

Εφαρμογή τεχνικών πρόβλεψης των χαρακτηριστικών των προγραμματιστών και των έργων λογισμικού αποθετηρίων ανοικτού λογισμικού

Εργασία στο μάθημα της
Αναγνώρισης Προτύπων

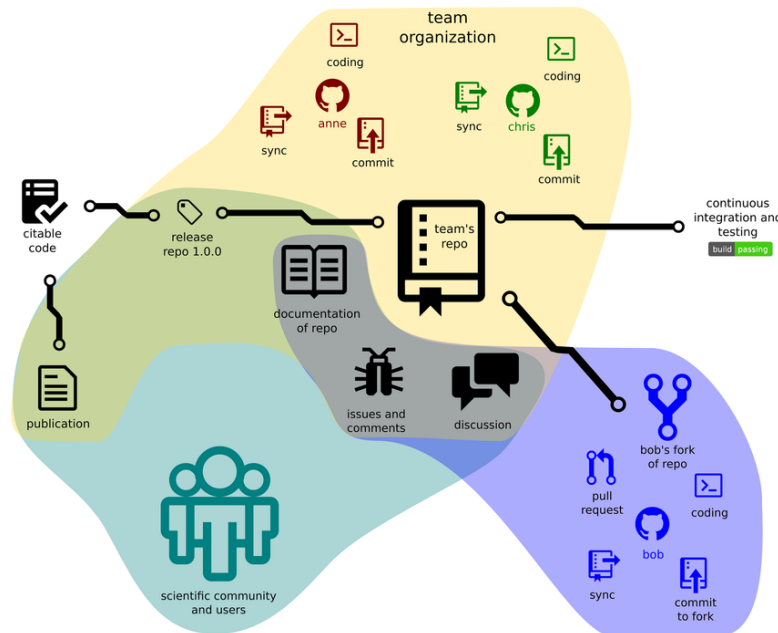
Έργα Ανοικτού Λογισμικού

- Ελεύθερα προσπελάσιμος κώδικας
- Συνεισφορά από κάθε ενδιαφερόμενο προγραμματιστή



Αποθετήρια Κώδικα

- Διατηρεί τον κώδικα
- Παρέχει πλήθος πληροφοριών (meta-data) για το project
- Παρέχει API* για την προσπέλαση των πληροφοριών



* Διεπαφή Προγραμματισμού Εφαρμογών (Application Programming Interface)

Application Programming Interface

- Προσπέλαση πληροφοριών για χρήστες, αποθετήρια Κ.Ο.Κ.
- Ελεύθερη πρόσβαση σε όλα τα open source αποθετήρια και τις πληροφορίες τους

```
[
  - {
    id: 136095779,
    node_id: "MDEwOlJlcG9zaXRvcnkxMzYwOTU3Nzk=",
    name: "agile-week",
    full_name: "timwheelercom/agile-week",
    private: false,
    - owner: {
      login: "timwheelercom",
      id: 17229444,
      node_id: "MDQ6VXN1c2E3MjI5NDQ0",
      avatar_url: "https://avatars1.githubusercontent.com/u/17229444?v=4",
      gravatar_id: "",
      url: "https://api.github.com/users/timwheelercom",
      html_url: "https://github.com/timwheelercom",
      followers_url: "https://api.github.com/users/timwheelercom/followers",
      following_url: "https://api.github.com/users/timwheelercom/following{/other_user}",
      gists_url: "https://api.github.com/users/timwheelercom/gists{/gist_id}",
      starred_url: "https://api.github.com/users/timwheelercom/starred{/owner}/{/repo}",
      subscriptions_url: "https://api.github.com/users/timwheelercom/subscriptions",
      organizations_url: "https://api.github.com/users/timwheelercom/orgs",
      repos_url: "https://api.github.com/users/timwheelercom/repos",
      events_url: "https://api.github.com/users/timwheelercom/events{/privacy}",
      received_events_url: "https://api.github.com/users/timwheelercom/received_events",
      type: "User",
      site_admin: false
    },
  },
]
```

Application Programming Interface (examples)

