종합설계프로젝트2
TEAM3

Log Analysis of IT System based on NLP

자연어 처리(NLP) 기반 IT 시스템 로그 분석

|Mentor| 정명훈 이사, Google Korea

|Presenter| 김다혜

|Team Member| 김보근 김나형 김다혜 옥명주 이현지

Check point

1. Project Overview
   - Necessity of Project promotion
   - Overview

2. Project Progress
   - Progress & Issue
   - Java Thread Dump Analysis
   - Write a Thesis

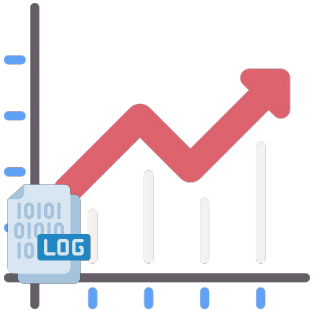
3. Project Schedule and Method
   - Meeting & Schedule Management
   - Gantt Chart

4. Expected Effect
   - Expected Effects
   - Expected Performance

01 Project Overview

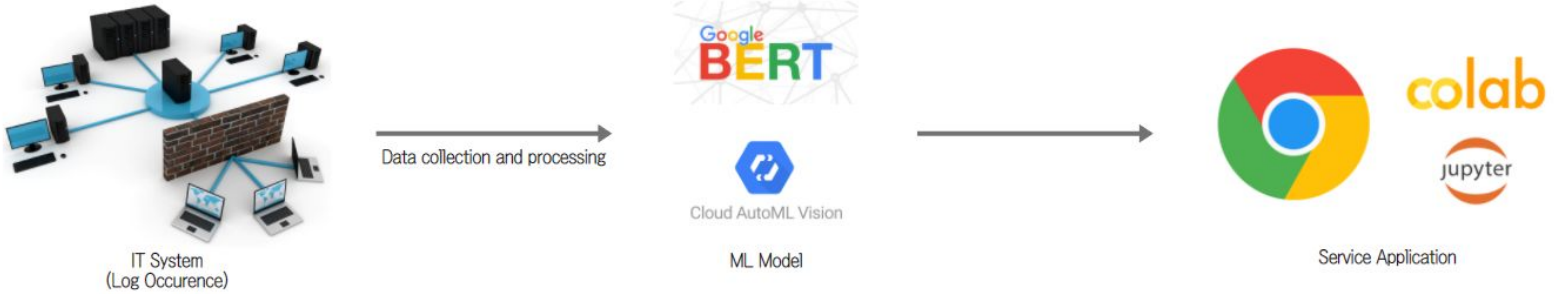
- Overview
- Necessity of Project promotion

01_1 Project Overview - Necessity of Project Promotion

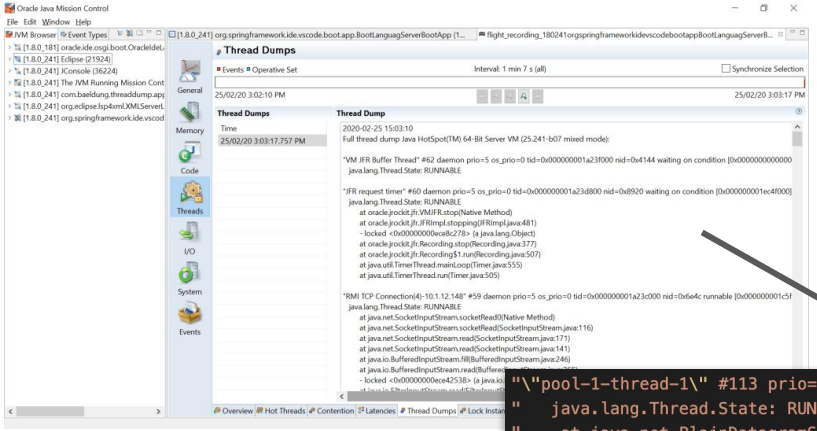


```
"\"pool-1-thread-1\" #113 prio=5 os_prio=0 tid=0x00007fc55c1b2000 nid=0x64c2 runnable [0x00007fc534686000]",
"   java.lang.Thread.State: RUNNABLE",
"    at java.net.PlainDatagramSocketImpl.receive0(Native Method)",
"    - locked <0x00000000c1c90308> (a java.net.PlainDatagramSocketImpl)",
"    at java.net.AbstractPlainDatagramSocketImpl.receive(AbstractPlainDatagramSocketImpl.java:143)",
"    - locked <0x00000000c1c90308> (a java.net.PlainDatagramSocketImpl)",
"    at java.net.DatagramSocket.receive(DatagramSocket.java:812)",
"    - locked <0x00000000c13c19b8> (a java.net.DatagramPacket)",
"    - locked <0x00000000c1c902b8> (a java.net.MulticastSocket)",
"    at org.jboss.modcluster.advertise.impl.AdvertiseListenerImpl$AdvertiseListenerWorker.run(AdvertiseListenerImpl.java:354)",
"    at java.lang.Thread.run(Thread.java:748)",
"",
"   Locked ownable synchronizers:",
"    - None",
""
```

