Accurate Troubleshooting Recommendations Using Online Forum and Language Models

온라인 포럼과 언어 모델을 활용한 시스템 문제 해결 방법 추천

Youyang Kim (Advisor: Byungchul Tak)

Kyungpook National University (KNU), Daegu, Republic of Korea

Problems in Cloud Systems

- Various cause of software problems
  - Misconfiguration
    - Mistaken by human, Environment changes, …
  - Hardware failure
    - Cpu/memory/storage/switch failure, temporary network partitioning
  - Resource shortage
    - Insufficient memory space, storage space, cpu capacity, network saturation
  - Network failure
  - Bug in source code

Researches on System Problems

- Anomaly detection
  - Building anomaly propagation graph [1]
  - Applying bayesian linear attribution [2]
  - Time-series metrics of sensor, network fingerprint, system calls [3,4,5]
  - Analyzing flow and pattern of log [6,7]
  - Workflow monitoring via interleaved logs [8]

- Fault localization
  - Network monitoring to localize fault component for OpenStack [9]
  - Combining multiple dimensions value to analyze abnormal KPI [10]
  - Building dependency graph using system calls [11]

Definition of troubleshooting

Most of the researches focus on detecting anomalies & fault localization.

Still, not enough work on finding root cause or direct solution of problem

- [1] Automated anomaly detection and root cause analysis in virtualized cloud infrastructures (NOMS'16)
- [2] BALANCE: Bayesian Linear Attribution for Root Cause Localization. (arxiv'23)
- [3] Root cause detection in a service-oriented architecture. (SIGMETRICS'13)
- [4] Sieve: Actionable insights from monitored metrics in distributed systems (Middleware'17)
- [5] Root cause localization for unreproducible builds via causality analysis over system call tracing (ASE'19)
- [6] Deeplog: Anomaly detection and diagnosis from system logs through deep learning (CCS'17)
- [7] Execution anomaly detection in distributed systems through unstructured log analysis (ICDM'09)
- [8] CloudSeer: Workflow monitoring of cloud infrastructures via interleaved logs (ASPLOS'16)
- [9] HANSEL: Diagnosing faults in OpenStack [CoNEXT'15]
- [10] CMMD: Cross-Metric Multi-Dimensional root cause analysis [KDD'22]
- [11] Root cause localization for unreproducible builds via causality analysis over system call tracing (ASE'19)

Research Problem

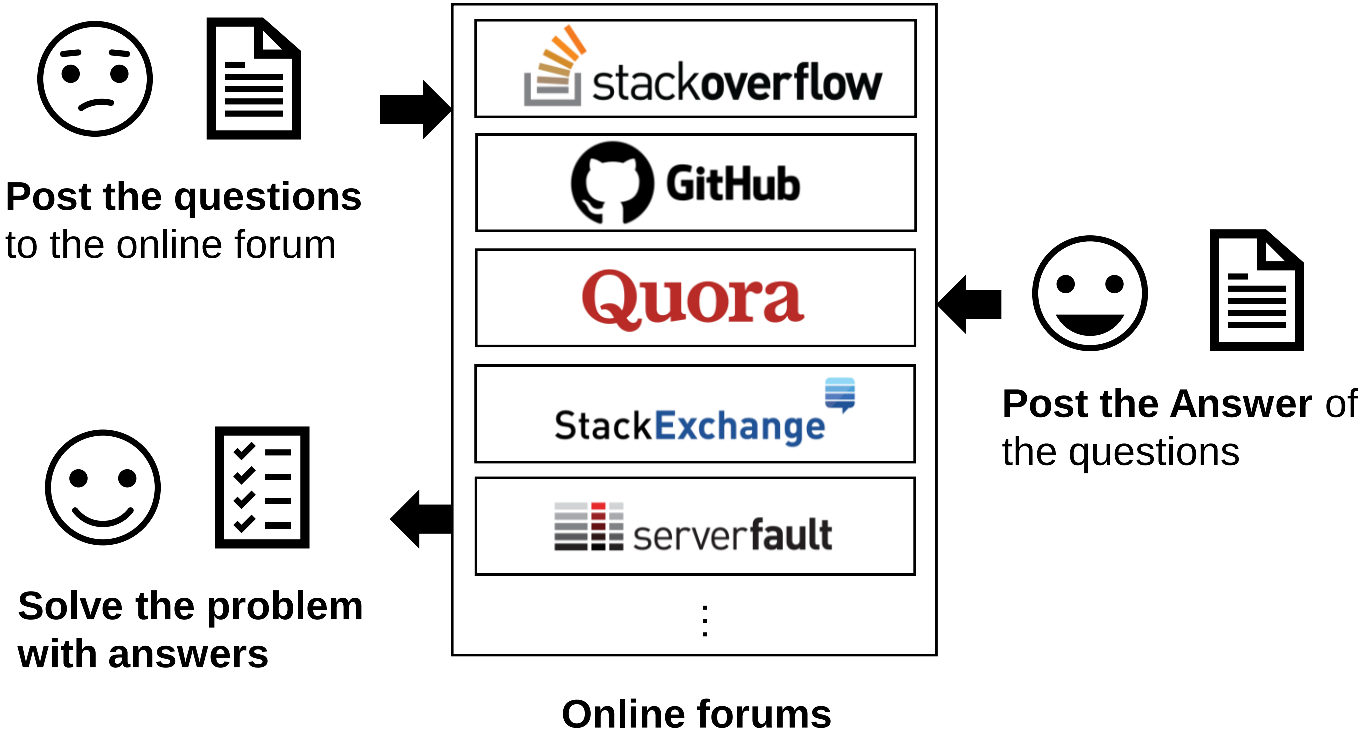
- Goal: Finding the ***root cause or direct solution*** to a given system problem

- Motivation:
  - Finding out the root cause or solution is too difficult and require high domain-knowledge
  - We often use sensor, log or system calls, but ***online forum has already vast amount of troubleshooting data.***
  - However, no suitable approaches exists to use forum data effectively.
  - Nowadays, new NLP and AI techniques help us to use online forum data effectively to system troubleshooting.

- Approach: *Using NLP and AI techniques, retrieving the most problem-relevant post which has root cause or solution in online forum*

What is Online Forum?

- An **online discussion site** where people can hold conversations in the **form of posted messages.**
  - Popular Online Forum: Stack Overflow, GitHub Issue, Quora, Serverfault, Stack Exchange, …



Post the questions to the online forum

Post the Answer of the questions

Solve the problem with answers

Online forums

A) Usage of online forum

stackoverflow

Total:

# 23,772,352 Questions

# 35+ million Answers

# 21+ million Users

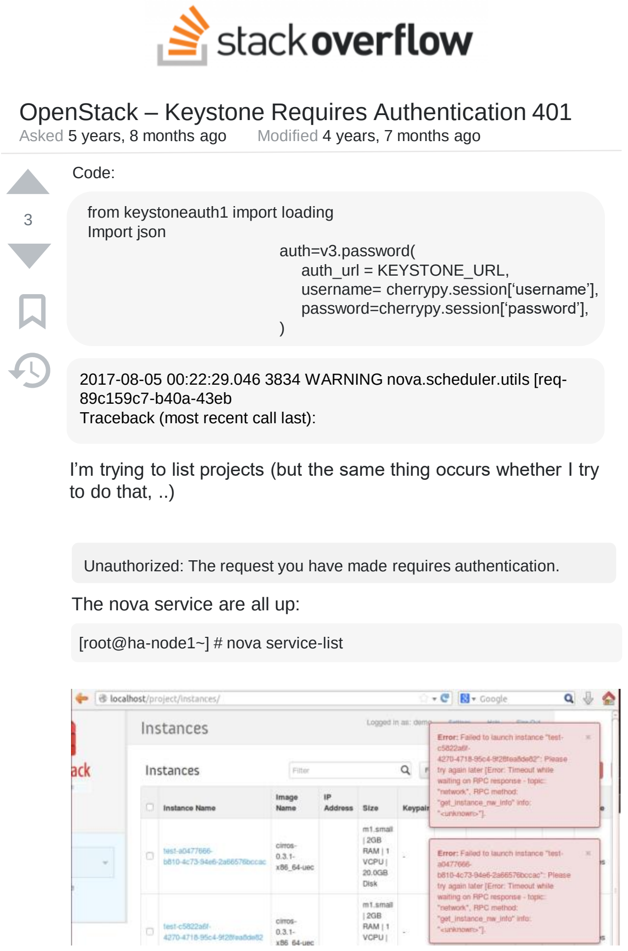
Per day:

# 7.6K new Questions

# 5.8M visitors

What Kinds of Questions are in Online Forum?

- Types of Information in Online Forum
  - ***Problems of system***
    - Sql, mongodb, docker, apache-spark,…
    - Q1. Openstack error when launching an instance
    - Q2. Spark.java.lang.OutOfMemoryError: java heap space
  - ***Coding problem***
    - Javascript, python, java, c++, …
    - Q1. How do I merge two dictionaries with one expression?
    - Q2. How can I remove a specific item from an array?
  - ***Conceptual question***
    - Q1. What and where are the stack and heap?
    - Q2. Is java ‘pass-by-value’ or ‘pass-by-reference’?



stackoverflow

OpenStack – Keystone Requires Authentication 401

Asked 5 years, 8 months ago Modified 4 years, 7 months ago

Code:

```
from keystoneauth1 import loading
Import json
                                    auth=v3.password(
                                        auth_url = KEYSTONE_URL,
                                        username= cherrypy.session['username'],
                                        password=cherrypy.session['password'],
                                    )
```

```
2017-08-05 00:22:29.046 3834 WARNING nova.scheduler.utils [req-
89c159c7-b40a-43eb
Traceback (most recent call last):
```

I’m trying to list projects (but the same thing occurs whether I try to do that, ..)

```
Unauthorized: The request you have made requires authentication.
```

The nova service are all up:

```
[root@ha-node1~] # nova service-list
```

Instances

	Instance Name	Image Name	IP Address	Size	Keypair
□	test-a0477666-b810-4c73-94e6-2a66576bccac	cirros-0.3.1-x86_64-uec		m1.small | 2GB RAM | 1 VCPU | 20.0GB Disk	-
□	test-c5822a6f-4270-4718-95c4-9f28fea8de82	cirros-0.3.1-x86_64-uec		m1.small | 2GB RAM | 1 VCPU |	-

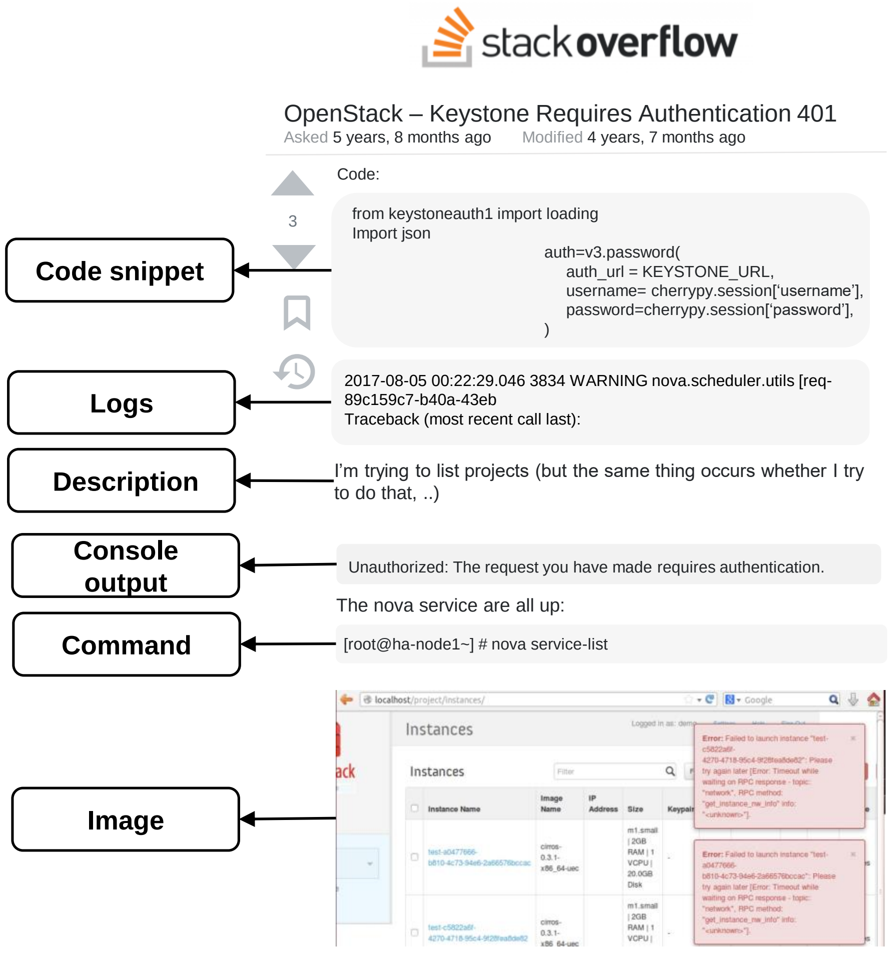
Error: Failed to launch instance "test-c5822a6f-4270-4718-95c4-9f28fea8de82": Please try again later [Error: Timeout while waiting on RPC response - topic: "network", RPC method: "get_instance_nw_info" info: "<unknown>"].

Error: Failed to launch instance "test-a0477666-b810-4c73-94e6-2a66576bccac": Please try again later [Error: Timeout while waiting on RPC response - topic: "network", RPC method: "get_instance_nw_info" info: "<unknown>"].

A) Example of stackoverflow questions about problem of system

What Kinds of Questions are in Online Forum?

- Types of Information in Online Forum
  - ***Problems of system***
    - Sql, mongodb, docker, apache-spark,…
    - Q1. Openstack error when launching an instance
    - Q2. Spark.java.lang.OutOfMemoryError: java heap space
  - ***Coding problem***
    - Javascript, python, java, c++, …
    - Q1. How do I merge two dictionaries with one expression?
    - Q2. How can I remove a specific item from an array?
  - ***Conceptual question***
    - Q1. What and where are the stack and heap?
    - Q2. Is java ‘pass-by-value’ or ‘pass-by-reference’?



OpenStack – Keystone Requires Authentication 401

Asked 5 years, 8 months ago Modified 4 years, 7 months ago

Code:

```
from keystoneauth1 import loading
Import json
                         auth=v3.password(
                            auth_url = KEYSTONE_URL,
                            username= cherrypy.session['username'],
                            password=cherrypy.session['password'],
                         )
```

```
2017-08-05 00:22:29.046 3834 WARNING nova.scheduler.utils [req-89c159c7-b40a-43eb
Traceback (most recent call last):
```

I'm trying to list projects (but the same thing occurs whether I try to do that, ..)

```
Unauthorized: The request you have made requires authentication.
```

The nova service are all up:

```
[root@ha-node1~] # nova service-list
```

localhost/project/instances/

Instances

Logged in as: demo

Instances

	Instance Name	Image Name	IP Address	Size	Keypair
□	test-a0477666-b810-4c73-94e6-2a66576bccac	cirros-0.3.1-x86_64-uec		m1.small | 2GB RAM | 1 VCPU | 20.0GB Disk	-
□	test-c5822a6f-4270-4718-95c4-9f28fea8de82	cirros-0.3.1-x86_64-uec		m1.small | 2GB RAM | 1 VCPU |	-

Error: Failed to launch instance "test-c5822a6f-4270-4718-95c4-9f28fea8de82": Please try again later [Error: Timeout while waiting on RPC response - topic: "network", RPC method: "get_instance_nw_info" info: "<unknown>"].