- <https://api.github.com/users/AuthEceSofteng>

```
1 {
2   "login": "AuthEceSoftEng",
3   "id": 6051438,
4   "node_id": "MDEyOk9yZ2FuaXphdGlvbG9wNTE0Mzg=",
5   "avatar_url": "https://avatars.githubusercontent.com/u/6051438?v=4",
6   "gravatar_id": "",
7   "url": "https://api.github.com/users/AuthEceSoftEng",
8   "html_url": "https://github.com/AuthEceSoftEng",
9   "followers_url": "https://api.github.com/users/AuthEceSoftEng/followers",
10  "following_url": "https://api.github.com/users/AuthEceSoftEng/following{/other_user}",
11  "gists_url": "https://api.github.com/users/AuthEceSoftEng/gists{/gist_id}",
12  "starred_url": "https://api.github.com/users/AuthEceSoftEng/starred{/owner}/{/repo}",
13  "subscriptions_url": "https://api.github.com/users/AuthEceSoftEng/subscriptions",
14  "organizations_url": "https://api.github.com/users/AuthEceSoftEng/orgs",
15  "repos_url": "https://api.github.com/users/AuthEceSoftEng/repos",
16  "events_url": "https://api.github.com/users/AuthEceSoftEng/events{/privacy}",
17  "received_events_url": "https://api.github.com/users/AuthEceSoftEng/received_events",
18  "type": "Organization",
19  "site_admin": false,
20  "name": "ISSEL Soft Eng Team",
21  "company": null,
22  "blog": "http://softeng.isSEL.ee.auth.gr",
23  "location": "Aristotle University of Thessaloniki, Greece",
24  "email": "softeng@olympus.ee.auth.gr",
25  "hireable": null,
26  "bio": "The software engineering team of the Intelligent Systems and Software Engineering Labgroup (ISSEL)",
27  "twitter_username": null,
28  "public_repos": 69,
29  "public_gists": 0,
30  "followers": 0,
31  "following": 0,
32  "created_at": "2013-11-27T15:51:25Z",
33  "updated_at": "2021-11-22T15:47:30Z"
34 }
```

Application Programming Interface (examples)

- <https://api.github.com/users/AuthEceSofteng/repos>

```
1 {
2   {
3     "id": 95766246,
4     "node_id": "MDEwO1JlcG9zaXRvcnk5NTc2NjI0Ng==",
5     "name": "abac-mde-teis",
6     "full_name": "AuthEceSoftEng/abac-mde-teis",
7     "private": false,
8     "owner": {
9       "login": "AuthEceSoftEng",
10      "id": 6051438,
11      "node_id": "MDEyOk9yZ2FuaXphdGlvbjYwNTE0Mzg=",
12      "avatar_url": "https://avatars.githubusercontent.com/u/6051438?v=4",
13      "gravatar_id": "",
14      "url": "https://api.github.com/users/AuthEceSoftEng",
15      "html_url": "https://github.com/AuthEceSoftEng",
16      "followers_url": "https://api.github.com/users/AuthEceSoftEng/followers",
17      "following_url": "https://api.github.com/users/AuthEceSoftEng/following{/other_user}",
18      "gists_url": "https://api.github.com/users/AuthEceSoftEng/gists{/gist_id}",
19      "starred_url": "https://api.github.com/users/AuthEceSoftEng/starred{/owner}/{/repo}",
20      "subscriptions_url": "https://api.github.com/users/AuthEceSoftEng/subscriptions",
21      "organizations_url": "https://api.github.com/users/AuthEceSoftEng/orgs",
22      "repos_url": "https://api.github.com/users/AuthEceSoftEng/repos",
23      "events_url": "https://api.github.com/users/AuthEceSoftEng/events{/privacy}",
24      "received_events_url": "https://api.github.com/users/AuthEceSoftEng/received_events",
25      "type": "Organization",
26      "site_admin": false
27    },
28    "size": 8983,
29    "stargazers_count": 1,
30    "watchers_count": 1,
31    "language": "Java",
32    "has_issues": true,
33    "has_projects": true,
34    "has_downloads": true,
35    "has_wiki": true,
36    "has_pages": false,
37    "forks_count": 0,
38    "mirror_url": null,
39    "archived": false,
40    "disabled": false,
41    "open_issues_count": 0,
42    "license": null,
43    "allow_forking": true,
44    "is_template": false,
45    "topics": [],
46    "visibility": "public",
47    "forks": 0,
48    "open_issues": 0,
49    "watchers": 1,
50    "default_branch": "master"
51  },
52 }
```

