

We are building an application that receives a text file/string from a user. The file/string acts as the text message sent to the end user. This text portion is then parsed by an interface. The interface parses the message and compares them to a collection/DB of key phrases. This DB of key phrases are typical key phrases that are already deemed threatening, harassment, abuse, etc. Based on the number of hits in this collection, the application sends a notification/message to the user.

Definitions

- **DB-database** - not necessarily a database, for now it is just a collection of all likely harmful phrases and keywords.
- **Key phrase-stored** in the DB - these phrases are seen as offensive or potentially offensive
- **Classification** - categorization of keywords; this breaks down into 3 categories: **discrimination, threat, harrasment**
- **Severity** - level of harm scaled: 1-5
- **JSON object** - where we store key phrase, classification, severity

Key phrases

- Stores phrases deemed as offensive only
 - Only stores these for speed inside DB
- Stores classification of these phrases → categorization
- Stores severity level of these phrases → 1-5 scale for now

Interface

- Retrieves a DB from global server
 - This DB will be kept as a local version only
 - Can have specific keywords
- Parses string
 - Changes string to lowercase only
 - Removes punctuation
 - Each word is tested against the DB
 - If there is a hit in the DB, add it to a collection
 - If there are enough hits in a time period, send a notification
- Admin addition
 - Admin can add key phrases and severity level
- User addition
 - Can suggest words to be added to DB
 - Can determine the severity of these words as well
 - These additions will be stored locally on the user's device

- Will be added to a different DB
 - If this phrase becomes trending/enough users suggest this phrase, phrase will be added to the main DB
- User deletion
 - Can say which words are not harmful to them
 - Severity of these words will be placed at 0
 - Or we can remove them completely from the local database

Notification

- Reactionary
- Receives JSON objs from Interface as well as beginning and ending timestamps
- Sends msg to user about alarming phrases and time stamp
 - If user deems this legitimate
 - Notif sends supportive message and list of possible resources, advice

Hotlines/Resources in response to the text:

Discrimination

<https://www.fordfoundation.org/the-latest/news/a-new-hotline-for-anonymous-harassment-and-discrimination-complaints/>

<https://www.sterlingattorneys.com/blog/2017/08/first-steps-to-take-if-you-are-being-discriminated-against.shtml>

<https://www.chwilliamsllaw.com/what-are-signs-that-i-may-be-being-discriminated-against-at-work/>

Threat

<https://www.runawayhelpline.org.uk/stages-of-running/im-being-hurt-abused-or-threatened/>

https://blogs.findlaw.com/law_and_life/2017/02/what-to-do-if-someone-threatens-you.html

Harassment

<https://www.thehotline.org/2016/04/21/reporting-to-police-options-tips-for-being-prepared/>

http://www.plea.org/legal_resources/?a=303

<https://www.beliefnet.com/wellness/galleries/7-ways-to-deal-with-harassment-of-any-kind.aspx>