Error: Failed to launch instance "test-a0477666-b810-4c73-94e6-2a66576bccac": Please try again later [Error: Timeout while waiting on RPC response - topic: "network", RPC method: "get_instance_nw_info" info: "<unknown>"].

A) Example of stackoverflow questions about problem of system

Researches on Online Forum

- Various kinds of research on online forum
  - Understanding characteristics of online forum [1, 2]
  - Analyzing coding aspects [3, 4]
  - Reformulating query [5]
  - What developers ask [6]

- ***Not many researches of online forum for troubleshooting***

- [1] Tagging and linking web forum posts (ACL'10)
- [2] Crowdsourced knowledge on stack overflow: A systematic mapping study. (EASE'17)
- [3] Sotorrent: Reconstructing and analyzing the evolution of stack overflow posts (MSR'18)
- [4] From query to usable code: an analysis of stack overflow code snippets (MSR'16)
- [5] Automated query reformulation for efficient search based on query logs from stack overflow (ICSE'21)
- [6] Going big: A large-scale study on what big data developers ask (FSE'19)

What We Do with Online Forum?

- Goal: Retrieving the most relevant online forum post related to system problem

What We Do with Online Forum?

- Goal: Retrieving the most relevant online forum post related to system problem

Failure-case data
(user-side data)

Command

```
yyk@DGX-Station~$: openstack image create –file cirros “cirros”
```

Code

```
yyk@DGX-Station~$  python3 spark.py

spark = SparkSession.builder.getOrCreate()

data = [("a1", 2), ("b1", 4)]
df = spark.createDataFrame(data, ["fruit_1", "quantity_1"])

condition = df.select("fruit_1"), when(col("quantity_1") > 2, col("fruit_1") == "a1")

df = df.withColumn("new_column_1", when(condition == "many", "Lots of
").otherwise("Not many ") + col("fruit_1"))
```

What We Do with Online Forum?

- Goal: Retrieving the most relevant online forum post related to system problem

<u>Failure-case data</u>
<u>(user-side data)</u>

Command

```
yyk@DGX-Station~$: openstack image create –file cirros “cirros”
```

Code

```
yyk@DGX-Station~$  python3 spark.py

spark = SparkSession.builder.getOrCreate()

data = [("a1", 2), ("b1", 4)]
df = spark.createDataFrame(data, ["fruit_1", "quantity_1"])

condition = df.select("fruit_1"), when(col("quantity_1") > 2, col("fruit_1") == "a1")

df = df.withColumn("new_column_1", when(condition == "many", "Lots of
").otherwise("Not many ") + col("fruit_1"))
```

Log

```
2017-08-05 00:22:29.246 WARNING
nova.scheduler.utils
[req-89c159c7-b40a-43eb-8]
Failed to compute_task_build_instances: No valid host was found. There are not
enough hosts available. Setting instance to ERROR state.
```

Console output

```
AttributeError:’DeleteResult’ object has no attribute ‘sort’
```

Description

I installed openstack on my centos vm and when I try to see the list of launched instances, I get this error.
This code creates a new virtual machine instance in the Openstack Cloud using the novaclient library.

What We Do with Online Forum?

- Goal: Retrieving the most relevant online forum post related to system problem

Failure-case data (user-side data)

Command

```
yyk@DGX-Station~$: openstack image create –file cirros “cirros”
```

Code

```
yyk@DGX-Station~$  python3 spark.py

spark = SparkSession.builder.getOrCreate()

data = [("a1", 2), ("b1", 4)]
df = spark.createDataFrame(data, ["fruit_1", "quantity_1"])

condition = df.select("fruit_1"), when(col("quantity_1") > 2, col("fruit_1") == "a1")

df = df.withColumn("new_column_1", when(condition == "many", "Lots of
").otherwise("Not many ") + col("fruit_1"))
```

Log

```
2017-08-05 00:22:29.246 WARNING
nova.scheduler.utils
[req-89c159c7-b40a-43eb-8]
Failed to compute_task_build_instances: No valid host was found. There are not
enough hosts available. Setting instance to ERROR state.
```

Console output

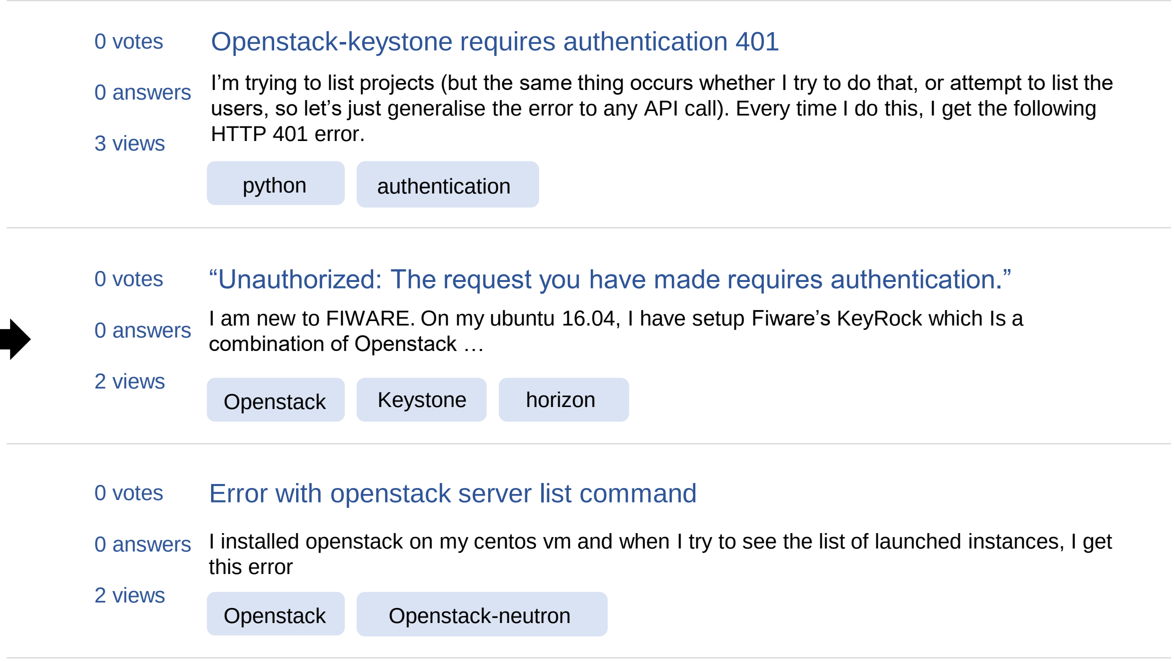
```
AttributeError:’DeleteResult’ object has no attribute ‘sort’
```

Description

I installed openstack on my centos vm and when I try to see the list of launched instances, I get this error.
This code creates a new virtual machine instance in the Openstack Cloud using the novaclient library.

Questions in forum (online forum side)

stackoverflow



0 votes
0 answers
3 views

Openstack-keystone requires authentication 401

I’m trying to list projects (but the same thing occurs whether I try to do that, or attempt to list the users, so let’s just generalise the error to any API call). Every time I do this, I get the following HTTP 401 error.

python | authentication

0 votes
0 answers
2 views

“Unauthorized: The request you have made requires authentication.”

I am new to FIWARE. On my ubuntu 16.04, I have setup Fiware’s KeyRock which Is a combination of Openstack …

Openstack | Keystone | horizon

0 votes
0 answers
2 views

Error with openstack server list command

I installed openstack on my centos vm and when I try to see the list of launched instances, I get this error

Openstack | Openstack-neutron

⋮

What We Do with Online Forum?

- Goal: Retrieving the most relevant online forum post related to system problem

Failure-case data (user-side data)

Command

```
yyk@DGX-Station~$: openstack image create –file cirros “cirros”
```

Code

```
yyk@DGX-Station~$  python3 spark.py

spark = SparkSession.builder.getOrCreate()

data = [("a1", 2), ("b1", 4)]
df = spark.createDataFrame(data, ["fruit_1", "quantity_1"])

condition = df.select("fruit_1"), when(col("quantity_1") > 2, col("fruit_1") == "a1")

df = df.withColumn("new_column_1", when(condition == "many", "Lots of
").otherwise("Not many ") + col("fruit_1"))
```

Log

```
2017-08-05 00:22:29.246 WARNING
nova.scheduler.utils
[req-89c159c7-b40a-43eb-8]
Failed to compute_task_build_instances: No valid host was found. There are not
enough hosts available. Setting instance to ERROR state.
```

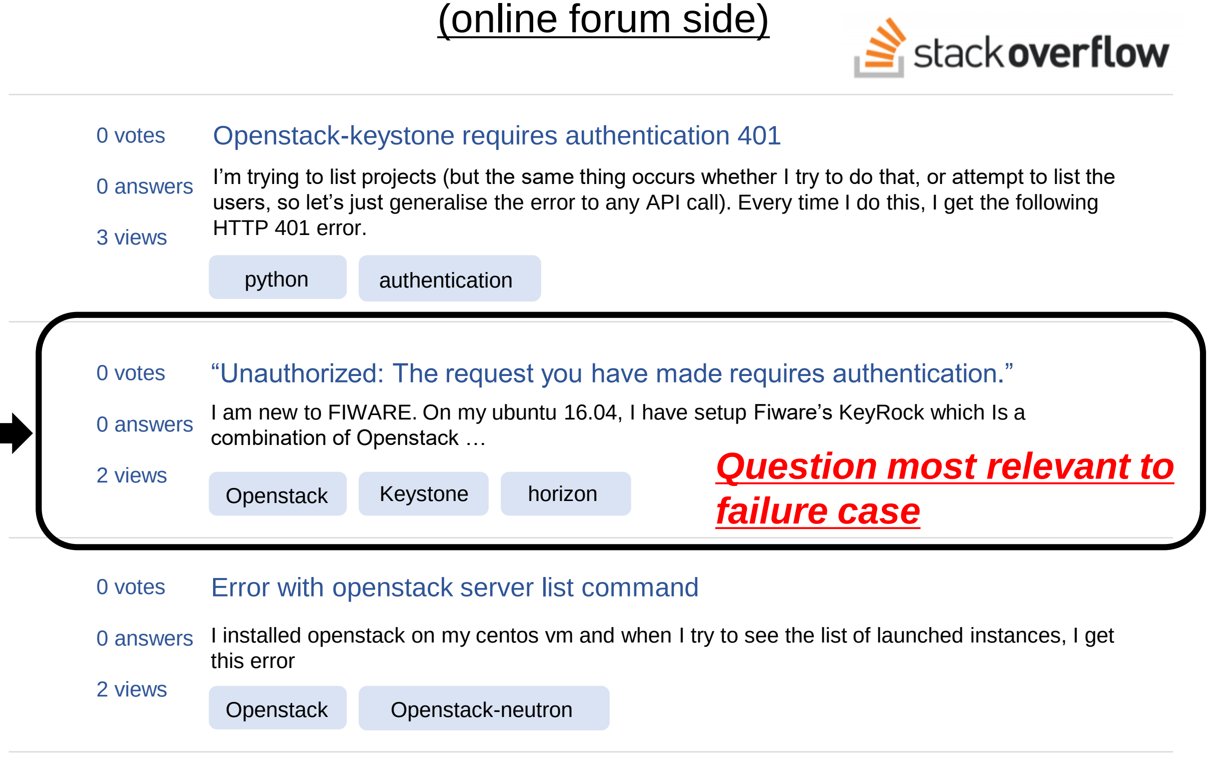
Console output

```
AttributeError:’DeleteResult’ object has no attribute ‘sort’
```

Description

I installed openstack on my centos vm and when I try to see the list of launched instances, I get this error.
This code creates a new virtual machine instance in the Openstack Cloud using the novaclient library.

Questions in forum (online forum side)



0 votes
0 answers
3 views

Openstack-keystone requires authentication 401

I’m trying to list projects (but the same thing occurs whether I try to do that, or attempt to list the users, so let’s just generalise the error to any API call). Every time I do this, I get the following HTTP 401 error.

0 votes
0 answers
2 views

“Unauthorized: The request you have made requires authentication.”

I am new to FIWARE. On my ubuntu 16.04, I have setup Fiware’s KeyRock which Is a combination of Openstack …

Question most relevant to failure case

0 votes
0 answers
2 views

Error with openstack server list command

I installed openstack on my centos vm and when I try to see the list of launched instances, I get this error

⋮

Naïve Search in Online Forum

- Naïve search 1. Searching using <u>*log or console output from failure case*</u>
  - Problem: Irrelevant document can have same logs, because one log appears in multiple operation

One log appears when multiple commands and codes are executed in OpenStack

Log	***Operations***
The request you have made requires authentication (HTTP 401)	Login to horizon dashboard
	Install Openstack, run ./stack.sh
	Make authentication with keystone
	Create domains
	List projects or domains

- Naïve search 2. Searching <u>*using code and command string*</u>
  - Problem: Code and command can be changed because of version update.
  - Problem: Code does not share the same word between the same tasks

- Naïve search 3. Searching with <u>*user-written description*</u>
  - Problem: Search results vary depending on how detailed the user explaining the situation.

Naïve Search in Online Forum

- Naïve search 1. Searching using *log or console output from failure case*
    - Problem: Irrelevant document can have same logs, because one log appears in multiple operation

One log appears when multiple commands and codes are executed in OpenStack

Log	***Operations***
The request you have made requires authentication (HTTP 401)	Login to horizon dashboard
	Install Openstack, run ./stack.sh
	Make authentication with keystone
	Create domains
	List projects or domains

Operation is also important when searching problem-related documents

- Naïve search 2. Searching *using code and command string*
    - Problem: Code and command can be changed because of version update.
    - Problem: Code does not share the same word between the same tasks

- Naïve search 3. Searching with *user-written description*
    - Problem: Search results vary depending on how detailed the user explaining the situation.

Does 3 Types of Naïve Search Works?