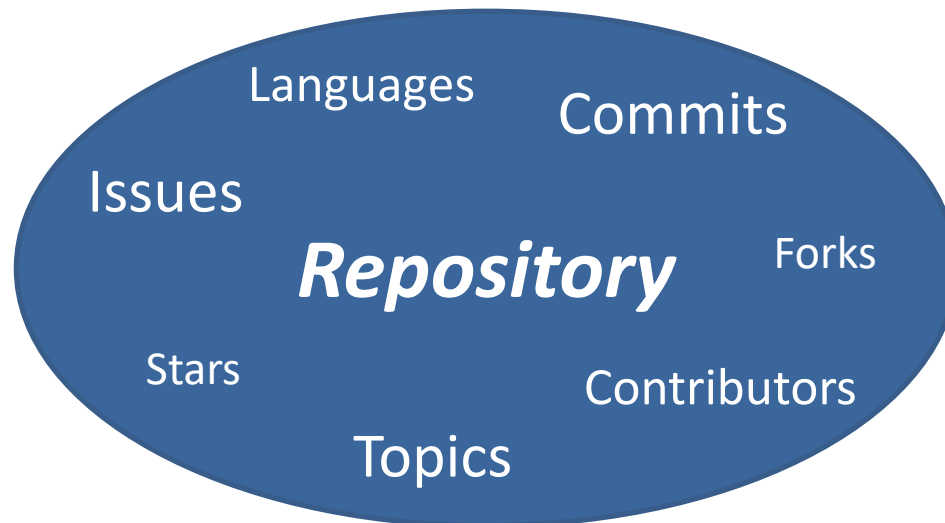
Application Programming Interface (examples)

- <https://api.github.com/repos/AuthEceSoftEng/emb-ntua-workshop>

```
1 {  
2   "id": 352953094,  
3   "node_id": "MDEwOlJlcG9zaXRvcnkzNTI5NTMwOTQ=",  
4   "name": "emb-ntua-workshop",  
5   "full_name": "AuthEceSoftEng/emb-ntua-workshop",  
6   "private": false,  
7   "owner": {  
8     "login": "AuthEceSoftEng",  
9     "id": 6051438,  
10    },  
11   "html_url": "https://github.com/AuthEceSoftEng/emb-ntua-workshop",  
12   "description": "Hands-on machine learning workshop",  
13   "fork": false,  
14   "size": 2910,  
15   "stargazers_count": 11,  
16   "watchers_count": 11,  
17   "language": "Python",  
18   "license": {  
19     "key": "mit",  
20     "name": "MIT License",  
21     "spdx_id": "MIT",  
22     "url": "https://api.github.com/licenses/mit",  
23     "node_id": "MDc6TGljZW5zZTEz",  
24   },  
25   "visibility": "public",  
26   "forks": 8,  
27   "open_issues": 1,  
28   "watchers": 11,  
29 }
```

Ορισμός του Προβλήματος

- Εξαγωγή χρήσιμης γνώσης από αποθετήρια
- Πρόβλεψης χαρακτηριστικών προγραμματιστών
- Πρόβλεψη χαρακτηριστικών έργων λογισμικού



Ενδεικτική Αντιμετώπιση (1/6)

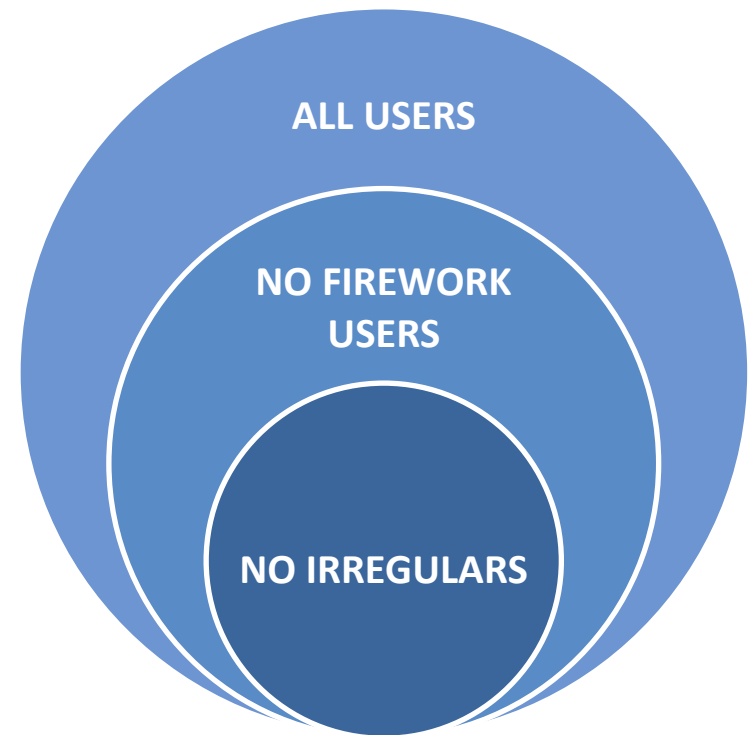
Εύρεση του ρόλου που κατέχει ένας χρήστης εντός του αποθετηρίου