01_1 Project Overview - Overview



01_1 Project Overview - Overview




```
"\"pool-1-thread-1\" #113 prio=5 os_prio=0 tid=0x00007fc55c1b2000 nid=0x64c2 runnable [0x00007fc534686000]",
"   java.lang.Thread.State: RUNNABLE",
"    at java.net.PlainDatagramSocketImpl.receive0(Native Method)",
"    - locked <0x00000000c1c90308> (a java.net.PlainDatagramSocketImpl)",
"    at java.net.AbstractPlainDatagramSocketImpl.receive(AbstractPlainDatagramSocketImpl.java:143)",
"    - locked <0x00000000c1c90308> (a java.net.PlainDatagramSocketImpl)",
"    at java.net.DatagramSocket.receive(DatagramSocket.java:812)",
"    - locked <0x00000000c13c19b8> (a java.net.DatagramPacket)",
"    - locked <0x00000000c1c902b8> (a java.net.MulticastSocket)",
"    at org.jboss.modcluster.advertise.impl.AdvertiseListenerImpl$AdvertiseListenerWorker.run(AdvertiseListenerImpl.java:354)",
"    at java.lang.Thread.run(Thread.java:748)",
"",
"   Locked ownable synchronizers:",
"    - None",
""
```

02 Project Progress

- Progress & Issue
- Java Thread Dump Analysis
- Write a Thesis

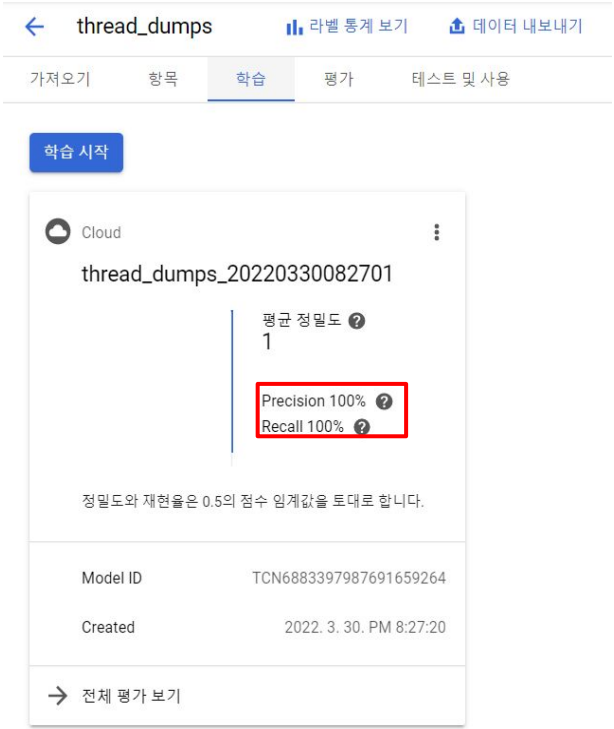
02_1 Progress & Issue

Progress : Gantt Chart

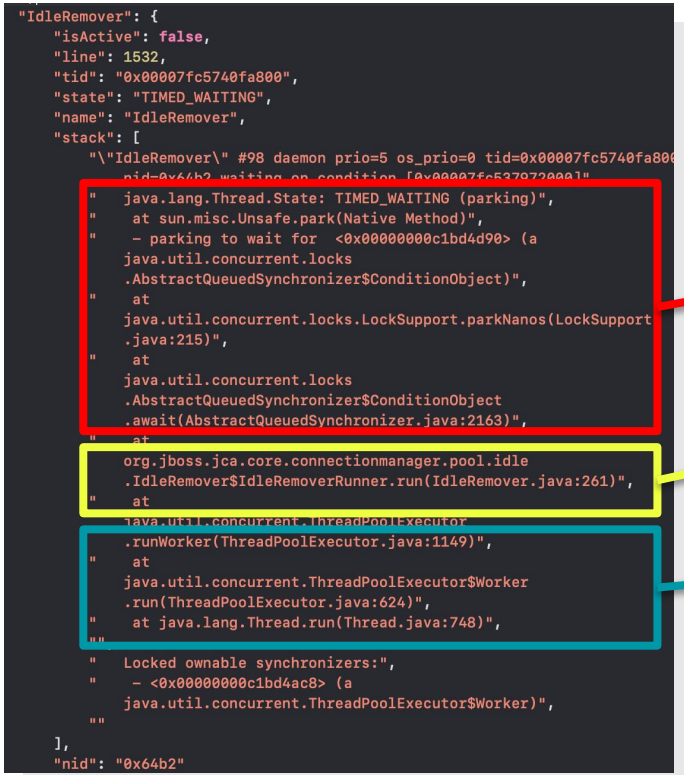
No	Task	담당자	상태	시작일	완료일	2022년 3월				2022년 4월				2022년 5월				2022년 6월	
						2주	3주	4주	5주	1주	2주	3주	4주	1주	2주	3주	4주	1주	2주
계획 및 설계																			
1	주제 선정 및 역할 분담	공동	Done	3/7	3/14														
2	계획서 작성	공동	Done	3/7	3/18														
데이터 분석																			
3	데이터 1차 분석	김다혜, 김보근	Done	3/21	3/27														
논문 작성																			