- *Naïve search does not find relevant document.*

Naïve search 1. Logs:

```
File “/usr/lib/python3/site-packages/keystoneauth1/session.py, line 484 in request
raise exceptions.from_response(resp, method, url)
…
Unauthorized: The request you have made requires authentication (HTTP 401)
```

Naïve search 2. Code:

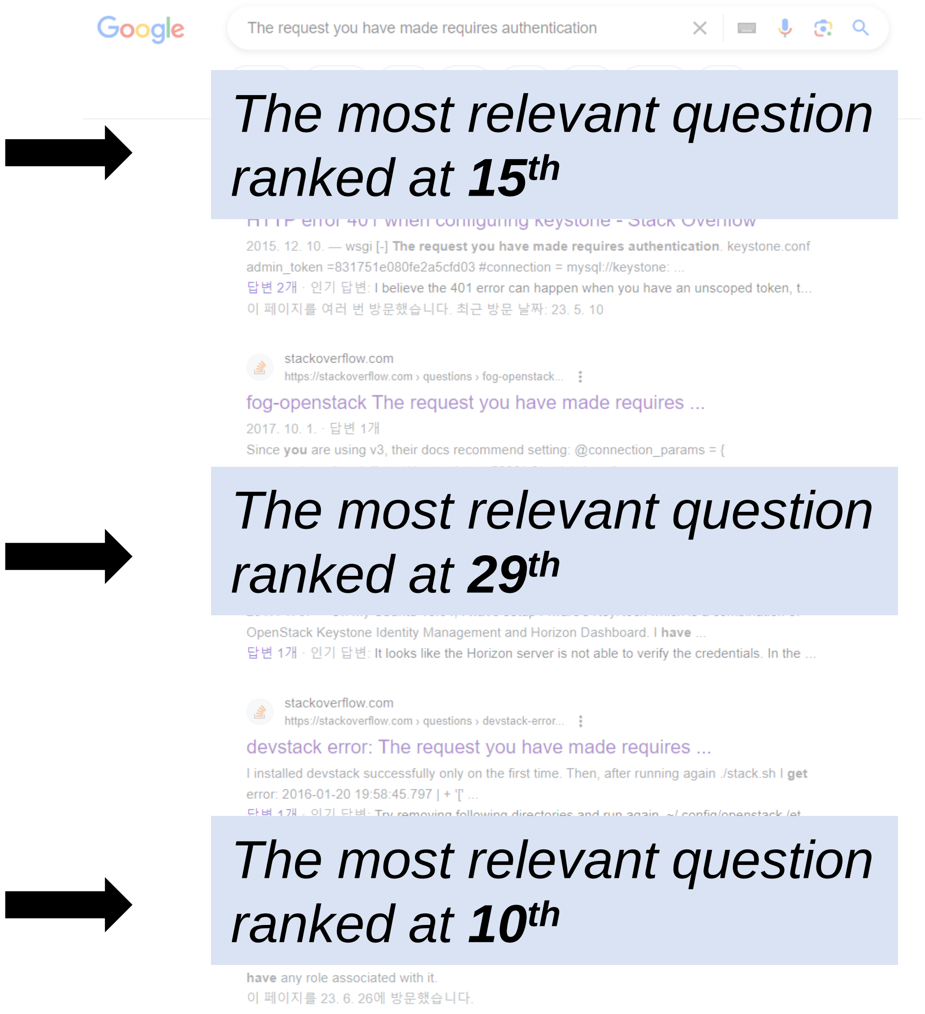
```
import openstack

os_connect = openstack.connect(
                    auth_url = AUTH_URL,
                    project_name = PROJECT_NAME,
                    username = USERNAME,
                    password = PASSWORD,
                    region_name = REGION_NAME,
                    user_domain_name = USER_DOMAIN_NAME)

for project in os_connect.list_projects():
                    print(“project list:”, project[‘name’])
```

Naïve search 3. User-written description

I tried to authenticate and list project in openstack.



The request you have made requires authentication

*The most relevant question ranked at **15**th*

HTTP error 401 when configuring keystone - Stack Overflow

2015. 12. 10. — wsgi [-] **The request you have made requires authentication**. keystone.conf admin_token =831751e080fe2a5cfd03 #connection = mysql://keystone: ...

답변 2개 · 인기 답변: I believe the 401 error can happen when you have an unscoped token, t...

이 페이지를 여러 번 방문했습니다. 최근 방문 날짜: 23. 5. 10

stackoverflow.com
https://stackoverflow.com › questions › fog-openstack...

fog-openstack The request you have made requires ...

2017. 10. 1. · 답변 1개

Since **you** are using v3, their docs recommend setting: @connection_params = {

*The most relevant question ranked at **29**th*

OpenStack Keystone Identity Management and Horizon Dashboard. I **have** ...

답변 1개 · 인기 답변: It looks like the Horizon server is not able to verify the credentials. In the ...

stackoverflow.com
https://stackoverflow.com › questions › devstack-error...

devstack error: The request you have made requires ...

I installed devstack successfully only on the first time. Then, after running again ./stack.sh I **get** error: 2016-01-20 19:58:45.797 | + '[' ...

답변 1개 · 인기 답변: Try removing following directories and run again. ~/.config/openstack /et

*The most relevant question ranked at **10**th*

have any role associated with it.

이 페이지를 23. 6. 26에 방문했습니다.

Why Naïve Search of Online Forum Does Not Work?

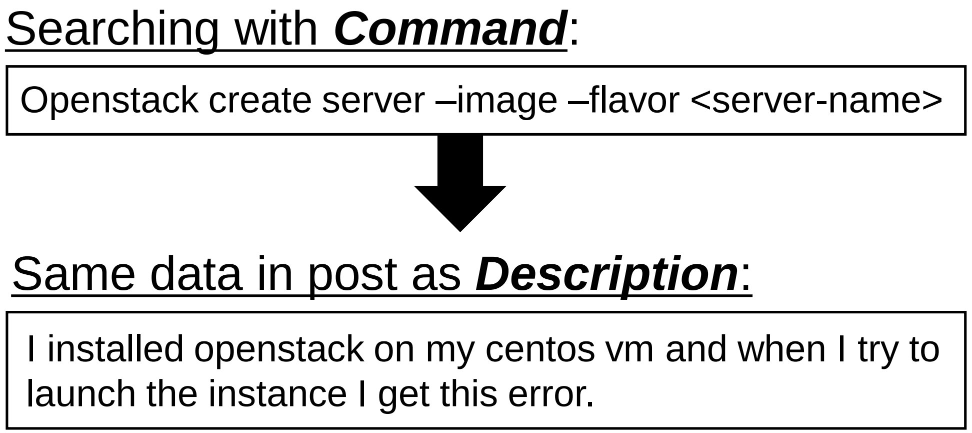
- Because of the assumption that ‘***the same data will exist in relevant forum posts***’.
  - When searching with command, user thinks that same command will exist in relevant post.
- But *relevant posts do not always contain the same data*, such as..
  - Users search with command, but same information is in the description.
  - Users search with description, but same information exists by code.

Searching with ***Command***:

```
Openstack create server –image –flavor <server-name>
```

Why Naïve Search of Online Forum Does Not Work?

- Because of the assumption that ‘***the same data will exist in relevant forum posts***’.
  - When searching with command, user thinks that same command will exist in relevant post.
- But *relevant posts do not always contain the same data*, such as..
  - Users search with command, but same information is in the description.
  - Users search with description, but same information exists by code.



Searching with ***Command***:

```
Openstack create server –image –flavor <server-name>
```

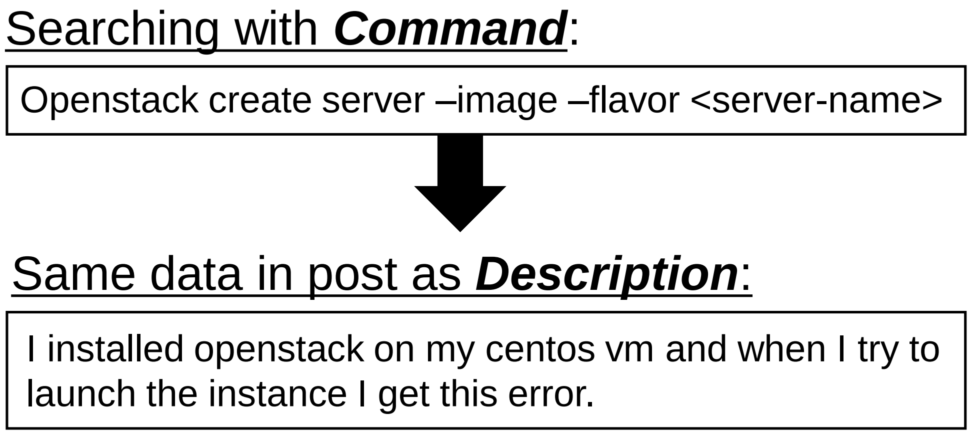
Same data in post as ***Description***:

I installed openstack on my centos vm and when I try to launch the instance I get this error.

a) Command and description share same information

Why Naïve Search of Online Forum Does Not Work?

- Because of the assumption that ‘***the same data will exist in relevant forum posts***’.
  - When searching with command, user thinks that same command will exist in relevant post.
- But *relevant posts do not always contain the same data*, such as..
  - Users search with command, but same information is in the description.
  - Users search with description, but same information exists by code.



Searching with ***Command***:

Openstack create server –image –flavor <server-name>

Same data in post as ***Description***:

I installed openstack on my centos vm and when I try to launch the instance I get this error.

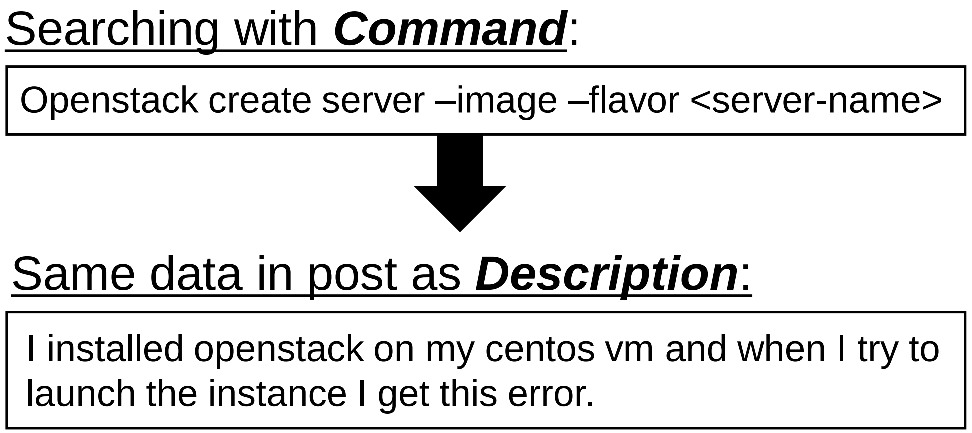
a) Command and description share same information

Searching with ***Description***:

Error with openstack server list command

Why Naïve Search of Online Forum Does Not Work?

- Because of the assumption that '***the same data will exist in relevant forum posts***'.
  - When searching with command, user thinks that same command will exist in relevant post.
- But *relevant posts do not always contain the same data*, such as..
  - Users search with command, but same information is in the description.
  - Users search with description, but same information exists by code.



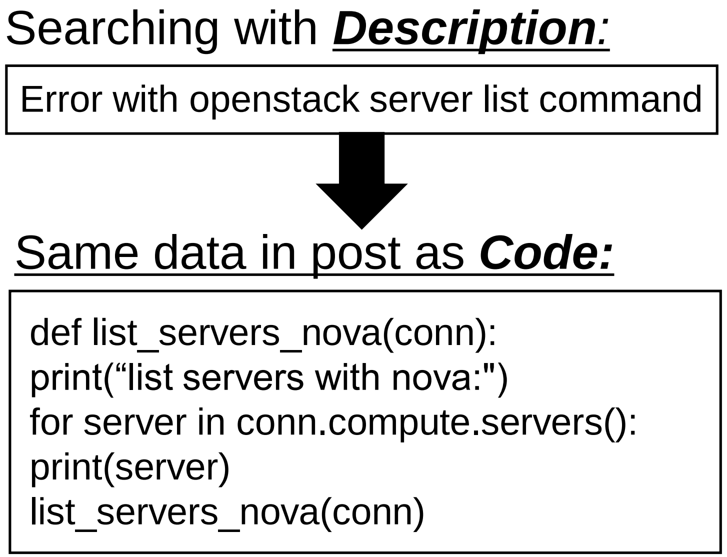
Searching with ***Command***:

Openstack create server –image –flavor <server-name>

Same data in post as ***Description***:

I installed openstack on my centos vm and when I try to launch the instance I get this error.

a) Command and description share same information



Searching with ***Description***:

Error with openstack server list command

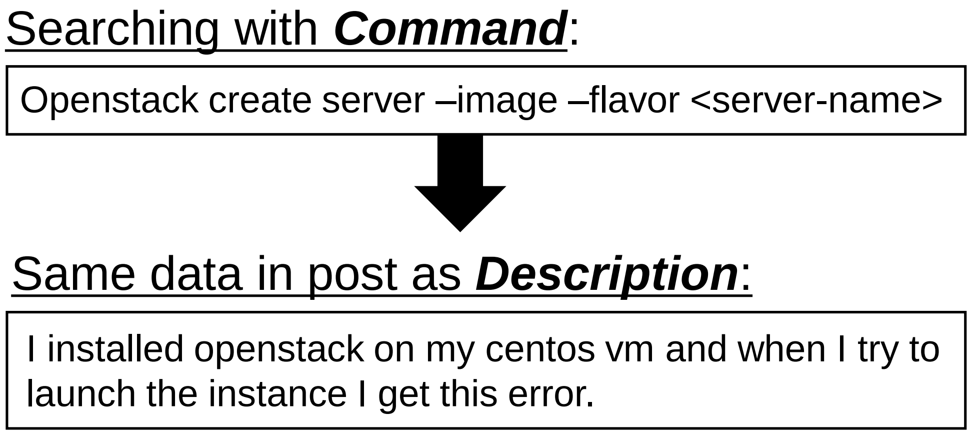
Same data in post as ***Code:***

```
def list_servers_nova(conn):
print(“list servers with nova:")
for server in conn.compute.servers():
print(server)
list_servers_nova(conn)
```

b) Description and code share same information

Why Naïve Search of Online Forum Does Not Work?

- Because of the assumption that '***the same data will exist in relevant forum posts***'.
  - When searching with command, user thinks that same command will exist in relevant post.
- But *relevant posts do not always contain the same data*, such as..
  - Users search with command, but same information is in the description.
  - Users search with description, but same information exists by code.



