

Shahjalal University of Science and Technology
Department of Computer Science and Engineering
CSE 400



Sachetan: A Crowdsourc Based Personal Safety Application
Shadman Habib
2013331026

Ozayer Islam
2013331062

Supervisor: **Sheikh Nabil Mohammad**

Assistant Professor
Department of Computer Science and Engineering

19th March, 2018

Sachetan: A Crowdsource Based Personal Safety Application



A Thesis/Project submitted to the Department of Computer Science and Engineering, Shahjalal University of Science and Technology, in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering.

Shadman Habib
201333102

Ozayer Islam
2013331062

Supervisor: **Sheikh Nabil Mohammad**

Assistant Professor

Department of Computer Science and Engineering

19th March, 2018

Recommendation Letter from Thesis Supervisor

The thesis

entitled “Sachetan: A Crowdsourced Based Personal Safety Application”

submitted by the student

1. Shadman Habib
2. Ozayer Islam

is a record of research work carried out under my supervision and I, hereby, approve that

the report be submitted in partial fulfillment of the requirements for the award of his/her/their Bachelor Degree(s).

Signature of the Supervisor: _____

Name of the Supervisor : Sheikh Nabil Mohammad

Date: 19th March, 2018

Certificate of Acceptance of the Thesis

The thesis

entitled “Sachetan: A Crowdsourced Based Personal Safety Application”

submitted by the student

1. Shadman Habib
2. Ozayer Islam

on 19th March, 2018

is, hereby, accepted as the partial fulfillment of the requirements for the award of his/her/their Bachelor Degree(s).

Head of the Department
Dr Mohammad Reza Selim
Professor
Department of Computer
Science and Engineering

Chairman, Exam. Committee
M. Jahirul Islam, PhD, PEng
Professor
Department of Computer
Science and Engineering

Supervisor
Sheikh Nabil Mohammad
Assistant Professor
Department of Computer
Science and Engineering

Abstract:

One of the main Barriers of development in any country is crime. Specially in a developing country like Bangladesh, crime and accidents are a major issue and concern. Nowadays mugging, theft, robbery, hijacking, pickpocketing, road accidents, sexual harassments etc. are pretty common in our country. These kinds of crimes are not only barring the development process, these are also taking the country one step backwards. Whereas there is no straightforward solution to these problems, we have worked on a project named “Sachetan”, Which we believe will reduce these occurrences and help the people to stay alert about these. It is an android based mobile application that points out the location of past incidents, frequency of incidents in a specific place and also suggest alternate path to any destination rather than the frequent crime zone. All these features are generated based on the user data provided through this app and from the newspaper crime articles. We have tried to find the relations between these crimes, their patterns and visualized them in graphical representation through this app.

Keyword: Mobile app, Safety, Crowdsourcing, Map, Social Awareness

ACKNOWLEDGEMENT

We would like to thank our supervisor Sheikh Nabil Mohammad for his instruction and guidance during this work. He's has been really patient in teaching me and it wouldn't have been possible without his help.

We would also like to thank my senior Sakhawat Hossain Saimon, Mustafizur Rahman Nebir, Mithun Das, Nishikanto Sarkar Shimul and Tanvir Islam Preom, for selflessly aiding us with their knowledge.

19th March, 2018
Shadman Habib
Ozayer Islam

Contents

Introduction:	1
Details on the Related work:	2
Limitation of these works:	3
Sachetan: A new Solution	3
Crowdsourcing:	4
What is Crowdsourcing?	4
Examples of Crowdsourcing	4
Types of Crowdsourcing	5
Different Types of Crowdsourcing	5
Crowdsourcing Design:	5
Crowdfunding	6
Microtasks	6
Open Innovation:	6
Advantages	7
Disadvantages	7
The Difference Between Crowdsourcing and Crowdfunding	7
Challenges in Crowdsourcing	8
Solution to these challenges	8
About our app:	10
Understanding the Sachetan app:	10
Front end:	10
Back end:	10
Our work on the app:	11
Cleaning the app	11
Additional Features:	14
Current Condition of the app:	15
Server Architecture	18
NoSQL Database	18
Java Spring Framework	19
Amazon Web Service	19
Methodology:	20
News Data Collection	20

Incident Types	20
Collection of News Article Data	21
Extraction of Date and Location Information from Articles	22
Future Work:	22
Criteria for Pattern generation	23
Data Collection	23
Filtering Technique Review:	23
User Notification	24
Goal:	24
Conclusion:	25
References	26

Introduction:

