INTEL INDEX REPORT

K.Y.Sai Haritha

13-09-2019
B.Tech Final Year
Computer Science and Engineering
IIIT Nuzvid,India

INTRODUCTION

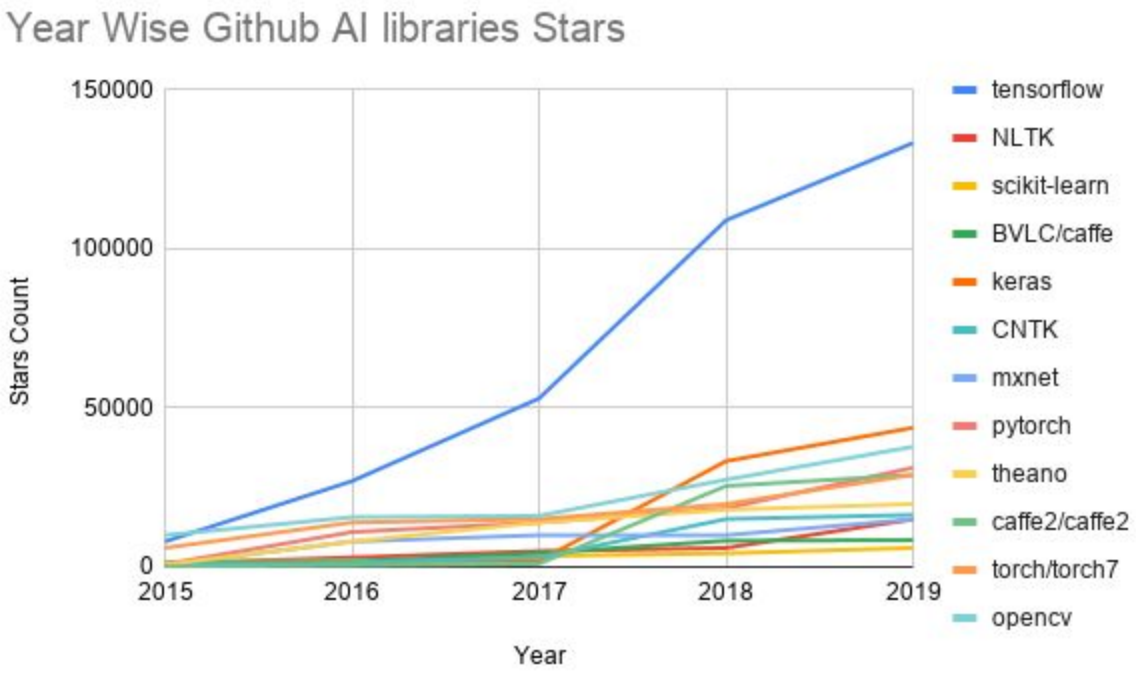
Drawing insights from the data stored in any public repositories is one of the tasks that many of the researchers working now-a-days . A lot of information can be gathered in this huge process and information can be used in many analytics . For example , based on daily commits of code and code quality , it will be useful for crediting the student along with the exam results so that actual skill of student can be identified .

HYPOTHESIS

Extracting the projects belonging to Artificial intelligence,machine learning,deep learning , data science , data analytics along with their stars on github.From this, yearly based , monthly based , daily based statistics of AI libraries which are more popular (case : in this project , based on stars count) can be drawn.

STATISTICS

The following graphs show the number of times various AI and ML software packages have been starred on gitlab-ce (in development phase) . This provides a measure of the popularity of various *AI programming frameworks* .



EXCEL DATA

	A	B	C	D	E	F
1		2019	2018	2017	2016	2015
2	tensorflow	133322	109031	53000	27000	8000
3	NLTK	15000	6000	4980	3000	1100
4	scikit-learn	6000	4300	3450	2000	1500
5	BVLC/caffe	8463	8260	4390	1800	1200
6	keras	43786	33284	1950	1500	700
7	CNTK	16375	15078	3000	1500	900
8	mxnet	15000	9992	10000	8000	700
9	pytorch	31231	18463	13970	11000	900
10	theano	19800	18000	14000	8000	800
11	caffe2/caffe2	28942	25466	1000	500	200
12	torch/torch7	29000	19800	15080	14000	6000
13	opencv	37798	27462	16098	15678	10000

Notes

Some of the data is manually entered into database to interpret the output.