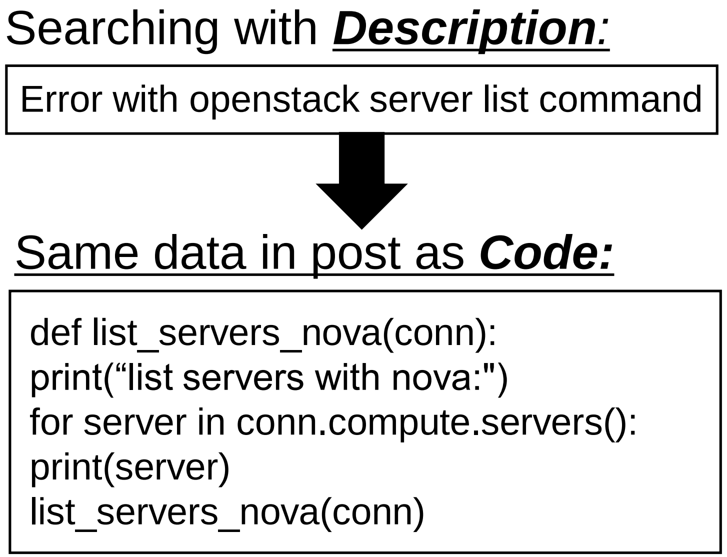
Searching with ***Command***:

Openstack create server –image –flavor <server-name>

Same data in post as ***Description***:

I installed openstack on my centos vm and when I try to launch the instance I get this error.

a) Command and description share same information



Searching with ***Description***:

Error with openstack server list command

Same data in post as ***Code:***

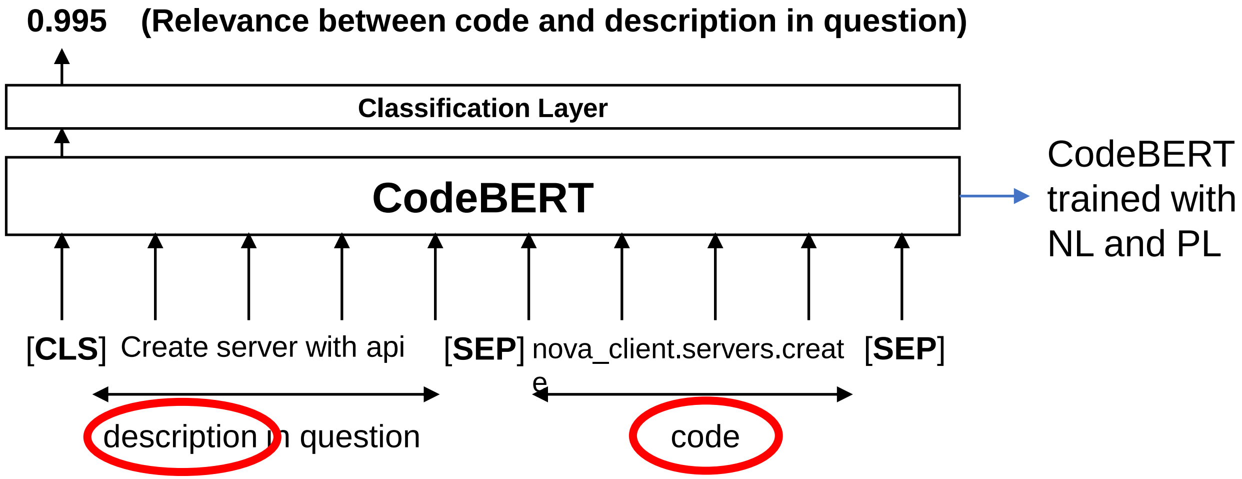
```
def list_servers_nova(conn):
print(“list servers with nova:")
for server in conn.compute.servers():
print(server)
list_servers_nova(conn)
```

b) Description and code share same information

Can't find relevant post!

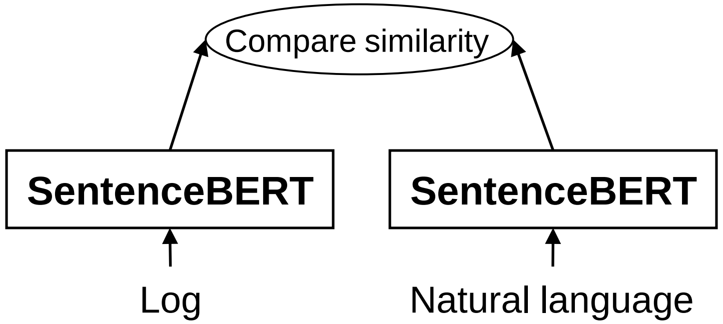
How Can We Solve This Problem?

- To resolve previous limitation, we need to compare between disparate data type.
  - comparison between command and description, or comparison between code and description
- Nowadays, bi-modal model such as **CodeBERT and SentenceBERT** enables the semantic search (enables to compute relevance) between disparate data types.
- We treat the online forum data *as combination of multiple disparate data*.
- We use this technique to **compare disparate types of data** for retrieving relevant online forum post



0.995
(Relevance between code and description in question)
Classification Layer
CodeBERT
CodeBERT trained with NL and PL
[CLS]
Create server with api
[SEP]
nova_client.servers.creat e
[SEP]
description in question
code

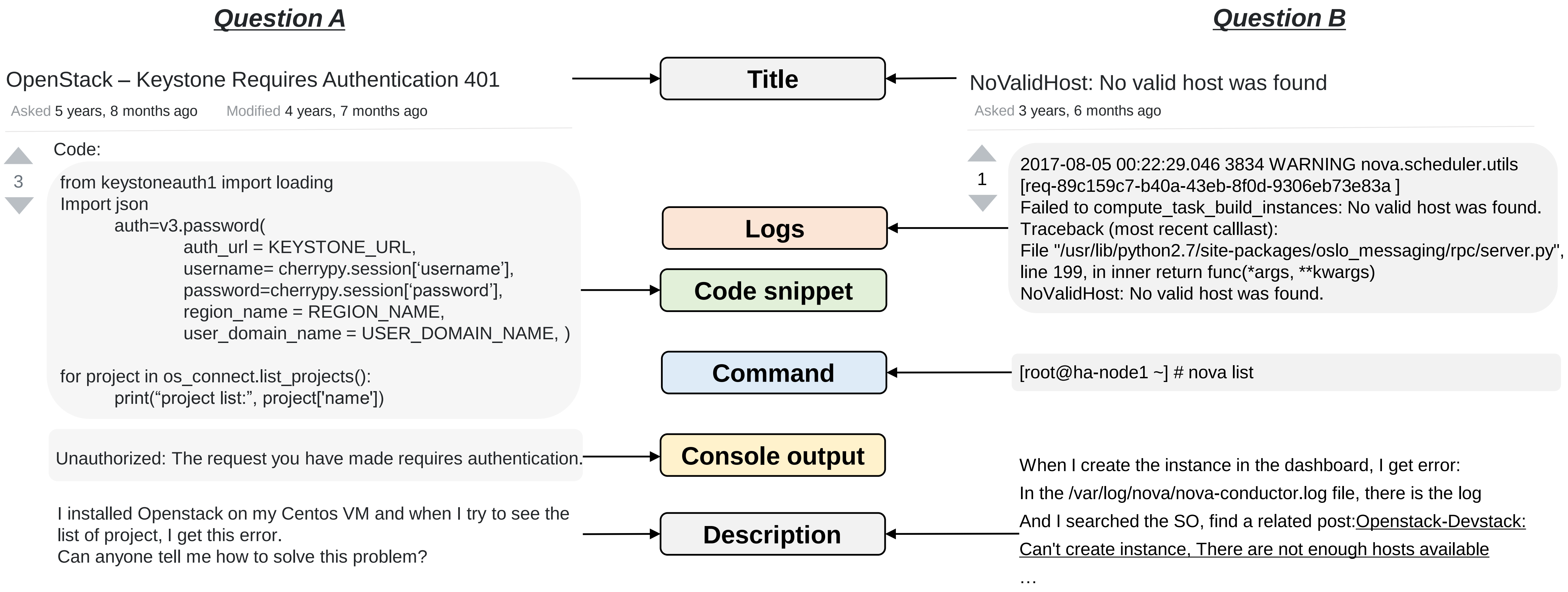
A) CodeBERT compute relevance between description and code



B) Comparing Similarity between log and NL using SentenceBERT

Multi-Modality of Online Forum Data

- Special aspect of online forum is **inclusion of multiple types of data**
  - Same type of data in failure case exist in forum posts.
  - machine languages (code), short phrases (command), natural language (log, console output, description)



Question A

OpenStack – Keystone Requires Authentication 401

Asked 5 years, 8 months ago Modified 4 years, 7 months ago

3

Code:

```
from keystoneauth1 import loading
Import json
        auth=v3.password(
                  auth_url = KEYSTONE_URL,
                  username= cherrypy.session[‘username’],
                  password=cherrypy.session[‘password’],
                  region_name = REGION_NAME,
                  user_domain_name = USER_DOMAIN_NAME, )

for project in os_connect.list_projects():
        print(“project list:”, project['name'])
```

```
Unauthorized: The request you have made requires authentication.
```

I installed Openstack on my Centos VM and when I try to see the list of project, I get this error.
Can anyone tell me how to solve this problem?

Title

Logs

Code snippet

Command

Console output

Description

Question B

NoValidHost: No valid host was found

Asked 3 years, 6 months ago

1

```
2017-08-05 00:22:29.046 3834 WARNING nova.scheduler.utils
[req-89c159c7-b40a-43eb-8f0d-9306eb73e83a ]
Failed to compute_task_build_instances: No valid host was found.
Traceback (most recent calllast):
File "/usr/lib/python2.7/site-packages/oslo_messaging/rpc/server.py",
line 199, in inner return func(*args, **kwargs)
NoValidHost: No valid host was found.
```

```
[root@ha-node1 ~] # nova list
```

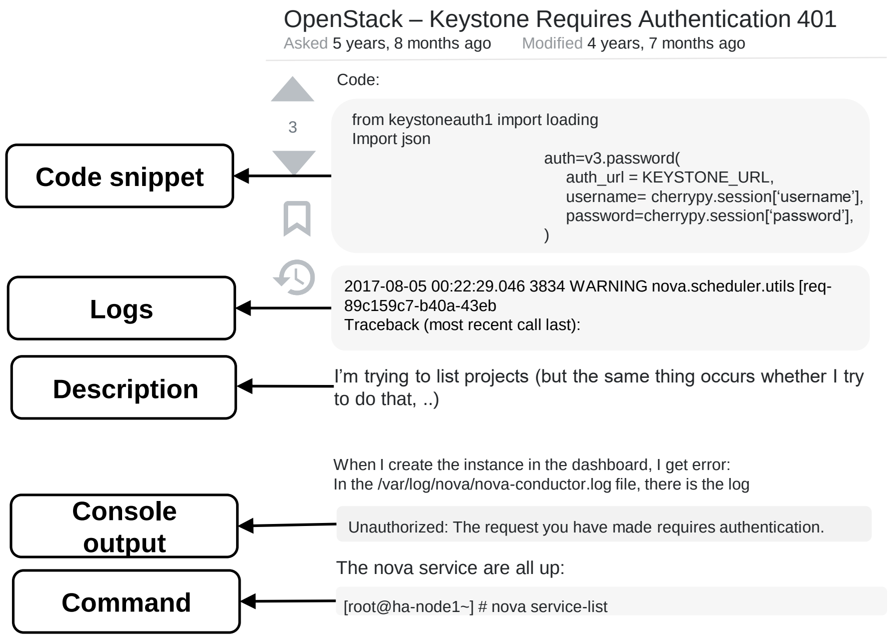
When I create the instance in the dashboard, I get error:
In the /var/log/nova/nova-conductor.log file, there is the log
And I searched the SO, find a related post:Openstack-Devstack: Can't create instance, There are not enough hosts available
…

How to Compare Failure-Case Data and Online Forum Posts?

- Approach: Separating forum posts and applying 3 types of model

stack overflow

Forum data
(online forum side)



OpenStack – Keystone Requires Authentication 401

Asked 5 years, 8 months ago Modified 4 years, 7 months ago

Code:

```
from keystoneauth1 import loading
Import json
                              auth=v3.password(
                                  auth_url = KEYSTONE_URL,
                                  username= cherrypy.session['username'],
                                  password=cherrypy.session['password'],
                              )
```

```
2017-08-05 00:22:29.046 3834 WARNING nova.scheduler.utils [req-
89c159c7-b40a-43eb
Traceback (most recent call last):
```

I'm trying to list projects (but the same thing occurs whether I try to do that, ..)

When I create the instance in the dashboard, I get error:
In the /var/log/nova/nova-conductor.log file, there is the log

```
Unauthorized: The request you have made requires authentication.
```

The nova service are all up:

```
[root@ha-node1~] # nova service-list
```

Code snippet

Logs

Description

Console output

Command

How to Compare Failure-Case Data and Online Forum Posts?

- Approach: Separating forum posts and applying 3 types of model

openstack

Failure-case Data
(user-side)

```
yyk@DGX-Station~$  python3 spark.py

spark = SparkSession.builder.getOrCreate()

data = [("a1", 2), ("b1", 4)]
df = spark.createDataFrame(data, ["fruit_1", "quantity_1"])

condition = df.select("fruit_1"), when(col("quantity_1") > 2, col("fruit_1") == "a1")

df = df.withColumn("new_column_1", when(condition == "many", "Lots of
").otherwise("Not many ") + col("fruit_1"))
```

Code snippet

```
2017-08-05 00:22:29.246 WARNING
nova.scheduler.utils
[req-89c159c7-b40a-43eb-8]
Failed to compute_task_build_instances: No valid host was found. There are not
enough hosts available. Setting instance to ERROR state.
```

Logs

I installed openstack and when I try to see list of launched instances, I get this error. This code creates a new virtual machine instance in the Openstack Cloud using the novaclient library.

Description

```
AttributeError:’DeleteResult’ object has no attribute ‘sort’
```

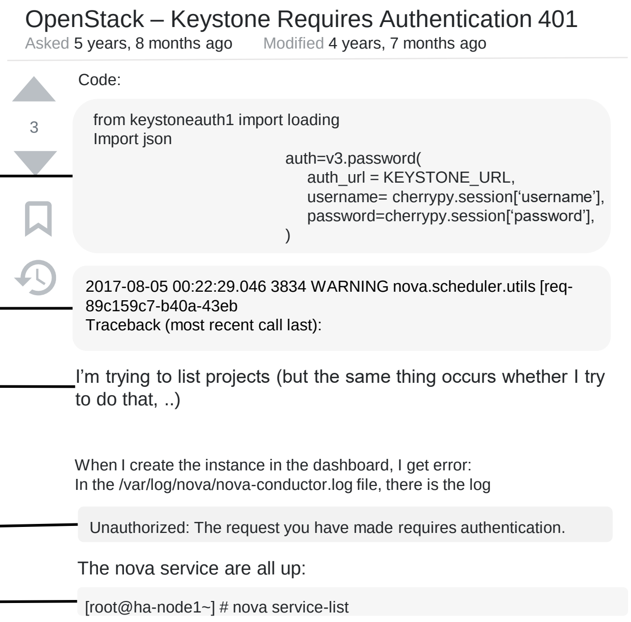
Console output

```
yyk@DGX-Station~$: openstack image create –file cirros “cirros”
```

Command

stackoverflow

Forum data
(online forum side)



OpenStack – Keystone Requires Authentication 401

Asked 5 years, 8 months ago   Modified 4 years, 7 months ago

Code:

```
from keystoneauth1 import loading
Import json
                              auth=v3.password(
                                  auth_url = KEYSTONE_URL,
                                  username= cherrypy.session[‘username’],
                                  password=cherrypy.session[‘password’],
                              )
```

```
2017-08-05 00:22:29.046 3834 WARNING nova.scheduler.utils [req-
89c159c7-b40a-43eb
Traceback (most recent call last):
```

I’m trying to list projects (but the same thing occurs whether I try to do that, ..)