Bangladesh is a developing country and its growth rate is quite remarkable over the years. Like any other developing country there are some hindrance towards the developments. While we are talking about the development of our country, criminals are doing their part to undone the good things. Recently, the crime rate is rising very quickly in every part of our country. Besides mugging, hijacking, sexual harassments, theft in public or remote places, recent news suggest that not even our own residence is safe anymore from becoming a headline! A few months back, reports came that some random thieves entered the 2nd and 3rd floor of SUST Ladies hall and robbed 4 laptops and 2 mobile phones! This news tells us how terrible the situation is in our country! Our government and different organizations are trying their best to mark the problems and find solutions of these. To add with their efforts, we are also trying to develop a platform from which a greater good is expected to come. Digitalization is a process that our country is going through for a better future and with the help of data mining, our app should be able to address a solution of these crime. To gather the information of the crimes is a big challenge which we are doing through this app and then analyze them for finding out patterns, frequency and other interesting stuffs.

Literature Review:

The fightback against these crimes are not new. Several efforts have been made earlier to tackle these. Several researches have been conducted to address the severity and harms of harassments and even legal remedies are there for that.

Different mobile apps are also developed to help these victims by sending signal or help message to security monitor center. Among these apps, SOS Response[3], Not your Baby App[4], Circleof6[5], Safetipin[7], Abhaya[8], OnWatch[9], SheSecure[10] etc. are mention worthy. In some of these apps, GPS location is also provided. This approach in some cases mitigates the harassment, but not necessarily prevents it. Therefore scholars were interested in analyzing crime hot spots[11] and burning times which are becoming a major component of the work of criminologists, crime analysts and crime prevention practitioners from the past decade. As social networks are the common platform of people to freely provide private information, for instance on their current situation, opinions etc. others try to use this opportunity to improve the prediction of crimes[12].

With the help of crowdsourcing[15] some attempts were made so that people can both provide crime-related information and be aware about their locations.

Hollaback![13], a social movement organization mainly works for raising awareness about street harassments against women. Protibadi[1] is also developed for sharing street harassment experiences for Bangladeshi women. It allows a user to report the location along with a description. HarassMap[14] collects and summarizes different types of harassments in a map mainly in Egypt. Another Bangladeshi movement, SafeStreet[2] is mentionable which empowers women in public places against sexual harassment. They also enable a woman to find a safe path that has less harassment hazard, at any point of time via their app.

Details on the Related work:

There are quite a number of works done before this app. Among those “**Protibadi**” and “**Safestreet**” are mention worthy. We read all the papers related to these works thoroughly and learnt some key features of these kinds of apps. Both these apps are focused on helping women against sexual harassments in public places. Those who faced harassments can share their experiences anonymously through the apps. Below are the special features that these apps had:

1. In the “**Safestreet**” app, background running is also enabled in order to capture real time harassments really quick. This way he/she can add report on that incident later.

2. In the “**Protibadi**” app, some additional features like web interfacing and SMS based reporting is also available which is great for crowdsourcing. These features brought some positive reviews from the female users according to the publisher.

Limitation of these works:

As a pioneering work, both of the apps did a marvelous job. But they also have some limitations which encouraged us to develop our very own app regarding crime. Among the limitations, the biggest one is that the apps are not available anymore! This is really bad to see that these apps are not available in Google Play Store. Even we started on this project 6 months earlier, the apps were not available even back then! We could only read the papers and learnt from them only!

Beside this, the app is focused on sexual harassments only to safe the women from facing the heinous crime. They do not deal with other daily crimes like mugging, hijacking etc. In our country these crimes occur parallely. So, a better solution was just a need of time.

Sachetan: A new Solution

“Sachetan” app provides a platform where people can see the crime location as well as report any crime location. Our main purpose right now is to gather report on crime through this app. This is a very well-known process, Crowdsourcing is now quite a familiar word in data collection method. Our app collects those crime related data using crowdsourcing for future analysis. The whole crowdsourcing is a big process and below are some key notes about it.

Crowdsourcing:

Our project Sachetan relies on data. Data collection is the first and foremost part of our research. Huge data is needed to process data mining and generate new information. We introduced the sources of data. News data can be extracted from different newspapers by using crawler. But these data are not enough as the quantity is less and also all news of the negative incidents occurring in different locations can't be found in only newspaper. Here comes the need of crowdsourcing, by which we can get user data. It is beneficial because user living in a specific place can give a rare information that the newspapers might not have. So, crowdsourcing plays a vital role in our research.

What is Crowdsourcing?