REQUIREMENTS

1. Python 3+ versions
2. Mysql database
3. Grafana

PROCEDURE

1. Flask will start running on http://localhost:5000.
2. /ai is the API triggered by webhook configured with push event . Here gitlab-ce used for testing purpose.
3. Information from webhook data which is useful will be storing into database.
4. Webhook data will be sent to our classification model to classify into AI or non-AI project.
5. Project information will be retrieved from gitlab using access token to know whether it is starred or not and if then its star count will be returned.
6. This all information will be storing in database to draw statistics using grafana.
7. We check if project is both AI and starred then it will be in our count as STARRED-AI-PROJECT.

EXECUTION

```
python3 init.py
```

INPUT DATA

The input data will be the webhook data . It consists of different data for different types of events . From Push event , data such as project id , project name , commit id , timestamp , git_http_url , user id etc which are useful are extracted.

PUSH event webhook data looks like :


```
{
 "object_kind": "push",
 "before": "95790bf891e76fee5e1747ab589903a6a1f80f22",
  "after": "da1560886d4f094c3e6c9ef40349f7d38b5d27d7",
  "ref": "refs/heads/master",
  "checkout_sha": "da1560886d4f094c3e6c9ef40349f7d38b5d27d7",
  "user_id": 4,
  "user_name": "John Smith",
  "user_username": "jsmith",
  "user_email": "john@example.com",
  "user_avatar":
"https://s.gravatar.com/avatar/d4c74594d841139328695756648b6bd6?s=8://s.gravata
r.com/avatar/d4c74594d841139328695756648b6bd6?s=80",
  "project_id": 15,
  "project":{
    "id": 15,
    "name":"Diaspora",
    "description":"",
    "web_url":"http://example.com/mike/diaspora",
```

```
    "avatar_url":null,
    "git_ssh_url":"git@example.com:mike/diaspora.git",
    "git_http_url":"http://example.com/mike/diaspora.git",
    "namespace":"Mike",
    "visibility_level":0,
    "path_with_namespace":"mike/diaspora",
    "default_branch":"master",
    "homepage":"http://example.com/mike/diaspora",
    "url":"git@example.com:mike/diaspora.git",
    "ssh_url":"git@example.com:mike/diaspora.git",
    "http_url":"http://example.com/mike/diaspora.git"
  },
  "repository":{
    "name": "Diaspora",
    "url": "git@example.com:mike/diaspora.git",
    "description": "",
    "homepage": "http://example.com/mike/diaspora",
    "git_http_url":"http://example.com/mike/diaspora.git",
    "git_ssh_url":"git@example.com:mike/diaspora.git",
    "visibility_level":0
  },
  "commits": [
    {
      "id": "b6568db1bc1dcd7f8b4d5a946b0b91f9dacd7327",
      "message": "Update Catalan translation to e38cb41.",
```

```
      "timestamp": "2011-12-12T14:27:31+02:00",
      "url":
"http://example.com/mike/diaspora/commit/b6568db1bc1dcd7f8b4d5a946b0b91f9dacd73
27",
      "author": {
        "name": "Jordi Mallach",
        "email": "jordi@softcatala.org"
      },
      "added": ["CHANGELOG"],
      "modified": ["app/controller/application.rb"],
      "removed": []
    },
    {
      "id": "da1560886d4f094c3e6c9ef40349f7d38b5d27d7",
      "message": "fixed readme",
      "timestamp": "2012-01-03T23:36:29+02:00",
      "url":
"http://example.com/mike/diaspora/commit/da1560886d4f094c3e6c9ef40349f7d38b5d27
d7",
      "author": {
        "name": "GitLab dev user",
        "email": "gitlabdev@dv6700.(none)"
      },
      "added": ["CHANGELOG"],
      "modified": ["app/controller/application.rb"],
      "removed": []
```

```
    }
  ],
  "total_commits_count": 4
}
```

