Can a Chatbot Support Software Engineers with Load Testing?

Approach and Experiences

Presentation at ACM/SPEC ICPE 2020, Apr. 2020

Samuel Beck
Lasse Merz
Christoph Zorn
Fabian Beck
Leonel Merino
Dušan Okanović
André van Hoorn

@samu_beck
@beck_fabian
@leonel_merino
@okanovic_d
@andrevanhoorn
Motivation

ChatBot

Performance Evaluation

System
Can a Chatbot Support Software Engineers with Load Testing?

**Goals**

- Evaluate the applicability of a chatbot to facilitate performance evaluations.
- Gather impressions from performance experts and novices about using a chatbot in this context.

I want to test Eve’s system.

I want to perform a load test for Eve’s system.

Perform a load test with 100 users and measure response time.
Can a Chatbot Support Software Engineers with Load Testing?

Approach

User Interaction

Language Processing

Analysis & Reporting

PerformoBot

Dialogflow

Eve

Mia

Colin
Great, let us set up a load test together. I need to know what load you want to use, the domain to be tested and the metric which should be measured.

Alright I created a load test for you with the following parameters: load: 100 users, domain: www.example.com, metric: response time. If you are satisfied I can execute the load test. Otherwise, you can change the parameters, add more parameters from the parameter list or reset and start over.
Hello! I am PerformoBot. I am here to help you with your performance evaluation. If you have any questions feel free to ask about anything you don't understand. Please state your performance evaluation concern.

@PerformoBot What can you do?

I can help you to evaluate any system you want. Currently, I can create, execute, and show the results of load tests. Feel free to set up a load test regarding your concern or ask any questions.

Then let's perform a load test.

Great, let us set up a load test together. You need to provide the load you want to use, the domain which should be tested and the metric to be measured.

What is a load?

The load defines the amount of work a load test is processing during its execution. The load is usually the number of (virtual) users that produce work on a web service (e.g 100 users).

Please use 500 users.

Please enter the domain (e.g. www.google.com).
Can a Chatbot Support Software Engineers with Load Testing?

Otherwise, you can change the parameters, add more parameters from the parameter list or reset and start over.

Eve 21:14 Uhr
@Performobot execute

Performobot APP 21:14 Uhr
I executed your load test. Well done, you successfully completed the task! I prepared a report about this problem with a very similar query. I examined 500 users at https://registration.uni-stuttgart.edu/login and measured the server status. After clicking the "I'm done" button you will get a link to this report. Take a look and fill out the survey afterwards. Thank you, it was a pleasure!

Eve 21:14 Uhr
@Performobot result

Performobot APP 21:14 Uhr
I created a report for your query. Click on the button below to see the details.
Summary: The server was online for 4 minutes and 45 seconds. After 4 minutes and 45 seconds, the server crashed and stayed offline for the rest of the experiment. There occurred 1 incident that led to the crash of the system

Here is the report. Click on the button for details. (57 kB)
Can a Chatbot Support Software Engineers with Load Testing?

**Performance Report**

Perform a load test with 500 users at https://registration.uni-stuttgart.edu/login and measure the server status.

**Server status**

- **From**: 08.04.2019 13:30 to 08.04.2019 14:34
- **Server status**: Online

**Analysis Summary**

- **Analysis Tool**: JMeter
- **Duration**: 4 m 59 s 322 ms
- **Collect samples**: 500
- **Metrics**: elapsed time, request status, server status, received traffic, latency, connection time, response time

**Experiment summary**

- JMeter was used as the load testing tool. During the execution of the experiment (4 m 59 s 322 ms), 500 virtual users produced a load of the tested domain (https://registration.uni-stuttgart.edu/login).

**Latency**

- **Summary**: The overall minimum was 220 ms after 2 m 37 s 723 ms and the overall maximum was 2523 ms after 4 m 5 s 322 ms. The mean was 657 ms, 244 ms was recorded 91.2% of the time, 220 ms was recorded 81.8% of the time.

- **Info**: Latency is the amount of time a message takes to traverse a system. In a computer network, it is the expression of how much time it takes for a packet of data to get from one designated point to another. It is measured as the time required for a request to be sent to the server and returned to its sender. Latency depends on the speed of the transmission medium and the delays in the transmission by devices along the way. A low latency indicates a high network efficiency.
Can a Chatbot Support Software Engineers with Load Testing?

Study: Research Questions

01 How can a chatbot help users do a performance evaluation through a load test?

02 How do users interact with PerformoBot?

03 What educational effect does PerformoBot have on participants?

04 What are the differences between inexperienced users and experts in the above categories?

Task + Survey

47 participants
Study: Task + Survey

- Participant
- Performance Evaluation
- Test System
- chat
- 8:00 min
- PerformoBot
- evaluate
- Survey

Can a Chatbot Support Software Engineers with Load Testing?
**Study: Results**

- **47 participants**
- **Median of 10 interactions per participant**
- **Average of 3.64 minutes session time**
- **87.2% finished the task**
- **80.9% correctly answered the scenario**

**Automatic execution & report creation**: Approved

**Explanations**: Approved

**Complex load testing capabilities**: Not approved

**Too little guidance fallback mechanisms & missing keywords**: Not approved

**No educational effect**: Not approved
Can a Chatbot Support Software Engineers with Load Testing?

Study: Inexperienced Users vs. Experts

Helpfulness of PerformoBot
- Novices and experts were equally able to successfully complete the task.
- Novices evaluated PerformoBot more positive and helpful.

Sentiment
- 22 novices and 25 experts
- Novices evaluated more positive and helpful.

Knowledge level
- Novice: 63.7%
- Expert: 44.0%

Sentiment distribution:
- Positive: 48.0%
- Neutral: 18.2%
- Negative: 8.0%
Can a Chatbot Support Software Engineers with Load Testing?

Conclusion

- Helpful for novice/inexperienced users
- Needs advanced functionality for experts

Artifacts:

- More performance concerns
- Improved usability features

DOI: 10.5281/zenodo.3662711
Icons from flaticon (https://www.flaticon.com/) made by dave-gandy (question mark), monkik (people), freepik, eucalyp, smashicons (analysis, computers, clocks), flat-icons (performobot), roundicons (check mark, x, smileys), and from https://icons8.de.

Chatbot gif: https://cdn.dribbble.com/users/37530/screenshots/2937858/drib_blink_bot.gif

Messages gif: https://i.gifer.com/origin/98/98447b873b927d46f752e9e0fc9c2910_w200.gif