Crowdsourcing involves obtaining work, information or opinions from a large group of people who submit their data via the Internet, social media and smartphone apps. People engaged in crowdsourcing sometimes work as paid freelancers, while others perform small tasks on a voluntary basis. For example, traffic apps encourage drivers to report accidents and other roadway incidents to provide real-time updated information to app users.

Examples of Crowdsourcing

Crowdsourcing is a vast used approach to many companies. Such as, Uber, which pairs available drivers with people who need rides, is an example of crowdsourced transportation.

Crowd opinions matter when companies desire to design new items. Companies can reach millions of consumers through social media, rather than rely on small focus groups, ensuring that the business obtains opinions from a variety of cultural backgrounds.

While crowdsourcing often involves breaking up a big job, businesses sometimes use crowdsourcing to assess how multiple people perform at the same job. For instance, if a company wants a new logo, it can have dozens of graphic designers

assemble samples for a small fee. The company can then pick a favorite and pay for a more complete logo package.

Types of Crowdsourcing

According to an article [4] the literature on crowdsourcing can be categorized into four parts. They are - application, algorithm, performance and dataset.

Application: Generally, there are two groups of users in a crowdsourcing site requesters and user. A list of available tasks is exhibited with associating time and period. Maintaining this condition, workers compete to give the best solution. Again, crowdsourcing application can be categorized into four types. They are voting system, game, information sharing system and creative system.

Algorithm: A crowdsourcing system design can be easily formalized by an algorithm. Solution of some theoretical challenging problems are provided by internet users. They can be game theory or any algorithm base problem.

Dataset: If a large dataset is needed for further research then crowdsourcing is a good solution to collect huge amount of data in short time. For example, in case of making a corpus huge amount of word is needed. We can offer users to provide word. In some case we can give them special offer.

Performance: It provides a chance to make an existing technique better.

Different Types of Crowdsourcing

Crowdsourcing Design:

When someone is looking for a logo design, he can tell a crowd of designers what he wants, how much he will pay, and the deadline of his work. In this case, interested designers will create a finished logo specifically for him. The requester will receive different finished logo designs more than he expects, and he can choose the best one. By doing design this way, crowdsourcing actually increases the quality & decreases the price, compared to online freelancing.

This type of crowdsourcing can also be used to get designs for furniture, fashion, advertisements, video & product design. Just about anything that can be designed can be crowdsourced.

Crowdfunding: This type of crowdsourcing involves asking a crowd of people to donate money to your project. For example, if someone wants to raise \$50,000 to pay for studio time to record a new CD, crowdfunding can help him raise that money. He can easily find a crowdfunding platform, set the goal amount, deadline, and any rewards offered to donors. He must raise 100% of his goal before the deadline, or all the donations are returned to the donors. Deadlines are typically less than 60 days.

Crowdfunding is mostly used by charities, artists & start-ups to raise money for projects such as filming a documentary, manufacturing an iPod watch, cancer research, or seed money.

Microtasks: Micro tasking involves breaking work up into tiny tasks and sending the work to a crowd of people. If someone has 5,000 photos on his website that need captions, he can ask 5,000 individual people to each add a caption to one photo. Break up the work and decide the payment for each completed task. With micro tasking, one can expect to see results within minutes. Micro tasking can involve tasks such as scanning images, proofreading, database correction and transcribing audio files.

Work is done faster, cheaper, and usually with less errors (when validation systems are in place). Additionally, microtasks can often be performed by people in less fortunate countries, including those with SMS capabilities but without computers.

Open Innovation:

If someone is confused to begin with an idea for a business opportunity, whether it's product design or perhaps a marketing firm, crowdsourcing can help through open innovation. With the help of Open innovation people from all aspects of business such as investors, designers, inventors, and marketers can collaborate into a functional profit-making reality. This can be done either through a dedicated web platform to gain outside perspective or used with only internal employees.

Open innovation brings together people from different parts of the world and different sectors of business to work together on a project. This is effectively a collection of different fields and levels of expertise that would not otherwise be available to any budding entrepreneur. It also elevates previously considered uninvolved parties, such as investors, to roll up their sleeves and impart their knowledge, essentially becoming more than just a cash cow!

Advantages

The advantages of crowdsourcing include cost savings, speed and the ability to work with people who have skills that an in-house team may not have. If a task typically takes one employee a week to perform, a business can cut the turnaround time to a matter of hours by breaking the job up into many smaller parts and giving those segments to a crowd of workers. Companies that need some jobs done only on occasion, such as coding or graphic design, can crowdsource those tasks and avoid the expense of a full-time in-house employee.