OUTPUT DATA

- ➔ Data regarding commits , stars count of a project and AI libraries are stored separately in mysql database.
- ➔ An overall output of AI libraries with stars count are files up in output/output.csv excel file .
- ➔ Can extract stats @yearly , @monthly ,@daily too from the retrieved information.

projectId	commitId	AI	Starred	Timestamp
1901	ed899a2f4b50b4370feeea94676502b42383c746	TRUE	TRUE	2019-09-11 9:13:56
551	aafa420b4b50bef7bc392a23545ab600cbbade37	FALSE	TRUE	2019-09-11 9:13:56
2304	6104942438c14ec7bd21c6cd5bd995272b3faff60	TRUE	TRUE	2019-09-11 9:13:56
1002	ae1d9fb46aa2b07ee9836d49862ec4e2c46fbbba	FALSE	FALSE	2019-09-14 10:14:45
1340	2dc6aa325a317eda67812f05600bdf0fcdc70ab0	FALSE	FALSE	2019-09-11 10:14:56
721	8b090c1b79a14f2bd9e8a738f717824ff53aebad	FALSE	TRUE	2018-09-12 5:35:46
2195	a738f717824ff53aebad8b090c1b79a14f2bd9e8	TRUE	FALSE	2018-08-12 21:14:14
2130	18f3e63d05582537db6d183d9d557be09e1f90c	FALSE	TRUE	2018-08-13 13:56:45
1612	af5b13261899fb2c0db30abdd0af8b07cb44fdc5	TRUE	TRUE	2017-08-12 11:18:25
1899	562e173be03b8ff2efb05345d12df18815438a4b	TRUE	FALSE	2019-09-11 21:36:46
1803	da1560886d4f094c3e6c9ef40349f7d38b5d27d7	FALSE	TRUE	2018-07-18 21:38:56
1732	bcbb5ec396a2c0f828686f14fac9b80b780504f20	TRUE	TRUE	2018-06-16 21:38:50
2106	2293ada6b400935a1378653304eaf6221e0fdb8f	TRUE	FALSE	2018-06-30 21:40:14
1105	48f2bc94e37e2f5bf000d9b3d8a10e4143335704	FALSE	TRUE	2018-06-27 22:15:26
2201	bdacd6241a7a81ba2e8f90399368037c611636ce	TRUE	TRUE	2017-09-24 4:27:40
2121	720d76208fe76c353df44ad8abc17ed65db8314f	FALSE	TRUE	2017-11-13 7:03:20
2231	13ee771de5b0532ccde8444621eea5e5154c766	TRUE	FALSE	2017-11-15 9:23:25
105	880e7f8ad09e13b6dbca37c3816e080a0c773ff6	TRUE	TRUE	2016-05-16 5:18:36
6	e59b81f15904be2237c419b61fbfde072ec91788	TRUE	FALSE	2016-05-16 22:05:28
13	c4daca529e3002c3249066ac0ddfe46cc155624c	TRUE	FALSE	2016-07-15 23:15:47

RESULTS

1. AI libraries and AI related projects are separated and stored successfully.
2. Projects with their **latest** stars count is calculated and stored.
3. Using grafana , year wise stats of AI Starred projects with star count is plotted and stored in mysql in a proper way

CONCLUSION

As already AI index report shows the github stars for AI libraries , but the stars count is not so accurate as they mentioned . With this approach of collecting information through commits , gives accurate results.

REFERENCES

1. https://docs.gitlab.com/ee/user/project/integrations/webhooks.html
2. https://www.tutorialspoint.com/mysql/index.htm
3. https://grafana.com/docs/
4. AI index report pdf