

cobot enhancement

Student Info

- Name : Meet Mangukiya
- GitHub Username: [@meetmangukiya](#)
- Alternative-/Nickname: Freenode IRC: mngkya
- Email: meet123mangukiya@gmail.com
- Which country will you reside in during this project?
 - India
- Which city, region or geographical boundary?
 - Mumbai, Maharashtra, India
- Time Zone: GMT +5:30
- GSoC blog RSS feed URL: <https://meetmangukiya.github.io/index.xml>

Code Sample

coala	Cool: https://github.com/coala/coala/pull/2895 https://github.com/coala/coala/pull/3610 - Yet to be merged All Merged: https://github.com/coala/coala/commits/master?author=meetmangukiya
coala-bears	Cool: https://github.com/coala/coala-bears/pull/877 https://github.com/coala/coala-bears/pull/1356 - Yet to be merged All Merged: https://github.com/coala/coala-bears/commits/master?author=meetmangukiya
cobot	https://gitlab.com/coala/cobot/commits/master
err-you-cobot	https://github.com/meetmangukiya/err-you-cobot/commits/master?author=meetmangukiya
Open PRs	https://github.com/pulls?utf8=%E2%9C%93&q=is%3Aopen+is%3Apr+author%3Ameetmangukiya+user%3Acoala
Other than coala	I have a patch at go-gitter https://github.com/sromku/go-gitter/pull/13

Project Info

1. How is your project helping coala and its community?

coala community has grown substantially over last few months. To make it easier on maintainers to cope up and still be a welcoming community for new contributors. Currently, cobot is a hubot instance, scripts for which are written in Coffeescript. This project is about porting the existing functionality of cobot to a python bot framework errbot and adding topic modelling to cobot to auto respond with correct(probably) answer when someone asks a question.

Another reason for porting: coala is python inclined community, cobot will receive more contribution if the devs can write plugins for cobot in python.

Addition of an interesting feature: Topic-Modelling for auto-responding FAQs

As previously mentioned, coala has been growing substantially and there are new contributors frequently and same questions are asked frequently. This feature aims at using Natural Language Processing and matching the asked questions with the sections in docs and then respond with a relevant section. This saves time for the users as well as developers significantly. Also, since this a bot interaction one doesn't tire out the opposite person responding.

2. What is the final goal for this project? What would make it a total and perfect success?

- a. Port cobot from hubot framework to errbot framework.
- b. Make the cobot code more readable and maintainable.

- c. Encourage contributions to cobot.
- d. Topic-Modelling using existing sections in Newcomers Guide.
- e. cobot responding to a question with less false positives

3. Are you already engaged with the project's possible mentors and to you have any preference for a particular mentor?

Lasse Schuirmann(@sils)

4. What parts of coala do you have to work with in order to complete this project? What else are you planning on using?

This project will be carried out of coala, inside a different repository dedicated to host the code of cobot.

5. Why are you the right person to work on this project?

I've worked on the existing cobot[hubot] and also started work on new cobot[errbot] a few months ago which will prove helpful in this project.

6. How do you plan to achieve completion of your project?

Abstract:

1. Port the existing scripts of cobot to errbot plugins.

All the commands that are available which can be obtained by `cobot help`

2. Ability to search API docs using cobot.

3. Ability to search User docs using cobot.

4. Ability to test cobot

Current cobot i.e. the hubot code is untested and there are many instances that it is deployed and it doesn't work, because it wasn't caught earlier due to lack of

tests.

Create an easier testing model for the errbot port and test the ported scripts.

5. Ability to list existing bears (using webservices)
6. Ability to show bear stats (using webservices)
 - a. No. of Bears
 - b. No. of Languages coala supports
7. Ability to show contribution stats (from webservices)
8. Ability to run bears on given code and return the results and diffs (from webservices)
9. cobot should work on private chats as well

This might require it to join in chat with the specified people. One workaround for now could be to join chat with all the people in the given org.

10. Easy deployment of cobot

Deployment is one of the tedious parts(most of the time). It should be easier to deploy cobot.

11. Ability to run hubot scripts in errbot (Stretch Goal)

Implementing this will allow errbot to use all the existing hubot scripts that already exists.

12. Topic modelling to answer user questions that have already been answered or point out to a relevant source on triggering a question (i.e. If ? is in the message).

How is this different from point 2 and 3?

- In point 2 and 3, a user issues a command with the exact search string that will be used for searching the docs

What are the benefits of this?

- cobot will be auto-responding to the questions whose answer it already knows (data is being fed beforehand).
- Questions don't have to be exactly same. cobot will be doing NLP and hence still return an answer for a differently put question. Which will really help the maintainers and even developers to get quicker answers.

What are the downsides?

- Since NLP is involved and also since there will be programming terms in the sentences as well, there will be false positives.

Why should we use this then?

- Sure, there will be false positives. But there will be times when the answer suggested by cobot will actually be correct. If this feature was not there, then the user anyway had to wait for an answer and if the answer

is incorrect then someone will probably point it out as well. So it is anyway a benefit.

Timeline

Community Period

Days	Milestones	Timeline Check
May 5 - May 18	<ul style="list-style-type: none">- Formulate a cEP, get reviews, and get it merged.- Get familiar with topic-modelling with gensim. And also weigh other possible alternatives.	A merged cEP
May 19 - May 22	If some change has to be introduced in the proposed timeline (since I'll have more insight after the cEP is merged) then discuss it with the mentors and do the appropriate changes.	A new timeline with incorporated changes.
May 5 - May 30	Mess around with gensim or the alternative that is being selected for this purpose.	