Disadvantages

It is unrealistic to expect that a group of people will give data at a same time. People want quick output of their input. But it needs some time to process crowdsourcing data. Sometimes to attract people into providing data, they must be offered with some advantages in return. There is a big problem of false information also. People provide false information often and that creates a huge problem of getting proper result.

The Difference Between Crowdsourcing and Crowdfunding

Crowdsourcing seeks information or product and crowdfunding seeks money to support individuals, charities or startup companies. People can contribute to crowdfunding requests with no expectation of repayment, or companies can offer shares of the business to contributors.

Challenges in Crowdsourcing

Crowdsourcing refers to solving large problems by involving human workers that solve component sub-problems or tasks. In data crowdsourcing, the problem involves data acquisition, management and analysis. But the main challenge in our project as well as any other crowdsourcing project is cheat detection. When thousands of users provide different kinds of data, it is very difficult to find out which information is right and which is not.

Solution to these challenges

For solving the different issues of crowdsourcing, there were some tried approaches in the previous work, which we call filtering techniques. The techniques are given below.

1. Majority Decision: In case of majority decision, a particular task is given. The solution which is submitted most is assumed the correct solution for this task.
2. Control Group: In control group approach, a single worker works on main task and there are some control groups. The control groups recheck the result again and again. And a result is considered valid if it is accepted by the maximum number of control groups.

3. Swift River

An open-source project to validate data. Ushahidi, a Kenyan company initiated this project. They gather and filter information from a variety of channels for example SMS, Email, twitter etc.

Combining both decentralized human filtering and natural language processing. Specifically developed natural language parser automate event-date extraction based on the parameters “to whom, where and when and Who did what?” More reports are collected to be sure about a specific event, the more likely it is that the event in question actually took place as described by the crowd. If any

user report many times about a specific event, 4 levels of filter validation is used.

Level -1:

The digital straw automates the extraction of place names, noun and verbs from each report for example location, actor and action. Time and date would automatically be coded based on the time of submission. Reports that use similar verbs (synonyms) and refer to the same or similar actors at the same location on the same day can then be clustered into appropriate event-clusters.

Level - 2:

Compare the content of the reports to check if they were exact replicas. If a match is found then Swift River could flag the particular message as suspicious. This means someone is trying to make its dataset polluted.

Level -3:

If an article comes in focus in many newspapers then it will get extra weight.

Level -4:

could identify the email addresses, IP addresses and mobile phone numbers in question to determine if they too were different.

4. Human Filtering:

Parsing the incoming reports and to clustering them into unique event-clusters is one of the main task of Swift River. Then human volunteer visualizes these event clusters to filter.

Flag icon is used to mark a report after the four level of filtering process mentioned above. The flag icons are green, red, and orange. Green flag indicates that the report is already verified as valid. Orange flag indicate the need for more review by the crowd and red flag indicates an immediate review. If a report is confirmed by the crowd as false then Swift river would note associating ip address, email or phone number. Besides Swift River can rate the credibility of users.

We are studying these filtering techniques more precisely as well as we are aware of some other useful techniques that would eliminate false information and redundant information properly.

About our app:

Our app Sachetan: Sachetan app was released a few years back considering all the crimes described in the upper sections. This is developed in android platform and is online serving for various purposes.

The app is a unique one where anyone can login to see the crime infiltrated areas marked on google maps. Detailed information about the app will be provided in different sections below.

Understanding the Sachetan app:

As we mentioned earlier, this app was developed a few years ago by our seniors. The development progress was not documented properly. So, we had to work quite hard to get the proper understanding of this project. The key things we understood is given below-

Front end:

The app is developed in android platform that contains almost 15 java classes and 10 layouts or designs. Whereas there are some common java files like **SignInActivity** or **LoginActivity**, others were mostly designed and developed for specific functionalities that were required in the app. As example, **MapActivity** and **ReportActivity** were designed as the names suggest for controlling the customized map on the homepage and for reporting crimes respectively. The layouts were designed in a quite user-friendly way where users can tell the functions of any button or tap just by seeing the screen!

We had to review all these java classes and layout first just to get the overview of the project.

Back end:

Backend was developed using the **Java spring framework** and **MongoDB** as database server. We are quite new in these platforms as we are still trying to figure out the whole backend process. Hopefully in the next few weeks we will be able to get hold of the whole backend system and use them in the best way possible for a better output. The server architecture will be described in another section briefly.

Our work on the app:

Cleaning the app

As partial work in this thesis progress, our main concern was on this app, to make the app more usable, more user friendly and most importantly, **Bug free**.

