Mining the trend of the most popular software tools of the last decade

CSPB 4502
Data Mining Project

Group 3

Ning Chih Chang

Description

This is preliminary trend analysis of the most popular programming languages and development tools in the last decade (from 2015 to 2024) and AI tools in the last two years (2023 and 2024) on different types of software jobs and age groups.

Example questions that could be answered by this project:

- What is the trend of the most popular programming languages for full stack web developers in the last decade and how does it differ from the trend for students?
- What is the trend of the most popular AI tools for age 35+ developers in the past two years?

Prior work related to trends in programming language and development tools

Yaofei Chen, Rose Dios, Ali Mili, Lan Wu and Kefei Wang. 2005. An empirical study of programming language trends. *IEEE Software* 22, 3 (June 2005), 72-79. https://doi.org/10.1109/MS.2005.55

Vishal Johri and Srividya Bansal. 2018. Identifying trends in technologies and programming languages using topic modeling. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. 391-396. https://doi.org/10.1109/ICSC.2018.00078

Harshit Gujral, Abhinav Sharma and Parmeet Kaur. 2018. Empirical investigation of trends in NoSQL-based big-data solutions in the last decade. In *2018 Eleventh International Conference on Contemporary Computing (IC3)*. 1-3. https://doi.org/10.1109/IC3.2018.8530582

Tanya Uppal, Saumitya Srivastava and Kavita Saini. 2022. Web development framework : future trends. In *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. 2181-2184. https://doi.org/10.1109/ICAC3N56670.2022.10074105

Datasets

- Stackoverflow Developer Survey 2015 – 2024 (10 sets of surveys in total)
- Downloaded from https://survey.stackoverflow.co/
- The 10 surveys are saved in Ning Chih's computer

Proposed work

- Data cleaning: handle missing values

- Data preprocessing: handle typos and other special characters

- Data integration: match the shared attributes and remove irrelevant attributes among 10 different surveys

- Data transformation: attribute construction for software jobs

Tools will be used:

Python, Pandas, Matplotlib, Scikit-learn...

Evaluation

- Use data visualization to provide alternative views of the dataset which can be used to evaluate the results

- The results of this study is preliminary, and therefore a further study with rigorous data sample methods is suggested