7	Draft 작성	공동	Doing	3/28	4/3														

02_1 Progress & Issue

Issue



02_2 Java Thread Dump Analysis



```
"IdleRemover": {
    "isActive": false,
    "line": 1532,
    "tid": "0x00007fc5740fa800",
    "state": "TIMED_WAITING",
    "name": "IdleRemover",
    "stack": [
        "\"IdleRemover\" #98 daemon prio=5 os_prio=0 tid=0x00007fc5740fa800
        nid=0x64b2 waiting on condition [0x00007fc537972000]",
        "   java.lang.Thread.State: TIMED_WAITING (parking)",
        "    at sun.misc.Unsafe.park(Native Method)",
        "    - parking to wait for  <0x00000000c1bd4d90> (a
        java.util.concurrent.locks
        .AbstractQueuedSynchronizer$ConditionObject)",
        "    at
        java.util.concurrent.locks.LockSupport.parkNanos(LockSupport
        .java:215)",
        "    at
        java.util.concurrent.locks
        .AbstractQueuedSynchronizer$ConditionObject
        .await(AbstractQueuedSynchronizer.java:2163)",
        "    at
        org.jboss.jca.core.connectionmanager.pool.idle
        .IdleRemover$IdleRemoverRunner.run(IdleRemover.java:261)",
        "    at
        java.util.concurrent.ThreadPoolExecutor
        .runWorker(ThreadPoolExecutor.java:1149)",
        "    at
        java.util.concurrent.ThreadPoolExecutor$Worker
        .run(ThreadPoolExecutor.java:624)",
        "    at java.lang.Thread.run(Thread.java:748)",
        ""
        "   Locked ownable synchronizers:",
        "    - <0x00000000c1bd4ac8> (a
        java.util.concurrent.ThreadPoolExecutor$Worker)",
        ""
    ],
    "nid": "0x64b2"
```