Throughout the last couple of months, we worked on this app to make it more acceptable. Besides providing the heatmap of crime and generating graphical interfaces, one of the main feature of this app is to collect crime reports from users. For this reason, we had to make sure that the users face no problem at all while going through this process. There are a few things listed below that we have worked to make this app more acceptable:

First of all, the app was full of bugs as its development work was not reviewed properly as it was supposed to. The reason is quite understandable as throughout the development process the main goal was to add all these must needed features

That the app now provides. We are 3rd group working on this project. Before us, 1st group worked on developing the app, 2nd group focused on adding statistical analysis. Now as 3rd group, while reviewing the app properly we found quite a number of bugs that we needed to fix.

While some of these bugs were visible from the first look at the app, some others were hard to find and we had to use the app from every angle possible. From the bugs that we fixed, some noticeable are mentioned below:

1. **Sign Up problem:** This activity runs without any problem for the mobiles that has a screen size of 5 inch or more. But anything lesser than this, the app encountered some problems as it could not load the necessary fields and buttons.

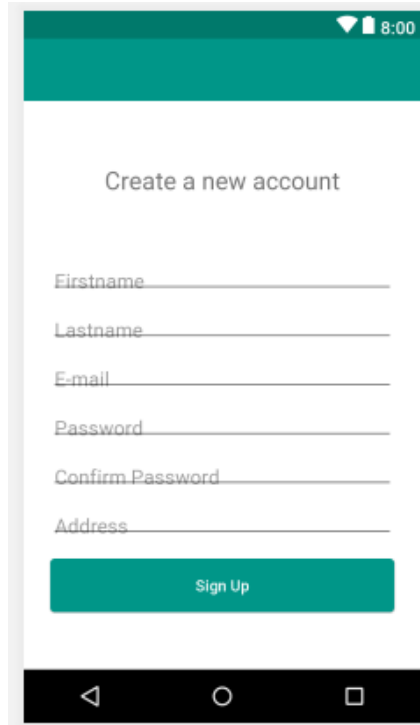


Figure: SignUp Layout

That's how it should look, but before fixing this, only the areas from "First Name" to "Confirm Password" appeared in the screen and the "Address" and "Sign Up" button disappeared.

The problem was solved using some layout change, version checking and also by changing some code of screen controlling.

2. **Crashing:** When the app is opened from minimized status, unfortunately the app crashed every time. The app didn't start properly after that crash. We had to check every function related to the activities to find the root of the problem. The problem almost looks like the below image:

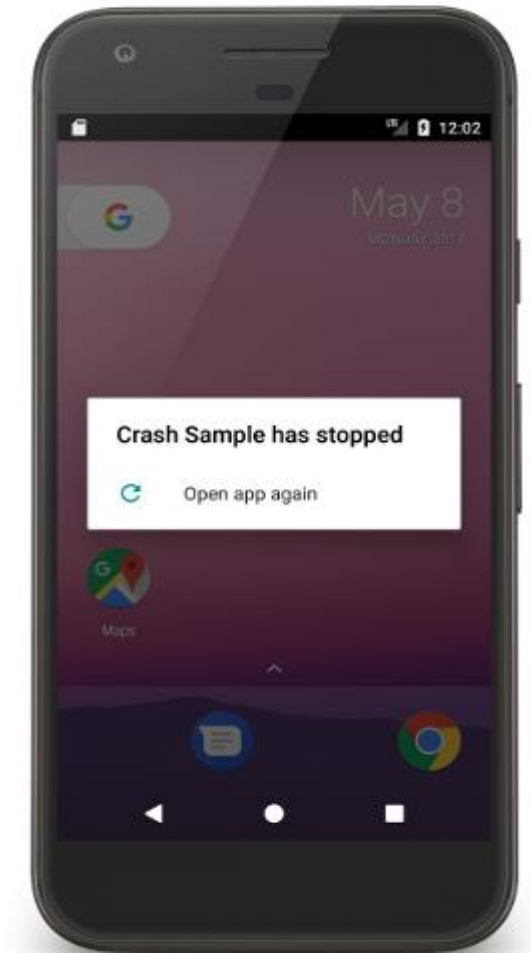


Figure: Crash Sample

We had to debug it properly keeping it in mind that solving this doesn't hamper the usual functionality.

3. **Alternate path:** The app suggests the user a path where the crime frequency is considerably less than some other path. Until this point there was only one path was shown in the map, we have tried to implement a method where the map will show multiple alternatives so that the users can choose the best one from them.