When I create the instance in the dashboard, I get error:
In the /var/log/nova/nova-conductor.log file, there is the log

```
Unauthorized: The request you have made requires authentication.
```

The nova service are all up:

```
[root@ha-node1~] # nova service-list
```

Code snippet ← code

Logs ← log

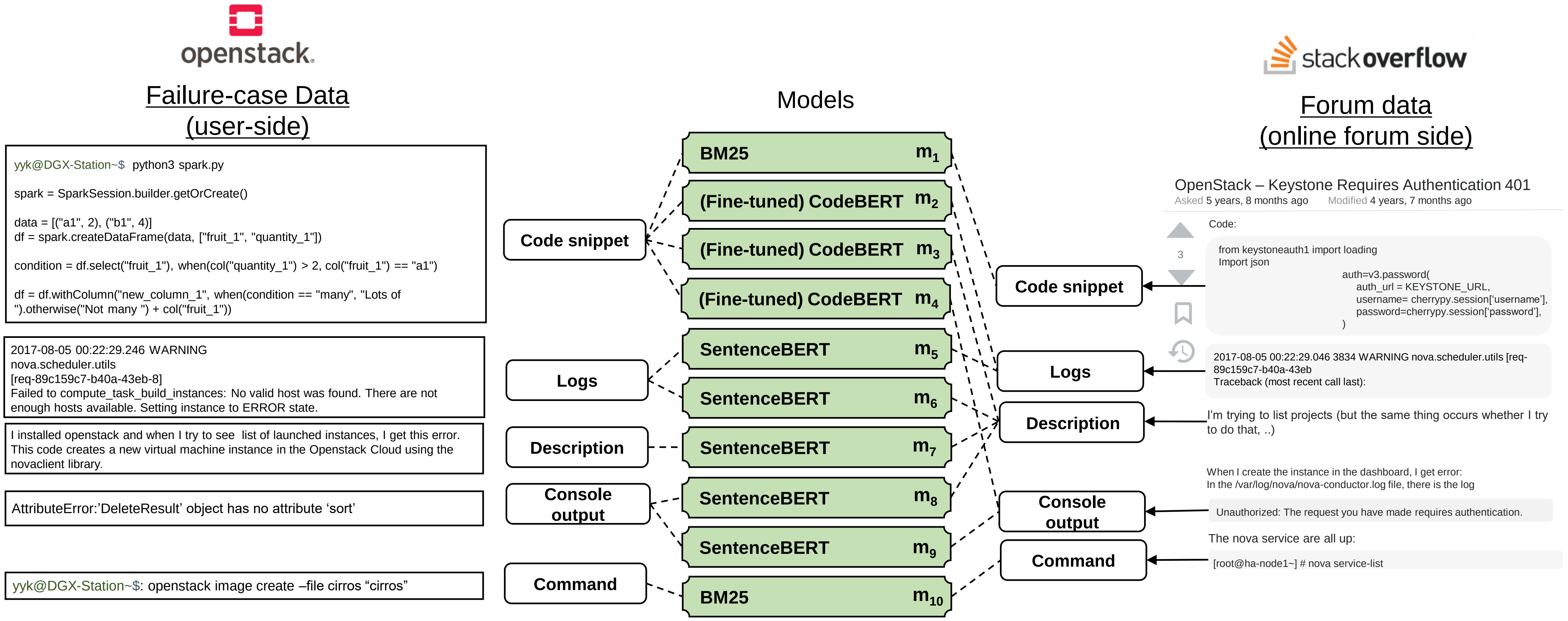
Description ← description

Console output ← error message

Command ← command

How to Compare Failure-Case Data and Online Forum Posts?

- Approach: Separating forum posts and applying 3 types of model



```
yyk@DGX-Station~$  python3 spark.py

spark = SparkSession.builder.getOrCreate()

data = [("a1", 2), ("b1", 4)]
df = spark.createDataFrame(data, ["fruit_1", "quantity_1"])

condition = df.select("fruit_1"), when(col("quantity_1") > 2, col("fruit_1") == "a1")

df = df.withColumn("new_column_1", when(condition == "many", "Lots of
").otherwise("Not many ") + col("fruit_1"))
```

```
2017-08-05 00:22:29.246 WARNING
nova.scheduler.utils
[req-89c159c7-b40a-43eb-8]
Failed to compute_task_build_instances: No valid host was found. There are not
enough hosts available. Setting instance to ERROR state.
```

I installed openstack and when I try to see list of launched instances, I get this error. This code creates a new virtual machine instance in the Openstack Cloud using the novaclient library.

```
AttributeError:'DeleteResult' object has no attribute 'sort'
```

```
yyk@DGX-Station~$: openstack image create –file cirros "cirros"
```

OpenStack – Keystone Requires Authentication 401

Asked 5 years, 8 months ago Modified 4 years, 7 months ago

Code:

```
from keystoneauth1 import loading
Import json
                    auth=v3.password(
                        auth_url = KEYSTONE_URL,
                        username= cherrypy.session['username'],
                        password=cherrypy.session['password'],
                    )
```

```
2017-08-05 00:22:29.046 3834 WARNING nova.scheduler.utils [req-
89c159c7-b40a-43eb
Traceback (most recent call last):
```

I'm trying to list projects (but the same thing occurs whether I try to do that, ..)

When I create the instance in the dashboard, I get error:
In the /var/log/nova/nova-conductor.log file, there is the log

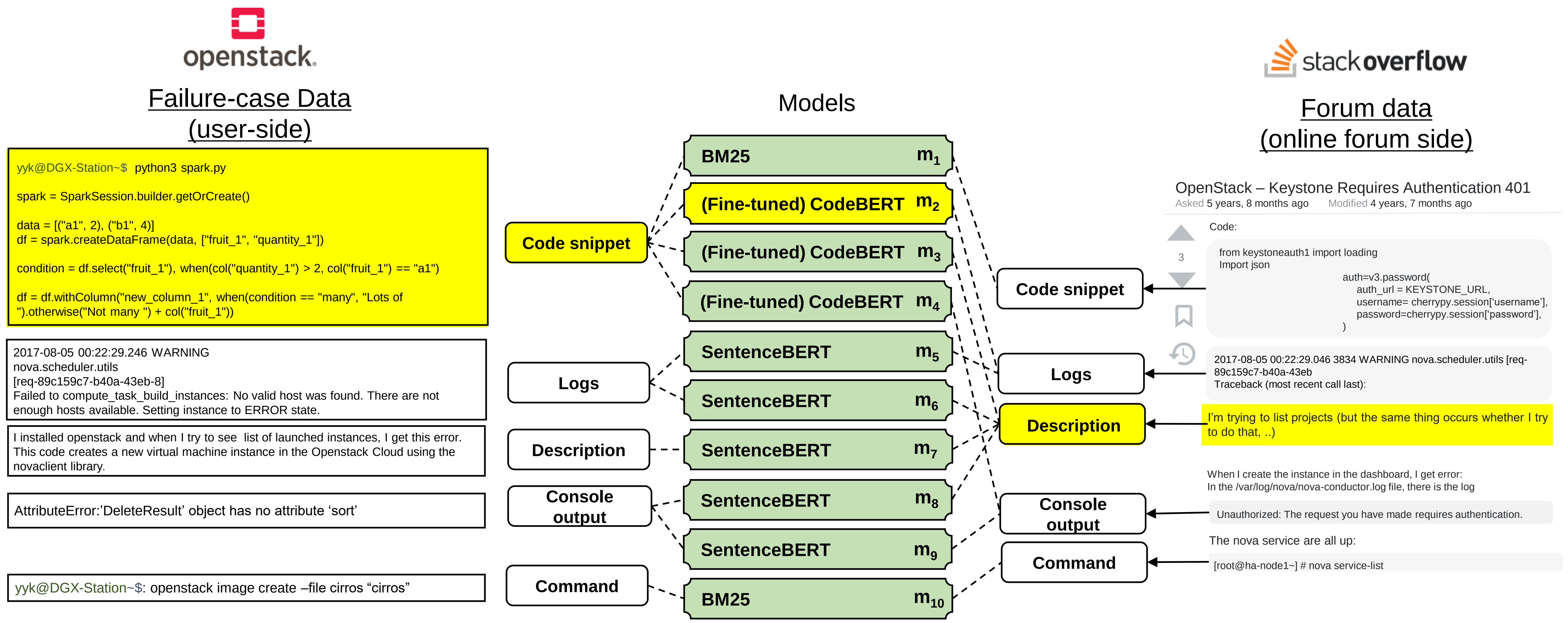
```
Unauthorized: The request you have made requires authentication.
```

The nova service are all up:

```
[root@ha-node1~] # nova service-list
```

How to Compare Failure-Case Data and Online Forum Posts?

- Approach: Separating forum posts and applying 3 types of model



```
yyk@DGX-Station~$: python3 spark.py

spark = SparkSession.builder.getOrCreate()

data = [("a1", 2), ("b1", 4)]
df = spark.createDataFrame(data, ["fruit_1", "quantity_1"])

condition = df.select("fruit_1"), when(col("quantity_1") > 2, col("fruit_1") == "a1")

df = df.withColumn("new_column_1", when(condition == "many", "Lots of
").otherwise("Not many ") + col("fruit_1"))
```

```
2017-08-05 00:22:29.246 WARNING
nova.scheduler.utils
[req-89c159c7-b40a-43eb-8]
Failed to compute_task_build_instances: No valid host was found. There are not
enough hosts available. Setting instance to ERROR state.
```

I installed openstack and when I try to see list of launched instances, I get this error. This code creates a new virtual machine instance in the Openstack Cloud using the novaclient library.

```
AttributeError:'DeleteResult' object has no attribute 'sort'
```

```
yyk@DGX-Station~$: openstack image create –file cirros “cirros”
```

OpenStack – Keystone Requires Authentication 401

Asked 5 years, 8 months ago   Modified 4 years, 7 months ago

Code:

```
from keystoneauth1 import loading
Import json
                    auth=v3.password(
                        auth_url = KEYSTONE_URL,
                        username= cherrypy.session[‘username’],
                        password=cherrypy.session[‘password’],
                    )
```

```
2017-08-05 00:22:29.046 3834 WARNING nova.scheduler.utils [req-
89c159c7-b40a-43eb
Traceback (most recent call last):
```

I’m trying to list projects (but the same thing occurs whether I try to do that, ..)

When I create the instance in the dashboard, I get error:
In the /var/log/nova/nova-conductor.log file, there is the log

```
Unauthorized: The request you have made requires authentication.
```

The nova service are all up:

```
[root@ha-node1~] # nova service-list
```

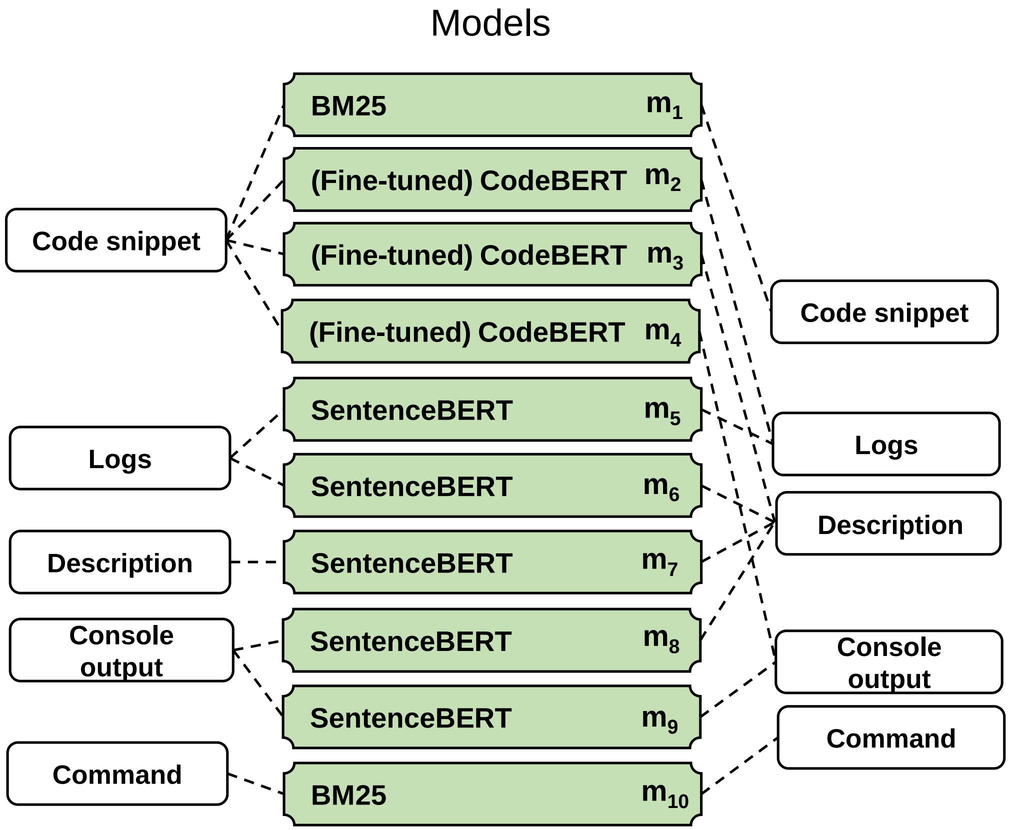
3 Models for Computing the Relevance Score

- Let’s learn what characteristics of each data make it use CodeBERT, S-BERT, BM25
- **Relevance score computing mechanism**

3 Models for Computing the Relevance Score

- Let’s learn what characteristics of each data make it use CodeBERT, S-BERT, BM25

- **Relevance score computing mechanism**
  - Searching ***related data*** in post: CodeBERT
    - Searching relevant data with the user side code
    - Model is fine-tuned before used
    - *Cod:Tnd, Cod:Log, Cod:Cns*

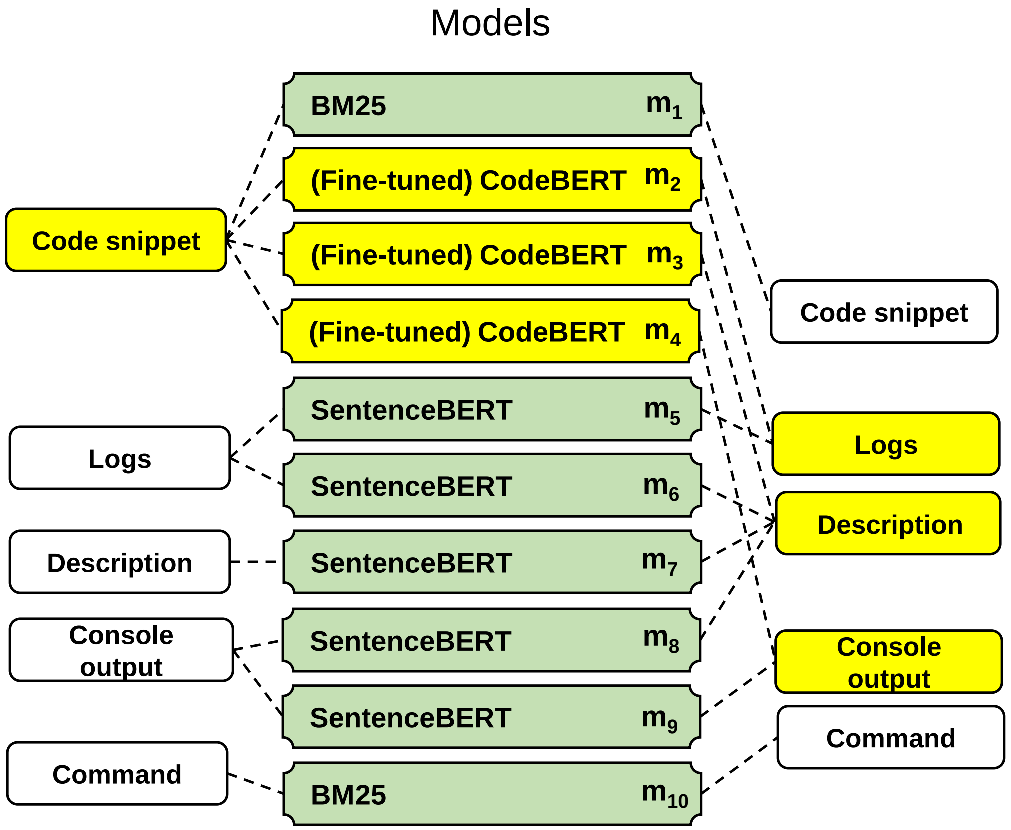


3 Models for Computing the Relevance Score

- Let's learn what characteristics of each data make it use CodeBERT, S-BERT, BM25

