

GENERAL ASSEMBLY DSI CAPSTONE PROJECT

WEEK 2 CHECK-IN

AARON KIM

PROGRESS REPORT

- DECISION :
USER ACTIVITY DEPENDENT PASSWORD
GENERATION
- PROJECT CODENAME: SG1

PROJECT OBJECTIVES

- REDUCE/ELIMINATE “FORGOT PASSWORD” REQUESTS
- REPLACE CURRENT CONCEPT OF A “PASSWORD”
- IMPROVE SECURITY BY “TRAINING” SECURITY TO THE USER

RISKS

- RISK: USER UNABLE TO REMEMBER SOME PAST BEHAVIORS
IE: NOT SOBER DURING TIME OF QUERIED ACTIVITY)
- RISK: QUESTIONS ARE PHRASED INCORRECTLY
EG: MODEL GENERATES QUESTION PHRASE THAT IS NOT VALID
FOR THE DESIRED RESPONSE
IE: GIVEN A NAME OF A CITY “ANAHEIM,” THE MODEL
INCORRECTLY GENERATES A QUESTION THAT IS NOT
RELEVANT TO A REFERENCE TO A LOCATION
- RISK: DATA SOURCES NOT AVAILABLE FOR PASSWORD TRAINING
IE: No SOCIAL MEDIA (FACEBOOK)/GOOGLE/iCLOUD/EMAIL ACCOUNT
AVAILABLE OR NOT ACCOUNT INFORMATION ACCESS IS GRANTED

ASSUMPTIONS

- ASSUMPTION: USER CAN RECALL INFORMATION ACCURATELY
- ASSUMPTION: AT LEAST ONE SOURCE OF USER PATTERN TRAINABLE DATA IS AVAILABLE

GOALS/CRITERIA

GOAL:

1. BASED ON USER GENERATED PATTERN DATA, GENERATE A LIST OF SECURITY QUESTIONS.
2. CREATE API THAT ONLY PROVIDES CHALLENGE TEXT WHEN REQUESTED AND RESPONDS TO AUTHENTICATION.
3. KEEP DATA ANONYMIZED – GOAL IS NOT TO IDENTIFY THE USER, JUST VERIFY THE LOGIN ATTEMPT AGAINST WHAT IS KNOWN OF THE AUTHORIZED USERS' BEHAVIOR.

GOALS/CRITERIA

PRIMARY SUCCESS CRITERIA:

1. GENERATE "NATURAL" WORDED QUESTIONS BASED OFF USER DATA
2. API AUTHENTICATION SUCCESS

SECONDARY SUCCESS CRITERIA:

1. "LEARN" WHICH TYPES OF QUESTIONS THE USER TENDS TO RECALL MORE ACCURATELY TO NARROW QUESTION GENERATING DATASET
2. ISOLATE AND ENCRYPT SG1 MODULE

DATA COLLECTION/CLEANING

DATA SOURCES

1. FACEBOOK (POSTS, CHECK-INS, FRIEND ACTIVITY – LIKES, EVENT CALENDAR)
2. GOOGLE (CALENDAR, BROWSING HISTORY, EMAIL, YOUTUBE, MAPS HISTORY)
3. ICLOUD (CALENDAR, CONTACTS, CALL/SMS HISTORY, BROWSER HISTORY)

DATA COLLECTION/CLEANING

PACKAGES

1. TfidfVectorizer, LabelEncoder, NLTK (POS_TAG, WordPunctTokenizer)
2. EncryptedPickle/Pickle
3. Facebook, Google, and iCloud Libraries (Mixed Python/PHP)
4. Geotext

PREPROCESSING

1. Tfidf – Vectorize words for NLP dictionary
2. POS_TAG – Get sentence structure breakdowns for analysis and question structure/type generation

DATA DICTIONARY SAMPLE - FACEBOOK

```
content
{
  "data": [
    {
      "actions": [
        {"link": "https://www.facebook.com/1512072122398732/posts/1775498886056053",
         "name": "Comment"},
        {"link": "https://www.facebook.com/1512072122398732/posts/1775498886056053",
         "name": "Like"}],
      "created_time": "2016-10-30T21:00:54+0000",
      [REDACTED]
      "icon": "https://www.facebook.com/images/icons/photo.gif",
      "id": "1512072122398732_1775498886056053",
      "is_expired": False,
      "is_hidden": False,
      "likes": {
        "data": [
          {"id": "506019182837045", "name": [REDACTED],
           "id": "10204925121235520", "name": [REDACTED],
           "id": "1012946735446628", "name": [REDACTED],
           "id": "10152097493151723", "name": [REDACTED],
           "id": "10204306259685191", "name": [REDACTED],
           "id": "744023000008", "name": [REDACTED],
           "id": "928789973805726", "name": [REDACTED],
           "id": "928789973805726", "name": [REDACTED]},
          {"data": [
            {"cursors": {"after": "OTI4Nzg50TczODA1NzI2",
             "before": "NTA2MDE5MTgyODM3MDQ1"}]}},
        "link": "[REDACTED]/photos/a.1523674677905143.1073741828.1512072122398732/1775496049389670/?type=3",
        "message": "Congratulations to the Orange County Conservation Corps on your ribbon cutting celebration for your new facility in Anaheim!",
        "name": "Photos from [REDACTED]",
        "object_id": "1775496049389670",
        "picture": "https://scontent.xx.fbcdn.net/v/t1.0-0/s130x130/14915597_1775496049389670_3335237246371269085_n.jpg?oh=d3326bcfad0c95bf2b957e4887118165&oe=58A71C8E",
        "privacy": {"allow": "u''",
        "deny": "u''",
        "description": "u''",
        "friends": "u''",
        "value": "u''"},
        "shares": {"count": 2},
        "status_type": "added photos",
        "story": "[REDACTED] added 8 new photos.",
        "story_tags": {"0": [{"id": "1512072122398732",
          "length": 23,
          "name": [REDACTED],
          "offset": 0,
          "type": "page"}]},
        "type": "photo",
        "updated_time": "2016-10-30T21:00:54+0000"}]
```

SG1 FLOW (v1)