- Data Gathering
 - Συλλογή χαρακτηριστικών που αφορούν τον ρόλο τού χρήστη εντός του αποθετηρίου
 - Αριθμός commits
 - Αριθμός issues που έχει ανοίξει
 - Αριθμός issues που έχει κλείσει
 - Αριθμός comments εντός του αποθετηρίου
 - Αριθμός αρχείων που τροποποιεί σε κάθε commit

Ενδεικτική Αντιμετώπιση (2/6)

Εύρεση του ρόλου που κατέχει ένας χρήστης εντός του αποθετηρίου

- Preprocessing
 - Αφαίρεση outliers, π.χ.
 - Πολύ μεγάλος – υπερβολικά μικρός αριθμός commits/issues
 - Users με abnormal συμπεριφορά
 - Κανονικοποίηση
 - Αξιολόγηση κανονικοποίησης



Ενδεικτική Αντιμετώπιση (3/6)

Εύρεση του ρόλου που κατέχει ένας χρήστης εντός του αποθετηρίου

- **Users Classification**

- SVMs, Artificial Neural Networks, Naïve Bayes, ...

- **Users Clustering**

- K-Means, DBSCAN, Hierarchical, GMMs, ...

- **Evaluation**

- Precision, Recall, ROC, F-measure, ...

- **Evaluation**

- Cohesion, Separation, Silhouette, Visualization, ...

Ενδεικτική Αντιμετώπιση (4/6)

Εύρεση του ρόλου που κατέχει ένας χρήστης εντός του αποθετηρίου

- Ερμηνεία αποτελεσμάτων
 - Αξιολόγηση ρόλων

Feature User	# Commits	# Issues Opened	# Issues Closed	# Comments	# Files per Commit	Role
#1	Many	Few	Many	Few	Many	Dev
#2	Many	Many	Many	Few	Few	Dev
#3	Many	Many	Few	Many	Few	DevOps
#4	Few	Many	Many	Many	Few	Ops
#5	Many	Many	Few	Many	Few	Ops
#6	Many	Many	Few	Many	Many	DevOps

Ενδεικτική Αντιμετώπιση (5/6)

Εύρεση του ρόλου που κατέχει ένας χρήστης εντός του αποθετηρίου

- Περαιτέρω αναλύσεις
 - Αξιολόγηση Dev χρηστών
 - ? Ποια είναι τα χαρακτηριστικά που υποδεικνύουν έναν Dev χρήστη
 - ? Πόσα commits κάνει ένας Dev χρήστης
 - ? Ποια χαρακτηριστικά πρέπει να μεταβληθούν για να θεωρηθεί ένας χρήστης Dev

Ενδεικτική Αντιμετώπιση (5/6)

Εύρεση του ρόλου που κατέχει ένας χρήστης εντός του αποθετηρίου

- Περαιτέρω αναλύσεις
 - Αξιολόγηση Ops χρηστών
 - ? Ποια είναι τα χαρακτηριστικά που υποδεικνύουν έναν Ops χρήστη
 - ? Πόσα commits κάνει ένας Ops χρήστης
 - ? Ποια χαρακτηριστικά πρέπει να μεταβληθούν για να θεωρηθεί ένας χρήστης Ops

Συζήτηση/Ιδέες

- Δημιουργία κατάλληλων research questions
 - Ποιες πληροφορίες μου παρέχει το API;
 - Ποιοι χρήστες έχουν συνδράμει στα πιο γνωστά και δημοφιλή αποθετήρια;
- Ομαδοποίηση ως προς το size και ως προς τα ιδιαίτερα χαρακτηριστικά
 - Small repositories in Java vs Small repositories in Python;
- Διαχείριση δεδομένων (dataset)
 - Δημιουργία χαρακτηριστικών, συνδυασμός χαρακτηριστικών, ...