- **Relevance score computing mechanism**

    - Searching ***related data*** in post: <u>CodeBERT</u>
        - Searching relevant data with the user side code
        - Model is fine-tuned before used
        - *Cod:Tnd, Cod:Log, Cod:Cns*



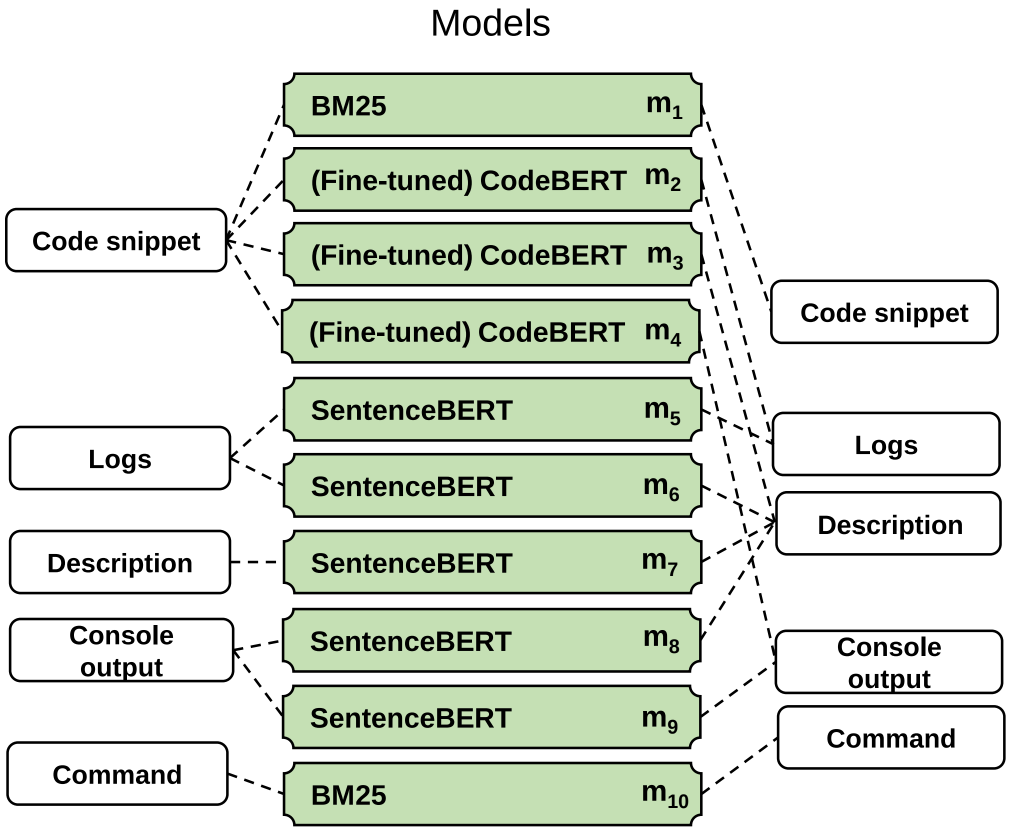
3 Models for Computing the Relevance Score

- Let's learn what characteristics of each data make it use CodeBERT, S-BERT, BM25

- **Relevance score computing mechanism**

  - Searching ***related data*** in post: CodeBERT
    - Searching relevant data with the user side code
    - Model is fine-tuned before used
    - *Cod:Tnd, Cod:Log, Cod:Cns*

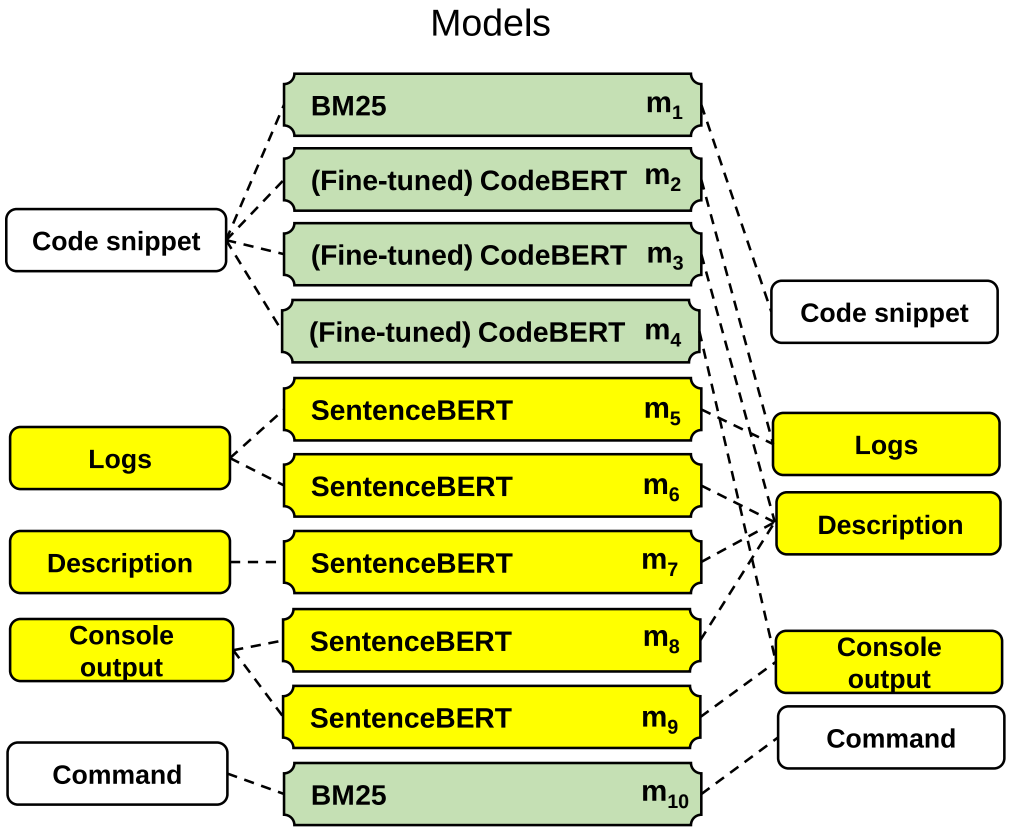
  - Searching ***similar sentence*** in post: SentenceBERT
    - Log, Cns: The same text is posted in an online forum
    - *Comparing data such as*
      - *Log:Log, Log:Tnd*
      - *Cns:Cns, Cns:Tnd*
      - *Des:Tnd*



3 Models for Computing the Relevance Score

- Let’s learn what characteristics of each data make it use CodeBERT, S-BERT, BM25

- **Relevance score computing mechanism**
  - Searching ***related data*** in post: CodeBERT
    - Searching relevant data with the user side code
    - Model is fine-tuned before used
    - *Cod:Tnd, Cod:Log, Cod:Cns*
  - Searching ***similar sentence*** in post: SentenceBERT
    - Log, Cns: The same text is posted in an online forum
    - *Comparing data such as*
      - *Log:Log, Log:Tnd*
      - *Cns:Cns, Cns:Tnd*
      - *Des:Tnd*

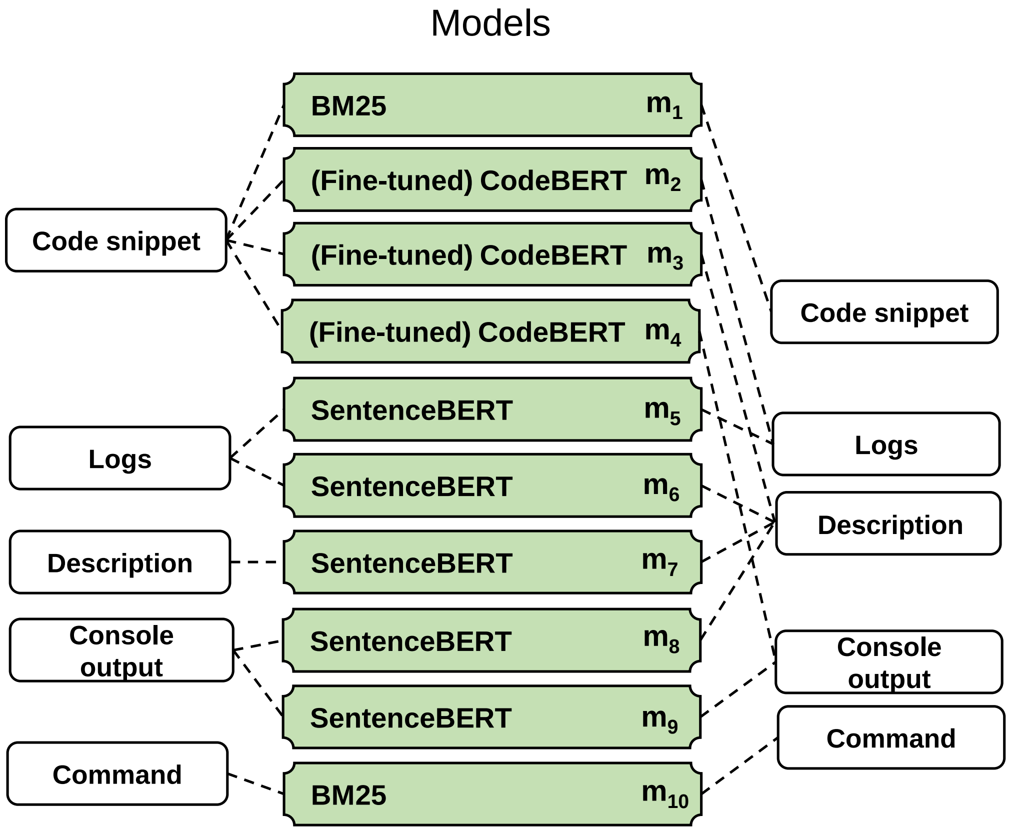


3 Models for Computing the Relevance Score

- Let’s learn what characteristics of each data make it use CodeBERT, S-BERT, BM25

- **Relevance score computing mechanism**

  - Searching ***related data*** in post: CodeBERT
    - Searching relevant data with the user side code
    - Model is fine-tuned before used
    - *Cod:Tnd, Cod:Log, Cod:Cns*
  - Searching ***similar sentence*** in post: SentenceBERT
    - Log, Cns: The same text is posted in an online forum
    - *Comparing data such as*
      - *Log:Log, Log:Tnd*
      - *Cns:Cns, Cns:Tnd*
      - *Des:Tnd*
  - Searching ***same token*** in post: BM25
    - Command : Fixed format so same data exist in post
      - e.g. “Openstack create image”
    - Code : Sharing same word when using API
    - Comparing data such as ***Cod:Cod, Cmd:Cmd***



3 Models for Computing the Relevance Score

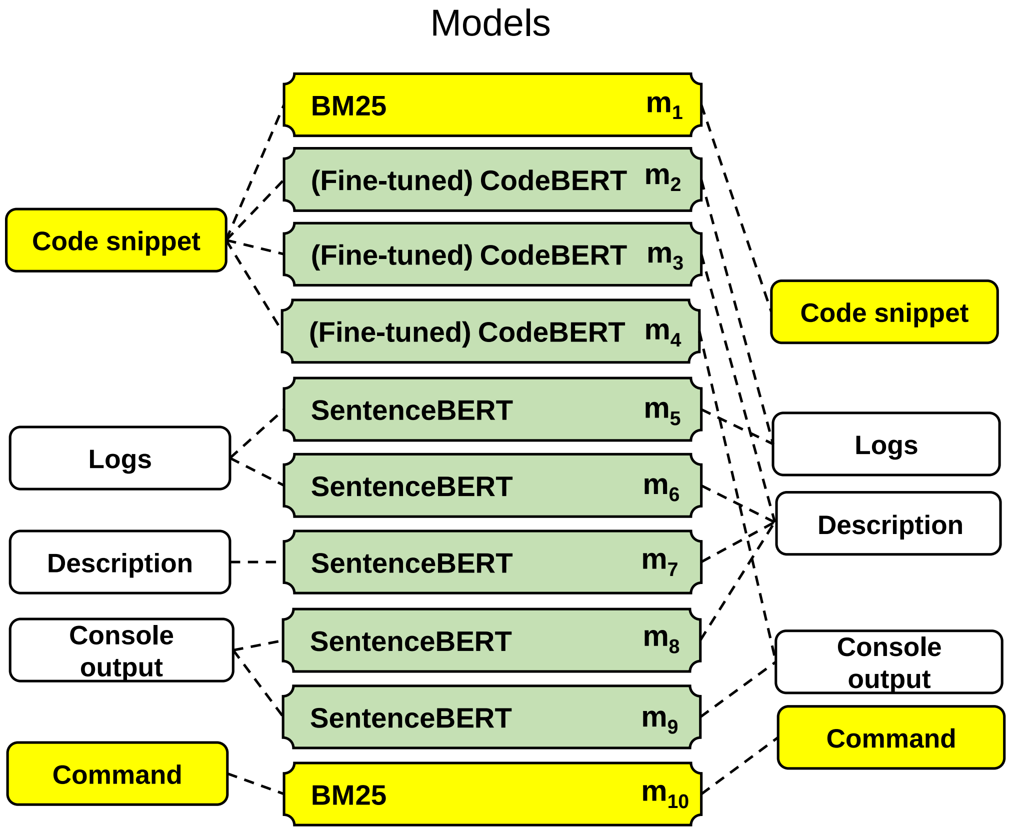
- Let's learn what characteristics of each data make it use CodeBERT, S-BERT, BM25

- **Relevance score computing mechanism**

  - Searching ***related data*** in post: CodeBERT
    - Searching relevant data with the user side code
    - Model is fine-tuned before used
    - *Cod:Tnd, Cod:Log, Cod:Cns*