4. **Login Page:** When the users register for the first time, after hitting the Sign-Up button, registration completes but it takes the user to the login page rather than taking the user inside the app! This could be very annoying for some users and so we fixed this problem upfront.

Beside these problems there were some other errors too that did not affect the app directly but could create problem eventually. We have tried to find them all and solve them.

5. **Server related Problem:** The app is currently hosted in Amazon hosting service, which was free of cost for one year. Now the free trial is over and running on paid service. We are trying to find any alternative as the service charge is quite high!

We have considered our university server, contacted with some of the people who knows well about the server only to find out that the server is quite faulty and maintenance takes a lot of time in case of any server damage! So, we can not host it in our university server and still trying to find a better solution where we can deploy the app and maintain regularly without facing any problem. Let's also not forget the payment issue that comes with deployment of the app to any live server!

Additional Features:

We have also added some necessary features in the app. From the google play store reviews, we have noticed some comments about not having a password recovery system.

1.Password Recovery:

The user Asif Mohaimen said: "I have lost my password. I do I recover it? There's no option for that."

We have worked on it and added the email verification and password recovery system in the app. So, from now on the users will be more authenticate and they can recover the password at any time.

2. Multiple alternative path:

As we mentioned in earlier section, users can now choose between paths that suits them best.

These were the features that were added so far!

Current Condition of the app:

The app is almost bug free! But in order to launch this nationwide, we are still trying to find (if any) the bugs and hopefully then we can solve those and make the app 100% ok. But without this precaution, we think our app is ready to be launched.

In the next few pages, some screenshots are provided of the most important activities that exists in the app.

1. Map Screen: It is the homepage of our app. When the is clicked open, users will see this page with the indication of his/her current position and all the crimes that have been reported around the user.



Figure: Map Activity

2. **New Report:** Authenticated user can report any kind of incidents that they experienced. They can select the category and report the incident.

A mobile app interface for reporting an issue. The title bar is teal with the text 'Report an issue'. Below it, a question asks 'What kind of problem did you experience?'. A text input field contains the word 'Mugging'. Below this is a larger text area with the placeholder 'Write what happened'. Another question asks 'How severe is the issue?'. A horizontal slider is positioned at the right end, with the text '4 - Very serious' below it. At the bottom are two buttons: 'Cancel' and 'Post'.

Figure: Report Activity

3. **User Report:** Users can see the reports that they have posted in this activity.

A mobile app interface titled 'My Reports'. It shows the user is logged in as 'john.doe@gmail.com' with the name 'John Doe' below it. A 'Sign Out' button is to the right. Below this is a section header 'Reports You've Posted'. There are three report entries, each with a colored square icon (4 in orange, 5 in red, 2 in yellow), a title, a location 'Unknown', and a description. The first report is 'Mugging' with the description '3 men held a knife to my throat and took all my belongings. Please avoid that area, stay safe!'. The second is 'Road Accident' with 'Two trucks collided, 2 were killed on spot.'. The third is 'Sexual Harassment' with 'Two eve-teasers were harassing school going girls. They fled when someone threatened to call the police.'. Each entry has a close button (X) to its right.

Figure: User Report Activity

Server Architecture

Sachetan was built using up-to-date technology and platforms to be able to serve a large number of users as seamlessly as possible. It was built using a client server architecture, where all of the data is stored in the server and delivered to the client per request.

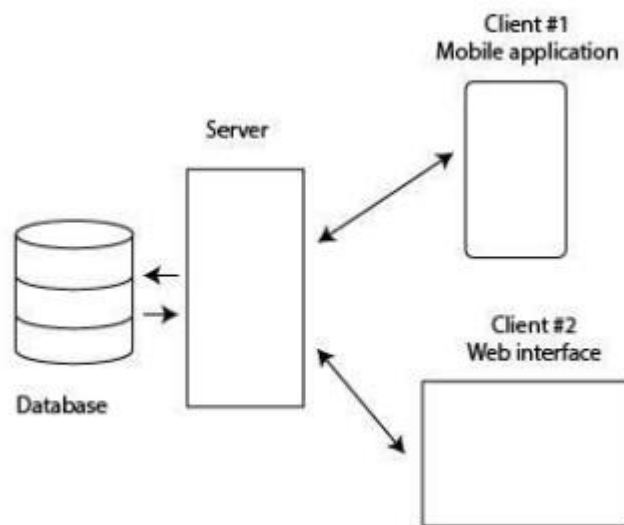


Figure: Server Architecture

NoSQL Database

Web servers, up until recently, have been using relational databases for storing and managing their data. However, they lack the ability to adapt to rapidly changing data types and meet the demands presented in the development of modern applications. Keeping this in mind, Sachetan was developed using a NoSQL database at its core. Based on ease of use and deployment, MongoDB was chosen from an available pool of NoSQL database systems. Another reason for picking MongoDB was its ability to store geolocation and perform spatial queries conveniently.