Java Library

Main Function

Base/ WAS code

02_2 Java Thread Dump Analysis

```
"DEADLOCK_TEST-1" daemon prio=6 tid=0x000000000690f800 nid=0x1820 waiting for monitor entry [0x000000000805f000]
   java.lang.Thread.State: BLOCKED (on object monitor)
   at com.nbp.theplatform.threaddump.ThreadDeadLockState$DeadlockThread.goMonitorDeadlock(ThreadDeadLockState.java:197)
   - waiting to lock <0x00000007d58f5e60> (a com.nbp.theplatform.threaddump.ThreadDeadLockState$Monitor)
   at com.nbp.theplatform.threaddump.ThreadDeadLockState$DeadlockThread.monitorOurLock(ThreadDeadLockState.java:182)
   - locked <0x00000007d58f5e48> (a com.nbp.theplatform.threaddump.ThreadDeadLockState$Monitor)
   at com.nbp.theplatform.threaddump.ThreadDeadLockState$DeadlockThread.run(ThreadDeadLockState.java:135)

   Locked ownable synchronizers:
   - None

"DEADLOCK_TEST-2" daemon prio=6 tid=0x0000000006858800 nid=0x17b8 waiting for monitor entry [0x000000000815f000]
   java.lang.Thread.State: BLOCKED (on object monitor)
   at com.nbp.theplatform.threaddump.ThreadDeadLockState$DeadlockThread.goMonitorDeadlock(ThreadDeadLockState.java:197)
   - waiting to lock <0x00000007d58f5e78> (a com.nbp.theplatform.threaddump.ThreadDeadLockState$Monitor)
   at com.nbp.theplatform.threaddump.ThreadDeadLockState$DeadlockThread.monitorOurLock(ThreadDeadLockState.java:182)
   - locked <0x00000007d58f5e60> (a com.nbp.theplatform.threaddump.ThreadDeadLockState$Monitor)
   at com.nbp.theplatform.threaddump.ThreadDeadLockState$DeadlockThread.run(ThreadDeadLockState.java:135)

   Locked ownable synchronizers:
   - None
```

Thread에 대한 정보가 출력되며, 현재 상태는 waiting for monitor entry 임

<0x00000007d58f5e60> lock에 대해 waiting 상태임을 알 수 있음

<0x000000007d58f5e46> lock에 획득하여 수행하는 것을 알 수 있음

- 현재 Thread의 수행 상태에 대한 Stack trace 정보
- 각 Stack별 수행되고 있는 class와 메소드, 그리고 소스의 라인이 출력됨
- Stack trace는 bottom-> top 방향

02_2 Java Thread Dump Analysis

상태	설명
allocated	Thread가 생성되기 위해 메모리 할당된 상태
initialized	Thread가 초기화 된 상태
runnable	Thread가 Java Virtual Machine 내에서 수행 중인 상태
waiting for monitor entry	Thread가 Monitor lock을 획득하기 위해 대기중인 상태
waiting on condition	Thread가 조건 변수(Condition Variable)에 의해 대기 중인 상태
in Object.wait()	Thread가 Object.wait()으로 인해 대기중인 상태
sleeping	Thread가 I/O등에 의해 대기 중에 sleep 상태가 된 경우

▲<Thread 상태 Sun java 1.5v>

▼<Thread 상태 Sun java 1.6v>

상태	설명
NEW	Thread가 생성되기 위해 메모리 할당된 상태
RUNNABLE	Thread가 Java Virtual Machine 내에서 수행 중인 상태
BLOCKED	Thread가 Monitor lock을 획득하기 위해 대기중인 상태
WAITING	Thread가 특정 작업을 위해 다른 Thread를 무기한 대기하는 상태
TIMED_WAITING	Thread가 특정 작업을 위해 정해진 시간만큼 다른 Thread를 대기하는 상태
TERMINATED	Thread가 exited 된 상태