Coding Period

Week 1: May 30 - June 5	<ul style="list-style-type: none">- Port all the existing functionality of cobot-hubot to cobot-errbot- Implement coding standard by using coafire and enforcing it in CI.
Week 2: June 6 - June 12 Week 3: June 13 - June 19	Implement: <ul style="list-style-type: none">- Bear Stats<ul style="list-style-type: none">- No. of bears- No. of languages supported- List existing bears

	<ul style="list-style-type: none"> - Additionally, filter by languages - Search API docs - Search User docs
	Implement: <ul style="list-style-type: none"> - Running bears on a given piece of code - Get review from the community about the UX of the new features added and introduce changes if any are requested and seem reasonable to add.
Week 4: June 20 - June 26 Week 5: June 27 - July 3	<ul style="list-style-type: none"> - Start writing the test helper for the plugins - Use that test helper and test all the existing plugins
	<ul style="list-style-type: none"> - Make cobot available on private chats. - Make cobot easily deployable
Week 6: July 4 - July 10	<ul style="list-style-type: none"> - Start laying the foundations of the topic-modelling.
Week 7: July 11 - July 17	To be extended during community bonding period[1] <ul style="list-style-type: none"> - Create corpus from existing docs - Train over the created corpus - Perform LDA and retrieve probable answer.
Week 8: July 18 - July 24	
Week 9: July 25 - July 31	
Week 10: Aug 1 - Aug 7	
Week 11: Aug 8 - Aug 14	<ul style="list-style-type: none"> - Buffer - Rigorous Testing - Getting comments from the community about the UX
Week 12: Aug 15 - Aug 21	<ul style="list-style-type: none"> - Document new cobot features. - Document the Topic-Modelling thing. - Explain the topic-modelling thing in a post to help future contributors - Buffer period.

Mockups

1. cobot list bears
| Languages | Bears |

| Python | ...bear... |

| Javascript | ...bears ..|

2. cobot list bears python

a. PEP8Bear

b. PythonPackageInitBear

c. ...

3. cobot search api rebase

Here are rebase docs:

http://api.coala.io/en/latest/search.html?q=rebase&check_keywords=yes&area=default

4. cobot search userdocs coafile

Here are coafile docs:

http://docs.coala.io/en/latest/search.html?q=coafile&check_keywords=yes&area=default

5. cobot bear stats

There are x Bears supporting y Languages

6. cobot user stats meetmangukiya

Issues opened : x

Commits: y

Reviews: z

7. cobot check Bear1 Bear2 ...

code

Corrected:

...

8. cobot auto-responding to questions

>>> What is a rebase?

<<< @user please have a look at this

http://api.coala.io/en/latest/Developers/Git_Basics.html#rebasing

>>> How to do a rebase?

<<< @user please have a look at this

http://api.coala.io/en/latest/Developers/Git_Basics.html#rebasing

>>> Rebase?

<<< @user please have a look at this

http://api.coala.io/en/latest/Developers/Git_Basics.html#rebasing

Other Commitments

1. Do you have any other commitments during the GSoC period, May 8th to August 29th?

a. Do you have exams or classes that overlap with this period?

- i. I will have my exams in the month of May.
- ii. I will have my vacations in the month of June and I'll try to work more in this period, including weekends. I hope to get most of the stuff done in this month.
- iii. My college will probably start from second or third week of July.

b. How do I plan to dedicate the time GSoC demands?

- i. I'd like to start working on the project from the last two weeks of bonding period of GSoC if possible, since I've already been around for quite some time and know the community as well as coala.

Let's crunch some numbers:

- 1. There are 84 days between May 30 and Aug 21.
- 2. That is equivalent to 12 weeks.
- 3. One has to work 5 days a week, that means 60 days

So if I start from may 18th(suppose) then I get 12 days headstart.

If I work June(all days), that'll be 30 days.

Assuming college starts in second week of July, that'll be 12 days.

That adds up to 54 days, and still almost 6 weeks is left. So even If I work on weekends for next weeks, then it will give 12 days, an extra one week.

But all this was given considering that I work each and every day from May 18 and not working on weekdays from July 3rd week.

But that is obviously not true, I'll be working on weekdays as well and distribute the work little over the remaining period increasing productivity.

So, given the above numbers it is approximately safe to assume that the possibilities about me getting time to complete my project is pretty good. But I'd require 12 days of headstart in bonding period.

- ii. In the week of July, I can work on GSoC after returning from my college, and if I find that I'm not getting enough time, then I will

skip not-so-important classes and the semester would've just been started, so it'll take time to settle.
And also I'll be working extensively on the weekends(college holiday).

- c. Do you have plans to have any other job or internship during this period?
No.
- d. Do you have any other short term commitments during this period?
No.
2. Have you applied to any other organizations? If so, to whom and do you have a preferred project/org?
No.

Extra information

OpenHub Account	https://www.openhub.net/accounts/meetmangukiya
University Name	K.J.Somaiya College of Engineering
Major	I.T. Engineering
Current Year & expected graduation year	First Year Engineering. 2020.
Degree	B.Tech.
Alternate Contact	meetmangukiya@coala.io
Homepage	meetmangukiya.github.io
Twitter	@meetmangukiya98
Instant Messaging	Telegram - @meetmangukiya Messenger - @meetbmangukiya

Footnotes

[1] : @sils spontaneously came up with this idea and there was not much time left for more specific implementation details to be discussed and added.