Java Spring Framework

A web framework is a software framework designed to support the development of web applications including web services, web resources and web APIs. Web frameworks are used to reduce the overhead associated with common tasks performed in the development process. Sachetan uses the Java Spring framework which offers a vast number of built-in modules that makes development, deployment and iteration fast, efficient and secure.

Amazon Web Service

Amazon Web Services (AWS), is one of the most popular web hosting services today. Amazon Elastic Compute Cloud, also known as EC2, is one of the most well-known among the suit of cloud-computing services provided by Amazon. It allows users to rent virtual computers on which they can run their own computer applications. Using the web service provided by EC2 a user can boot an Amazon Machine Image and configure a virtual machine. Such a virtual machine is called an instance, and it encourages scalable deployment of applications. Keeping scalability in mind, the backend web application of Sachetan was chosen to be deployed in an EC2 instance.

Methodology:

News Data Collection

On the previous work, A crawler extracted news data from online newspaper and also from different newspapers our seniors extracted news data.

1. Prothom Alo
2. Bhorer Kagoj
3. Sylhetnews24
4. SylhetSongbad

Incident Types

In our work, we have chosen to collect data about the following types of incidents:

1. Mugging: attacking and robbing someone of their valuables in a public place
2. Robbery: stealing the property of one or multiple individual or organization by threatening to use force.
3. Road Accident: motor vehicles colliding or going off road resulting in personal injury.
4. Sexual harassment: bullying of sexual nature, or inappropriately demanding sexual favors in exchange of something.

5. Theft: the act of stealing valuable goods or money, burglary.

Collection of News Article Data

Web crawlers or internet bots were programmed to systematically browse news sites.

If an article consisted of one or more predefined keywords, it was extracted and appended to a text file for the relevant incident type. Each incident type had its own set of keywords.

The following major newspaper websites were selected for crawling based on popularity:

Prothom Alo: <http://www.prothom-alo.com>

Bhorer Kagoj: <http://www.bhorerkagoj.net>

Shomokal: <http://bangla.samakal.net>

SylhetNews24: <http://sylhetnews24.com/>

SylhetSongbad: <http://sylhetsangbad.com/>

The primary crawling process yielded the following results in the selected categories:

	Number of articles found				
	Mugging	Road Accident	Robbery	Sexual Harassment	Theft
Prothom Alo	360	364	327	119	485
Bhorer Kagoj	722	561	670	275	1385
Shomokal	531	400	391	160	1513
SylhetNews24	67	43	43	25	98
SylhetSongbad	38	11	26	12	46

The collected dataset was very noisy and unreliable. Human intervention was necessary as computer aided filtering methods were not capable of eliminating all the noise from the dataset. After meeting certain criteria, the article was added to the refined list of articles.

Extraction of Date and Location Information from Articles

After cleaning up the data the next task was extracting date and location information. There were two types of data in the newspaper. One is the description of particular incident type and another is the full overview of the crime of one particular area for several days. As both types of news articles were into consideration, instead of trying to calculate the exact date of the incident, each article was flagged with the date of its publication.

The next task was to extract location information from the articles. In order to detect the name of a place occurring within a sentence, each word of the sentence has to be matched with a list of places. Tagging an article with location information turned out to be an optimization issue because there were more than one Upazila names on some articles and much of the Upazila names were included with suffixes. So, the matching algorithm could not match all the Upazila tags. Careful methods were applied to avoid errors such as- if the name was in the headline then it mostly be the location of the incident and also if the name was nearby to the keyword, it should be the desired location and

Future Work:

In order to make this app more efficient, we have still a lot of work to do. Some of them are listed below:

Add more features:

It is quite understandable that a user may not use the app quite often as compared to other normal apps. So, we have to find a way to make the user stay with us even

though he/she may not report any crime or see the map for any reason. We have considered adding Feeds of news, specially crime news featuring in the app. This way the user will not have to visit any other news portal for any crime information.

Adding SOS button:

We have considered running the app on background and add quick button to send SOS messages to the server! We have thought that the SOS message will contain the current location of the user and a common help message.

Notification on the phone:

In case of any recent crime reports, the app will send notifications to all the users. This may happen via the mobile app or via the Mobile SMS that is registered against the users' account.