02_3 Data Processing

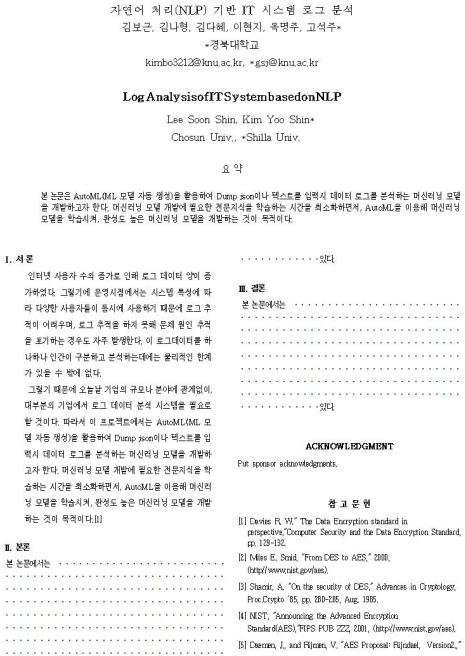
ThreadDump_뉴발란스_SystemLog남길때_Lock걸려느림_parsed.json

ThreadDump_뉴발란스_SystemLog남길때_Lock걸려느림_parsed.json 〉 No Selection

```
19    "Exec Stream Pumper": {
20        "isActive": false,
21        "line": 10,
22        "tid": "0x00007f95c80a8000",
23        "state": "RUNNABLE",
24        "name": "Exec Stream Pumper",
25        "stack": [
26            "\"Exec Stream Pumper\" #496 daemon prio=10 os_prio=0
                tid=0x00007f95c80a8000 nid=0xeea runnable [0x00007f95763a3000]",
27            "   java.lang.Thread.State: RUNNABLE",
28            "    at java.io.FileInputStream.readBytes(Native Method)",
29            "    at java.io.FileInputStream.read(FileInputStream.java:255)",
30            "    at
                java.io.BufferedInputStream.fill(BufferedInputStream.java:246)",
31            "    at
                java.io.BufferedInputStream.read1(BufferedInputStream.java:286)",
32            "    at
                java.io.BufferedInputStream.read(BufferedInputStream.java:345)",
33            "    - locked <0x00000007e9b43788> (a
                java.lang.UNIXProcess$ProcessPipeInputStream)",
34            "    at java.io.FilterInputStream.read(FilterInputStream.java:107)",
35            "    at
                com.opennaru.khan.agent.org.apache.commons.exec.StreamPumper
                .run(StreamPumper.java:107)",
36            "    at java.lang.Thread.run(Thread.java:748)",
37            "",
38            "   Locked ownable synchronizers:",
39            "    - None",
40            ""
41        ],
42        "nid": "0xeea"
```

```python
if key != 'created':
    thread = jsonData[key]
    stack = '\n'.join(thread['stack'])
    file_st = open('dumps/threads/{0:07d}.txt'.format(idx), 'w')
    idx = idx + 1
    file_st.write(stack)
    file_st.close()
    # thread의 이름에서 영어, 띄워쓰기만 남기고 나머지(숫자, 기호 등) 삭제
    name = re.sub('[^ a-zA-Z]','',thread['name'])
    stack = re.sub('^\"[^\"]*\"', name, stack)

    #스레드의 고유번호(#1234), 포인터 저장 위치(형식 [0x000], <0x000>) 삭제
    stack = re.sub('#[0-9]+|\[[^)]*?\]|\<[^)]*?\>', '', stack)

    #공백, .(온점), (, ), ,(반점) 기준으로 split
    words = re.split(' |\.|\(|\)|\n|,', stack)

    dehydrated_words = []
    for word in words:
        #비워있는 word 스킵
        if word == '' : continue

        #java: 형식으로 되어있는 코드 행 삭제, prio=, os_prio, tid, nid 가 포함된 단어 삭제
        #javax, at, java, com, org 처럼 의미없이 자주 출현하는 단어 삭제(불용어 처리)
        if word.find('java:') == -1 and word.find('prio=') == -1 and word.find('id=') == -1 and \
        word != 'javax' and word != 'at' and word != 'java' and word != 'com' and word != 'org':
            dehydrated_words.append(word)
```

02_4 Write a Thesis



자연어 처리(NLP) 기반 IT 시스템 로그 분석

김보근, 김나형, 김다혜, 이현지, 옥명주, 고석주*

*경북대학교

kimbo3212@knu.ac.kr, *gsj@knu.ac.kr

Log Analysis of IT System based on NLP

Lee Soon Shin, Kim Yoo Shin*

Chosun Univ., *Shilla Univ.

