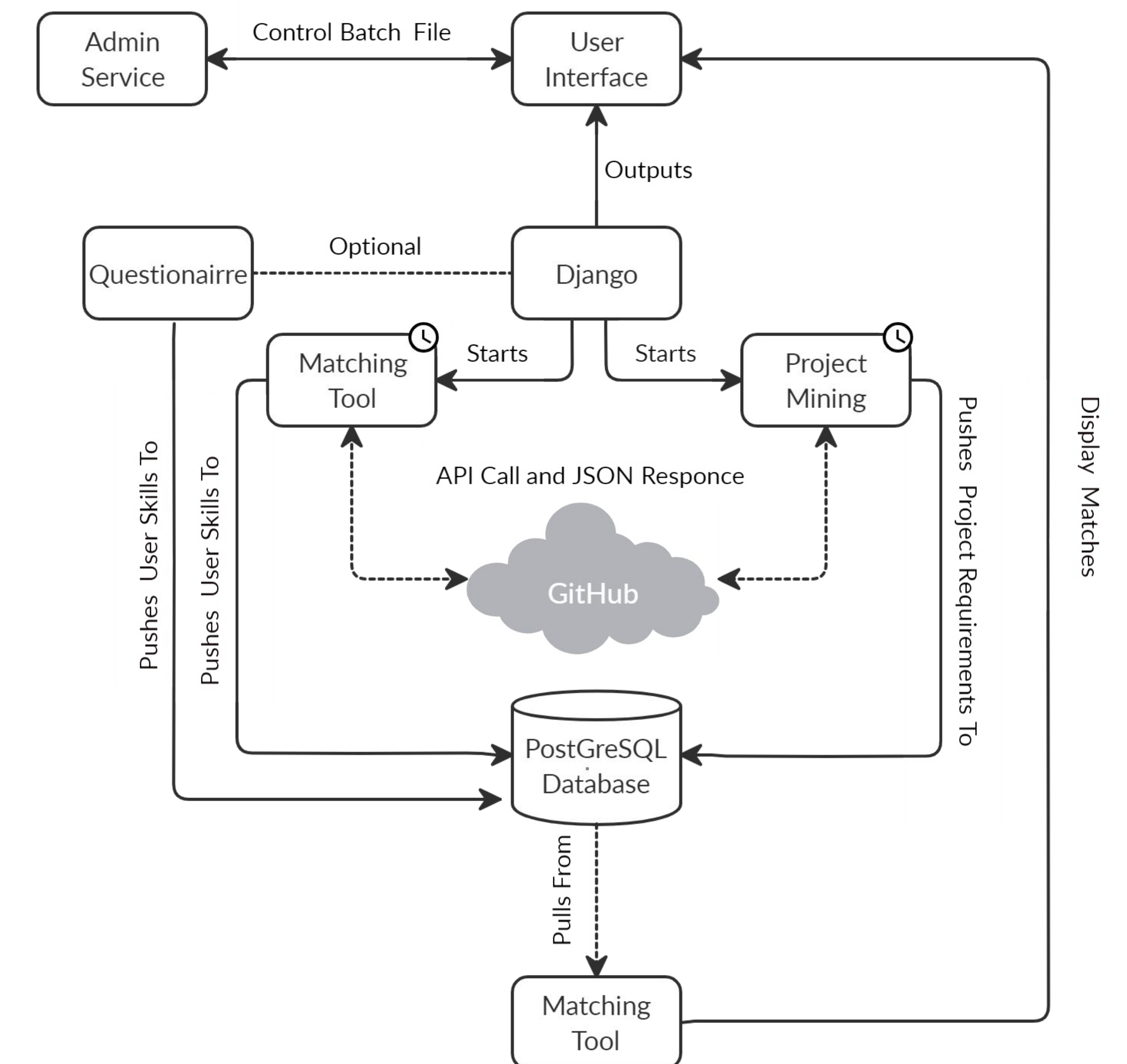
FUTURE WORK

- Due to time constraints, our client only asked us to match users to Java projects. In the future, this will expand to more languages and skills.
- Skills are determined from files users contribute to, not individual code snippets as that it out of our range. As our client continues this project, this may change.

OUTCOMES

This project will aid our client, Dr. Steinmacher in his efforts to support open source newcomers. By creating a starting point for these developers, the application will ease the information overload initially felt, hopefully leading to more contributions.

REFERENCES

[1] https://opensource.com/resources/what-open-source
[2] https://blog.panoply.io/postgresql-vs-mongodb.
[3] https://hostingfacts.com/hosting-reviews/digitalocean/