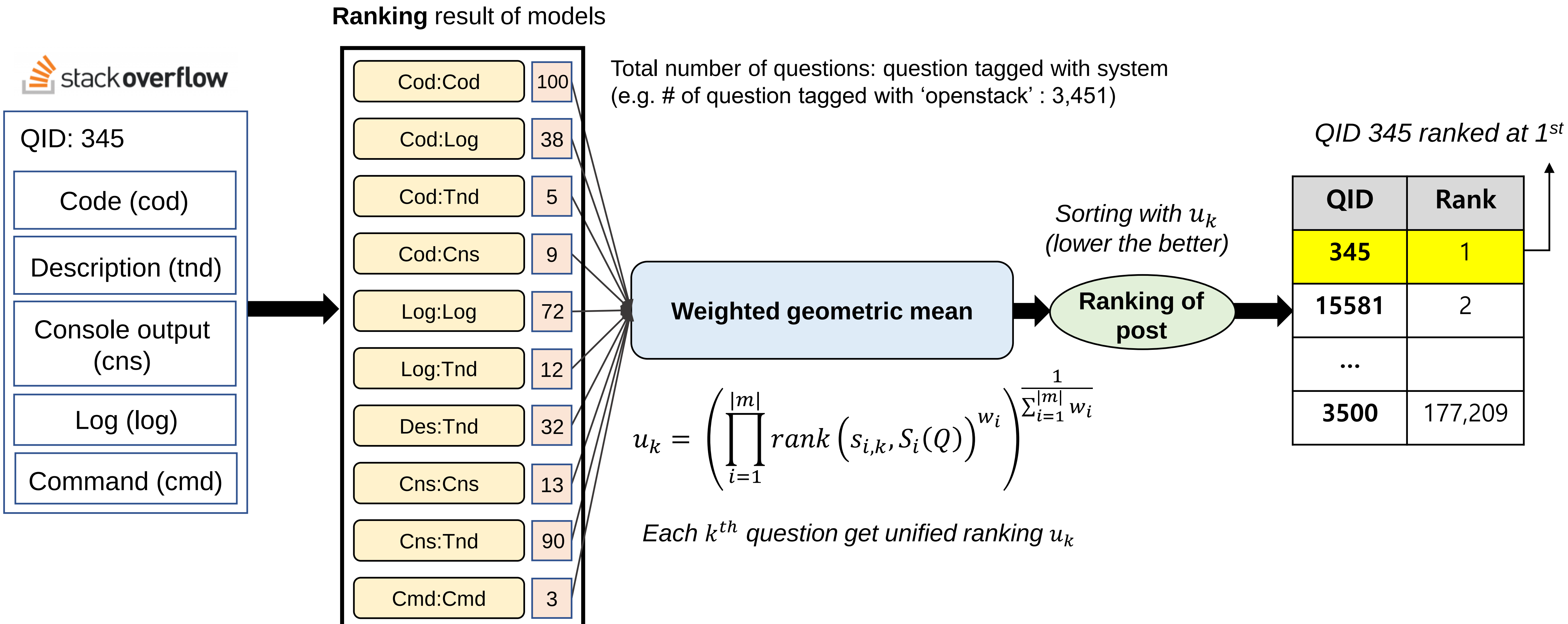
  - Searching ***similar sentence*** in post: SentenceBERT
    - Log, Cns: The same text is posted in an online forum
    - *Comparing data such as*
      - *Log:Log, Log:Tnd*
      - *Cns:Cns, Cns:Tnd*
      - *Des:Tnd*

  - Searching ***same token*** in post: BM25
    - Command : Fixed format so same data exist in post
      - e.g. "Openstack create image"
    - Code : Sharing same word when using API
    - Comparing data such as ***Cod:Cod, Cmd:Cmd***



How to Combine Results of Multiple Models?

- Combining results with weighted geometric mean and ranking documents



Ranking result of models

Cod:Cod	100
Cod:Log	38
Cod:Tnd	5
Cod:Cns	9
Log:Log	72
Log:Tnd	12
Des:Tnd	32
Cns:Cns	13
Cns:Tnd	90
Cmd:Cmd	3

Total number of questions: question tagged with system
(e.g. # of question tagged with ‘openstack’ : 3,451)

Sorting with u_k
(lower the better)

QID 345 ranked at 1st

QID	Rank
345	1
15581	2
…	
3500	177,209

$$u_k = \left(\prod_{i=1}^{|m|} rank\left(s_{i,k}, S_i(Q)\right)^{w_i} \right)^{\frac{1}{\sum_{i=1}^{|m|} w_i}}$$

Each k^{th} question get unified ranking u_k

1) Question in Stack Overflow

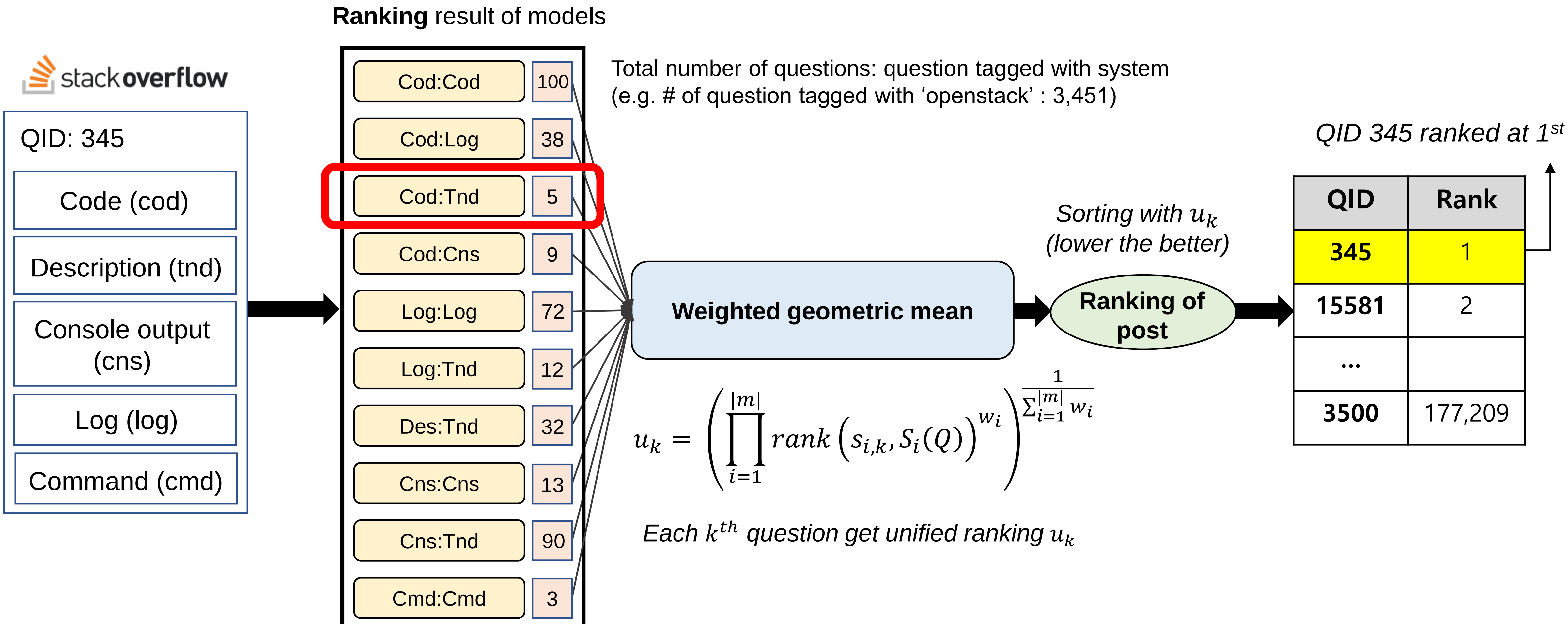
2) Get ranking result

3) Combining rankings of model using weighted geometric mean

4) Ranking

How to Combine Results of Multiple Models?

- Combining results with weighted geometric mean and ranking documents



Total number of questions: question tagged with system
(e.g. # of question tagged with ‘openstack’ : 3,451)

Sorting with u_k
(lower the better)

QID 345 ranked at 1st

QID	Rank
345	1
15581	2
…	
3500	177,209

$$u_k = \left(\prod_{i=1}^{|m|} rank\left(s_{i,k}, S_i(Q)\right)^{w_i} \right)^{\frac{1}{\sum_{i=1}^{|m|} w_i}}$$

Each k^{th} question get unified ranking u_k

1) Question in Stack Overflow

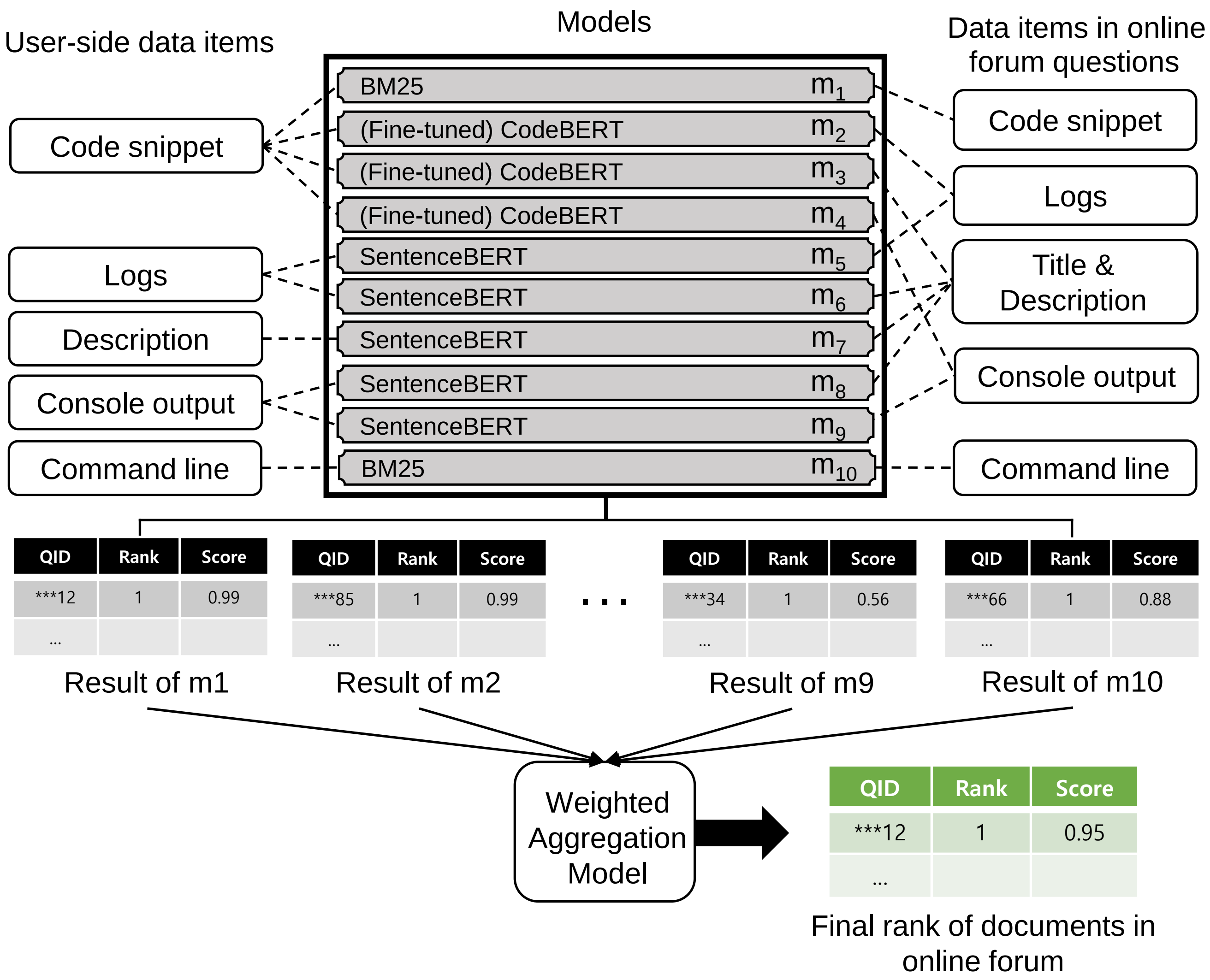
2) Get ranking result

3) Combining rankings of model using weighted geometric mean

4) Ranking

Overall Architecture of FDD

- Goal: Improving the accuracy of searching online forum post related to problem



QID	Rank	Score
***12	1	0.99
...		

Result of m1

QID	Rank	Score
***85	1	0.99
...		

Result of m2

. . .

QID	Rank	Score
***34	1	0.56
...		

Result of m9

QID	Rank	Score
***66	1	0.88
...		

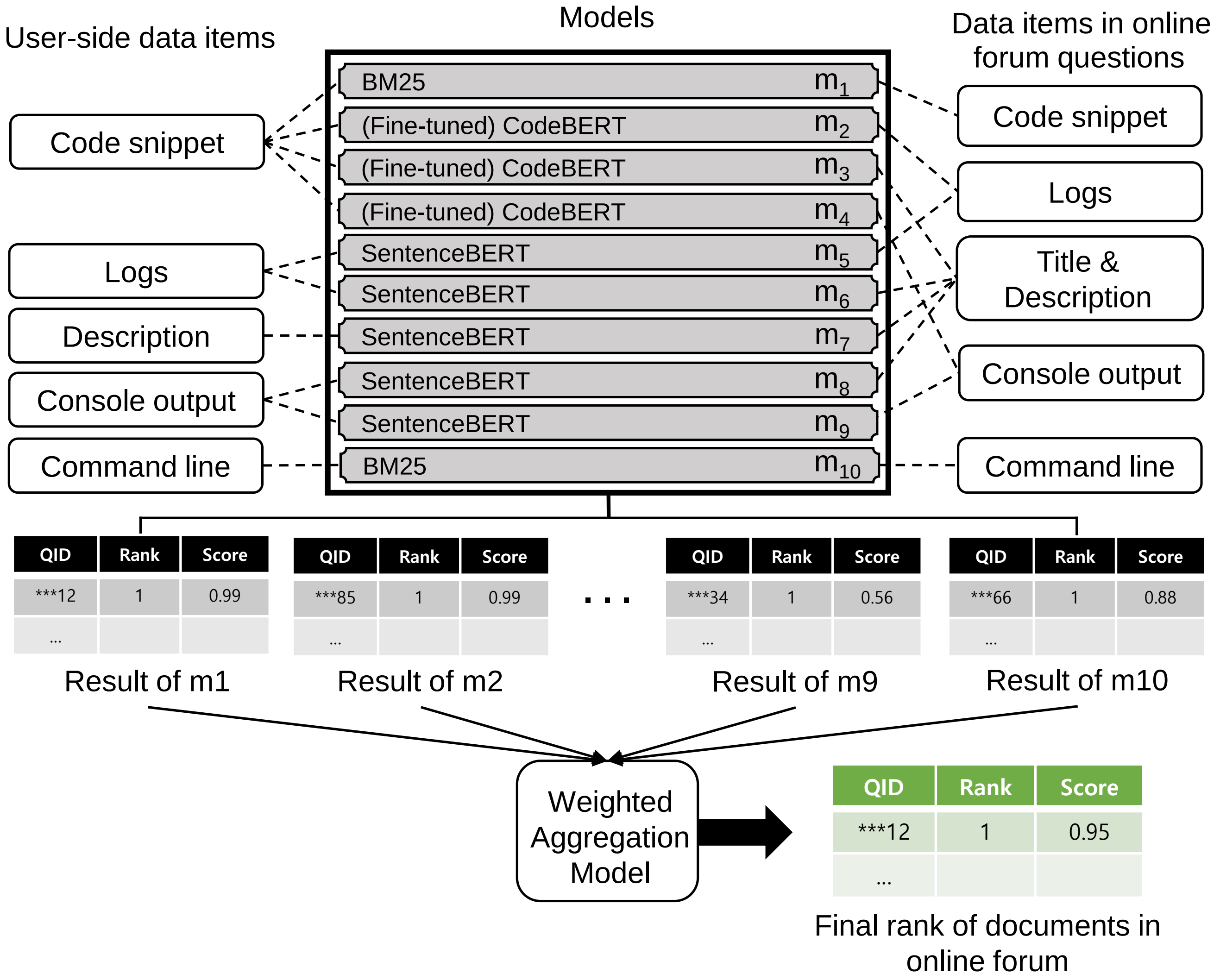
Result of m10

QID	Rank	Score
***12	1	0.95
...		