요 약

본 논문은 AutoML/ML 모델 자동 생성)을 활용하여 Dump json이나 텍스트를 입력시 데이터 로그를 분석하는 머신러닝 모델을 개발하고자 한다. 머신러닝 모델 개발에 필요한 전문지식을 학습하는 시간을 최소화하면서, AutoML을 이용해 머신러닝 모델을 학습시켜, 완성도 높은 머신러닝 모델을 개발하는 것이 목적이다.

I. 서 론

인터넷 사용자 수의 증가로 인해 로그 데이터 양이 증가하였다. 그렇기에 운영시점에서는 시스템 특성에 따라 다양한 사용자들이 등시에 사용하기 때문에 로그 추적이 어려우며, 로그 추적을 하지 못해 문제 원인 추적을 포기하는 경우도 자주 발생한다. 이 로그데이터를 하나하나 인간이 구분하고 분석하는데에는 물리적인 한계가 있을 수 밖에 없다.

그렇기 때문에 오늘날 기업의 규모나 분야에 관계없이, 대부분의 기업에서 로그 데이터 분석 시스템을 필요로 할 것이다. 따라서 이 프로젝트에서는 AutoML/ML 모델 자동 생성)을 활용하여 Dump json이나 텍스트를 입력시 데이터 로그를 분석하는 머신러닝 모델을 개발하고자 한다. 머신러닝 모델 개발에 필요한 전문지식을 학습하는 시간을 최소화하면서, AutoML을 이용해 머신러닝 모델을 학습시켜, 완성도 높은 머신러닝 모델을 개발하는 것이 목적이다.[1]

II. 본론

본 논문에서는 . . . . . 있다.

III. 결론

본 논문에서는 . . . . . 있다.

ACKNOWLEDGMENT

Put sponsor acknowledgments.

참 고 문 헌

[1] Davies R. W." The Data Encryption standard in perspective,"Computer Security and the Data Encryption Standard, pp. 129-132.

[2] Miles E. Smid, "From DES to AES," 2000, (http://www.nist.gov/aes).

[3] Shamir, A. "On the security of DES," Advances in Cryptology, Proc.Crypto '85, pp. 280-285, Aug. 1985.

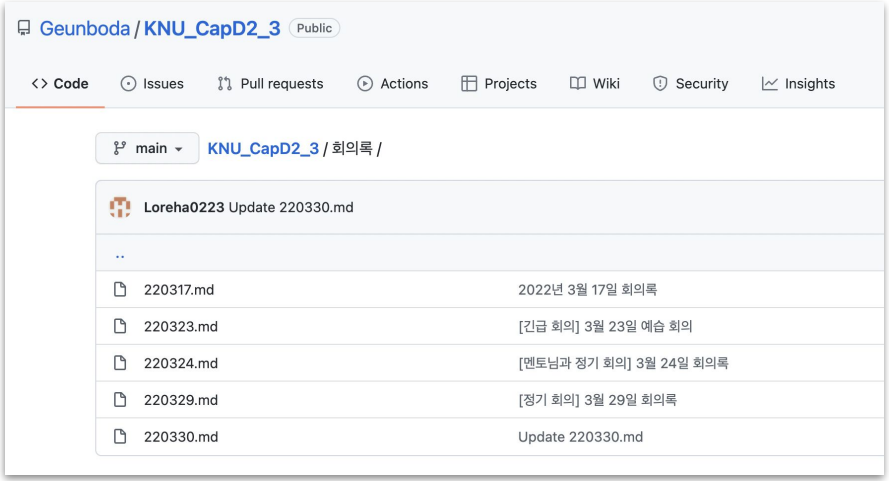
[4] NIST, "Announcing the Advanced Encryption Standard(AES),"FIPS PUB 222, 2001, (http://www.nist.gov/aes).