Various reporting Method:

If a user loses his own mobile phone then it would not be possible for him to report the accident himself through the mobile app. So, we are considering reporting system through Email or Mobile SMS. If we can complete this, then we believe more data will be added in the app and more information we will get.

Criteria for Pattern generation: Right now, the app uses some basic criteria like frequency of crime in an area for finding the patterns of the crime. If we want to find more interesting patterns then we have to come out with new ideas about how we can connect the dots between these crimes!

Data Collection: This is the most important part of our thesis program. If we are to find the patterns between these crimes, we have to analyze them more and more and with more data. So, we have to make sure that the app is used for providing crime reports and only then we will be able to do the full analysis on them and come up with any new pattern!

Filtering Technique Review:

Crowdsourcing is one way for collection crime related data. The other available process for this is to crawl newspaper data related to this topic! This is a very important part of our thesis as it brings us the data we need to analyze.

The current crawler added in our app is not quite efficient as it is often unable to filter the correct data from the wrong. We have to modify the crawling system properly so that only the mugging, robbery, theft, sexual harassments etc. are filtered rather than any other news.

We also have to filter some news that uses the keywords related to crime but represents a different meaning!

As example: Here is a news published on a daily newspaper:

“Bangladesh snatches victory over Sri Lanka in a nervy fight”

Here words “Snatches” and “fight” both are used in an inspiring way rather than in any negative way. But the way our news crawler crawls data right now, it is unable to differentiate these two meanings of the same word. So, we have to modify our crawler and increase the accuracy.

Another important task that has to be done is the detection of correct or wrong information. Crowdsourcing is a process where users can share anything they want. In order to filter the correct information, we have to apply some existing technique related to this and test the accuracy. If the accuracy is not satisfactory then we have to find a better solution.

Interface Modification: The user interface in our app is already very eye-catching. But change is always needed if we want to give the users a taste of something new! So, if needed, we will customize the interface for a better experience.

User Notification: We have some incident data that was collected in hard copy. We will create account manually for those people who volunteered in sharing the incidents. Then the data need to be inserted in our database and make them ready for generating the pattern.

Goal: Our ultimate goal is to collect as much data as possible and then analyze them more and more to find the pattern of these crimes and also help the user to be safe from these crimes.

Conclusion:

“Sachetan” app is developed with the view of doing something for the greater good in our society. All it takes is a few minutes of the user who can share their own or other peoples’ terrible experience through the app to give us the opportunity to analyze those data and find any available patterns to stop the ever-growing crimes in our neighborhood. We hope that we will soon be able to get a breakthrough in our goal.

References

1. Ahmed, S.I., Jackson, S.J., Ahmed, Nova., Ferdous, H.S., Rifat, Md. R., Rizvi, A.S.M., Ahmed, S., Mansur, R. S.: Protibadi: A platform for fighting sexual harassment in urban Bangladesh(2014)
2. Ali, M.E., Rishta, S.B., Ansari, L., Hashem, T., Khan, A.I.: SafeStreet: empowering women against street harassment using a privacy-aware location based application(2015)
3. SOS Response, <https://globalnews.ca/news/540110/app-aims-to-increase-safety-for-b-c-women-at-risk-from-domestic-violence/>
4. Not your Baby App, <http://www.gender-focus.com/2013/05/18/new-apps-tackle-dating-violence-street-harassment/>
5. Circleof6 website: <http://www.circleof6app.com/>
6. Bowman, C. G.: Street harassment and the informal ghettoization of women. Harvard Law Review, 106(3), 517–580, 1993.
7. Safetipin App: <http://safetipin.com/>
8. Yarrabothe, R.S., Thota, B.: Abhaya: An Android App for the safety of women. India Conference, 2015 Annual IEEE
9. OnWatch, <http://www.onwatchoncampus.com/>
10. Vairaperumal, N., P, Saravannan, S., Kalpana.: SheSecure Safety App–The Hexa Umbilical Cord
11. Brantingham, P.L., Brantingham, P. J.: A theoretical model of crime hot spot generation. Studies on Crime & Crime Prevention, 7--26 (1999)
12. Bendler, J., Brandt, T., Wagner, S., Neumann, D.: Investigating Crime-to-twitter Relationships In Urban Environments - Facilitating A Virtual Neighborhood Watch(2014)
13. Hollaback! Together We Have the Power to End Harassment, <https://www.ihollaback.org/>
14. HarassMap - Stop sexual harassment, <http://harassmap.org/en/what-we-do/the-map/>
15. Daily Crowdsourcing, <https://dailycrowdsourcing.com/training/crowdsourcing/what-is-crowdsourcing/>