Final rank of documents in online forum

Overall Architecture of FDD

- Goal: Improving the accuracy of searching online forum post related to problem



QID	Rank	Score
***12	1	0.99
...		

Result of m1

QID	Rank	Score
***85	1	0.99
...		

Result of m2

QID	Rank	Score
***34	1	0.56
...		

Result of m9

QID	Rank	Score
***66	1	0.88
...		

Result of m10

QID	Rank	Score
***12	1	0.95
...		

Final rank of documents in online forum

1. Separating the data in online forum posts and user-side

2. Applying different types of model for different types of data

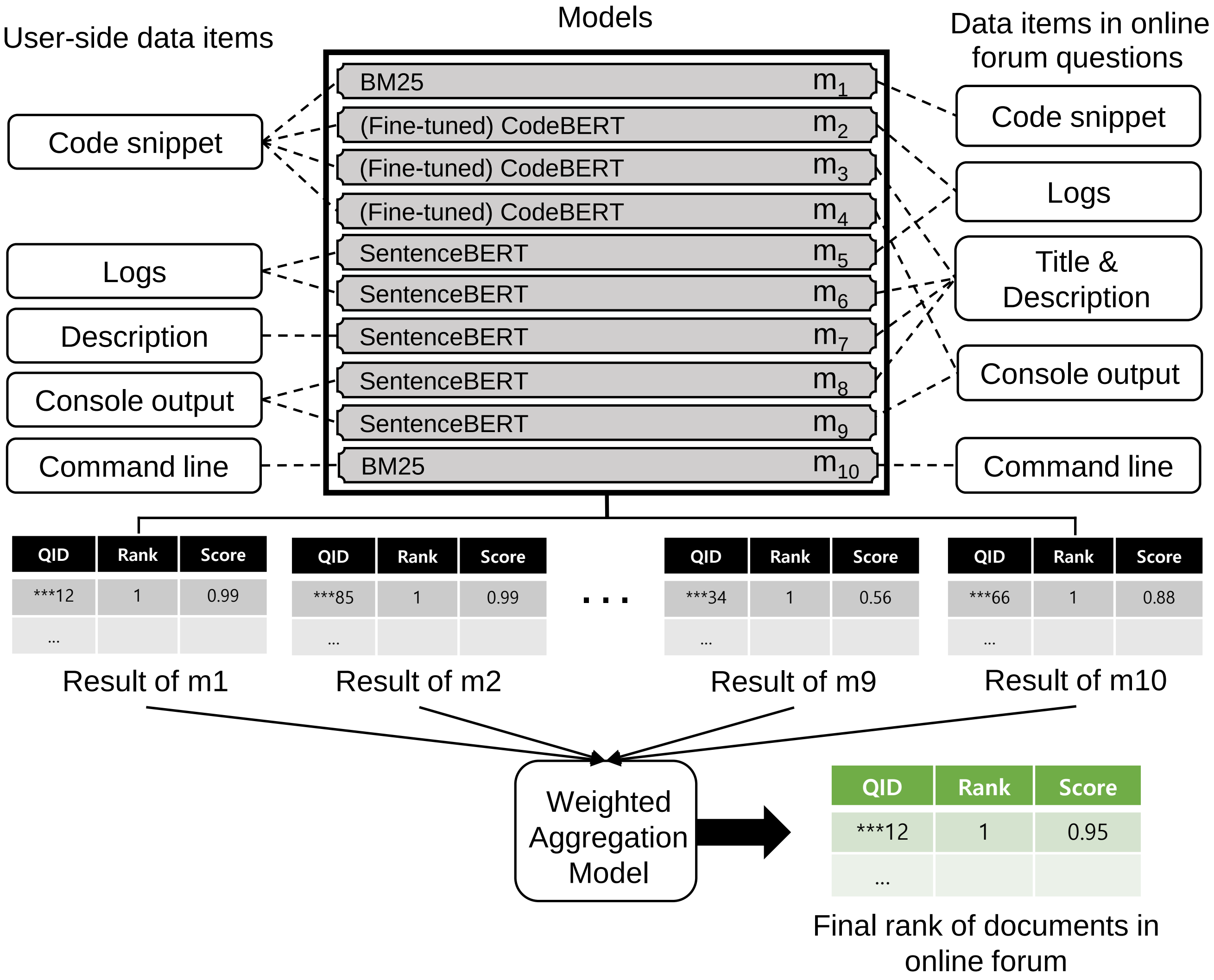
3. Aggregating result of each model and get final ranking of posts

Overall Architecture of FDD

- Goal: Improving the accuracy of searching online forum post related to problem

FDD:

Forum **D**ata **D**istiller



QID	Rank	Score
***12	1	0.99
…		

Result of m1

QID	Rank	Score
***85	1	0.99
…		

Result of m2

QID	Rank	Score
***34	1	0.56
…		

Result of m9

QID	Rank	Score
***66	1	0.88
…		

Result of m10

QID	Rank	Score
***12	1	0.95
…		

Final rank of documents in online forum

1. **Separating the data** in online forum posts and user-side

2. **Applying different types of model** for different types of data

3. **Aggregating result** of each model and get final ranking of posts

How to Evaluate FDD?

- Benchmark cases
  - 5 Target application: Openstack, MongoDB, Spark, Elasticsearch, Kafka
  - **77 Failure cases** are manually reproduced

stackoverflow

QID
5675
5676
…
172,209

1) Selecting question

2) Reproducing failure cases

Case No. 15	
Command	Openstack create image –file cirros
Console output	The request you have made requires authentication
Description	I want to create image in openstack

3) Collecting operation context & operation output

- Evaluation metric: MRR / SuccessRate@K

MRR	SuccessRate@K
$\mathrm{MRR} = \frac{1}{\lvert Q\rvert}\sum_{i=1}^{\lvert Q\rvert}\frac{1}{\mathrm{rank}_i}$	$SuccessRate@K = \frac{1}{\lvert C\rvert}\sum_{c=1}^{c}\delta(Rank_c \le k), k \in \{1,5,10\}$
• ***How high the question is ranked*** • If all questions ranked 1st, MRR = 1 • Higher the better	• ***How many questions are ranked within K*** • If all questions ranked 1st, SuccessRate@K = 1 • Higher the better

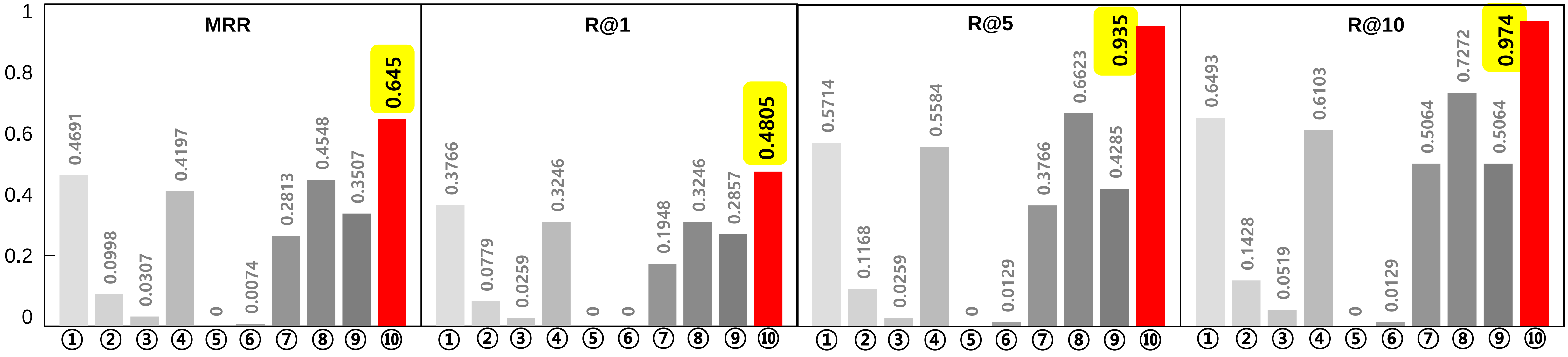
Result of Spark Using FDD (Total postings: ***100,729***)

CaseNo.	Type of Case	BM25	Doc2vec	BERT	SBERT	BERT(F)	CBERT(F)	G (a)	G (b)	ST	ChatGPT	FDD
S1	Kmeans with python	10	486	18078	35	71576	656	NF	7	1	X	2
S2	Connect Mongodb	1	33857	28359	18	86309	3099	20	1	NF	X	1
S3	Calling java/scala function	17	3119	11332	1	67103	8053	69	69	4	X	3
S4	Create dataframe	3	7025	6975	867	31954	7731	NF	NF	NF	X	1
S5	Use multiple conditions	6	271	354	1	19115	23058	26	26	6	X	1
S6	Pyspark mongodb connect	432	14512	2360	4	80537	3795	13	13	NF	O	4
S7	Check dataframe	21340	288	74353	4462	29628	12809	23	23	14	X	3
S8	Create spark session	1	13857	8987	1	66978	17214	3	1	1	X	2
S9	Create udf	1	4526	47966	39	75390	421	10	4	NF	O	1
S10	Read and write table	6	3101	36404	2	76320	16773	1	1	1	O	1
S11	Set spark configuration	2	86	21152	158	86366	32885	1	1	1	X	1
S12	Delete and recreate context	5	27741	6652	87	63470	27724	21	2	17	X	4
S13	Create dataframe	1	4536	3538	94	78517	38046	NF	NF	NF	X	4
S14	Run logistic regression	140	1786	21206	26	73910	4474	1	1	1	O	1
S15	Run mllib in pyspark	55988	19610	37652	189	76428	8223	NF	NF	NF	X	5
S16	Initializing spark context	1	235	4189	2	50009	7538	4	3	5	O	1
S17	Run spark on remote cluster	2	911	4689	5	82692	12404	6	2	9	X	1
S18	Run spark in cluster mode	47350	82365	29435	16193	98211	26704	40	12	NF	X	3

Improved Searching Accuracy With FDD

- Ranking quality comparison
  - Comparing ranking of ground-truth with 9 baselines

① BM25 ② Doc2vec ③ BERT(CLS) ④ SentenceBERT ⑤ BERT(Fine-tuned)
⑥ CodeBERT(Fine-tuned) ⑦ Google Search (a) ⑧ Google Search (b)
⑨ Stack Overflow Search ⑩ FDD

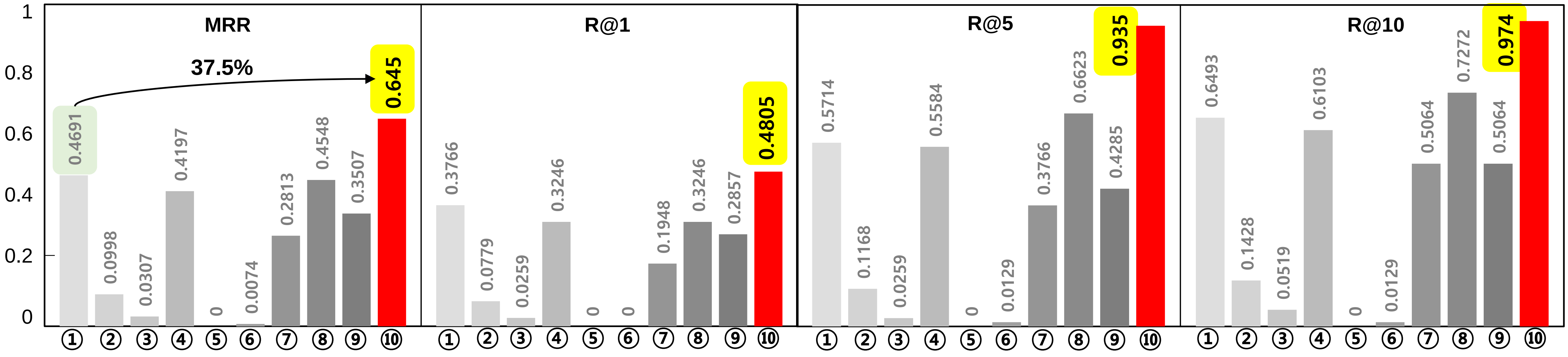


- Result analysis
  - **FDD** produced better results than all baselines. FDD ***improves accuracy by 37.5%*** compared with BM25.
  - FDD ***ranked most of the cases(75/77) within 10th .***
  - **BM25** showed the highest MRR among the baseline. High ranking only in **command-based cases**.
  - By Google Search (b), 30% of cases were ranked at 1st. But the **variance is too high** (too many ‘Not Found’)

Improved Searching Accuracy With FDD

- Ranking quality comparison
  - Comparing ranking of ground-truth with 9 baselines

① BM25 ② Doc2vec ③ BERT(CLS) ④ SentenceBERT ⑤ BERT(Fine-tuned)
⑥ CodeBERT(Fine-tuned) ⑦ Google Search (a) ⑧ Google Search (b)
⑨ Stack Overflow Search ⑩ FDD



- Result analysis
  - **FDD** produced better results than all baselines. FDD ***improves accuracy by 37.5%*** compared with BM25.
  - FDD ***ranked most of the cases(75/77) within 10th .***
  - **BM25** showed the highest MRR among the baseline. High ranking only in **command-based cases**.
  - By Google Search (b), 30% of cases were ranked at 1st. But the **variance is too high** (too many 'Not Found')

Conclusion

- We try to troubleshoot ***by finding the most relevant online forum post*** with system failure.
- The approach of ***recognizing multiple disparate data items within the online posting data*** and ***performing cross-data search*** shows high effectiveness.
  - On 5 diverse applications, for 75 reproduced failure cases, our technique gave **single-digit ranking** of true online posts.
  - The accuracy improved 37.5% compared against best performing competitor.
- Limitation of FDD
  - Cross-data type search may not always work
    - logs or console output evolved too much across versions
  - Our accuracy is bound by the performance of the modeling technique
    - CodeBERT in FDD only accept code with less than 512 tokens.