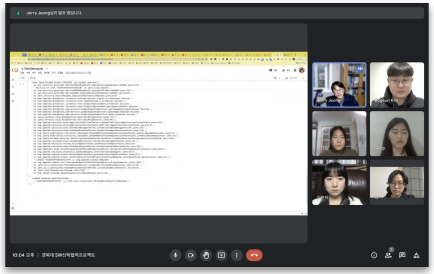
[5] Daemen, J., and Rijmen, V. "AES Proposal: Rijndael, Version2.,"

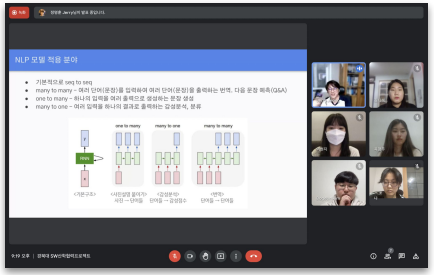
03 Project Schedule and Method

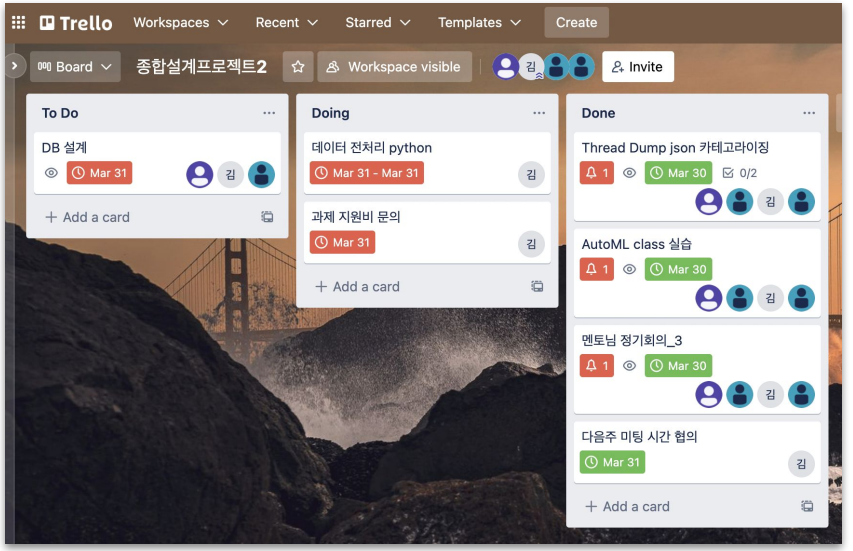
- Meeting & Schedule Management
- Gantt Chart

03_1 Project Schedule & Method - Meeting & Schedule Management









Github Google Meet Trello

03_1 Project Schedule & Method - Meeting & Schedule Management

오픈마루 - 경북대 NLP 기반 로그 분석 프로젝트

Google Calendar에서 보기

시간 2022년 4월 5일 (화) 오후 3시 – 오후 4시 (KST)

참석자 javalove93@gmail.com, rhdlwmal123@gmail.com, kimbo3212@gmail.com, okmj0731@gmail.com...

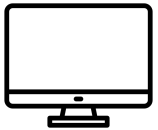
03_2 Project Schedule & Method - Gantt Chart

1.  : 

Model

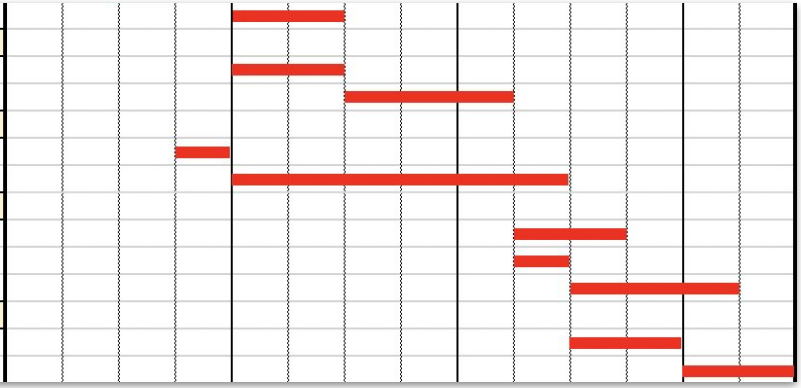
2. 

Paper

3. 

API / UI

4	데이터 2차 분석	김나형, 이현지, 옥명주	to Do	4/4	4/17
모델링					
5	모델 개발(AutoML)	김다혜, 김보근	to Do	4/4	4/17
6	모델 개발(BERT)	김나형, 이현지, 옥명주	to Do	4/18	5/8
논문 작성					
7	Draft 작성	공동	Doing	3/28	4/3
8	최종 논문 작성	공동	to Do	4/4	5/12
응용 프로그램 개발					
9	API 개발	이현지	to Do	5/9	5/22
10	UI / UX 설계 및 기획	김나형, 김다혜	to Do	5/9	5/15
11	웹 애플리케이션 개발	김나형, 김다혜	to Do	5/16	6/5
연동					
12	업체 시스템과 모델 연동	김보근, 옥명주	to Do	5/16	5/29
13	Testing 및 수정	공동	to Do	5/30	6/12



04 Expected Effects

- Realization
- Expected Performance

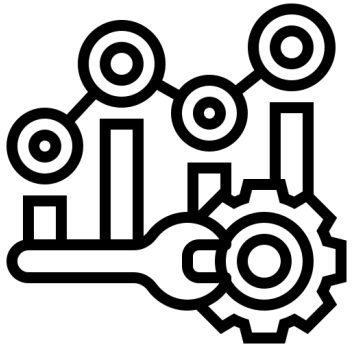
04_1 Expected Effects - Realization

- Difficulty of Data Analysis

```
    },
    "default task-97": {
        "isActive": true,
        "line": 4323,
        "tid": "0x0000000003b1a000",
        "state": "TIMED_WAITING",
        "name": "default task-97",
        "url": "/",
        "duration": 20175,
        "cpuTime": 3.98,
        "stack": [
            "\"default task-97\" #314 prio=5 os_prio=0 tid=0x0000000003b1a000
                nid=0xc24 runnable [0x00002ac1d2dd0000]",
            "   java.lang.Thread.State: TIMED_WAITING (parking)",
            "    at sun.misc.Unsafe.park(Native Method)",
            "    - parking to wait for  <0x00000007414aefa8> (a
            java.util.concurrent.locks
            .AbstractQueuedSynchronizer$ConditionObject)",
            "    at
            java.util.concurrent.locks.LockSupport.parkNanos(LockSupport
            .java:215)",
            "    at
            java.util.concurrent.locks
            .AbstractQueuedSynchronizer$ConditionObject
            .awaitNanos(AbstractQueuedSynchronizer.java:2078)",
            "    at
            java.util.concurrent.ArrayBlockingQueue.poll(ArrayBlockingQueue
            .java:418)",
            "    at
            org.infinispan.affinity.impl.KeyAffinityServiceImpl
            .getKeyForAddress(KeyAffinityServiceImpl.java:140)",
            "    at
            org.wildfly.clustering.web.infinispan.AffinityIdentifierFactory
            .createIdentifier(AffinityIdentifierFactory.java:51)",
            "    at
            org.wildfly.clustering.web.infinispan.session
            .InfinispanSessionManager
            .createIdentifier(InfinispanSessionManager.java:244)",
```

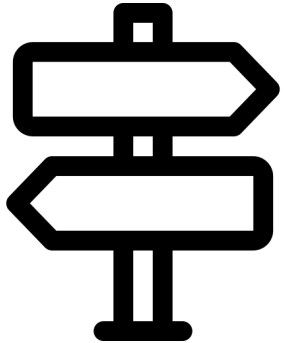
Thread Dump Data

04_2 Expected Effects - Expected Performance



Optimization

To Enterprise



Direction

To Developer

Satisfaction

To Consumer

Thank you
for
your attention!

From. TEAM3