Abstract—Terrorism is a complex and evolving phenomenon, and terrorist groups are highly dynamic in terms of cooperation and locality. Terrorism data collection organizations such as the study of terrorism and response to terrorism (START) release terrorism incident data on a yearly basis. Such data is valuable for historical terrorism incident analysis. However, it is crucial that researchers have access to real time terrorism incidents data to perform analysis such as safety level estimation, monitoring or projections. We developed a novel system that automatically collects news data and analyzes them in real time for terrorism incidents. We developed a safety assessment formula based on recent terrorism incidents and time. Our system displays the safety level of countries and cities on a dynamic map, and it also allows to tracks the events of a selected terrorist group. Our results show that countries like Israel, Nigeria, Ukraine, Iran, and Afghanistan are currently among the most unsafe places in the world. Our system can assist counter-terrorism experts to estimate the potential risks of countries, cities and other geo-locations. A country is unsafe when it undergoes many time terrorism incidents data, we developed a formula to calculate the safety level of countries and cities to estimate their risks. A country is unsafe when it undergoes many terrorism incidents. We developed a method for automatically gathering terrorism incident data from news sources around the world and storing them in the Global Terrorism Database (GTD). START integrated into GTD the databases of other terrorism data collection organizations such as Pinkerton Global Intelligence Service (PGIS), and the Center for Terrorism and Intelligence Studies (CETIS). The 2013 version of GTD [1] contains over 125,000 terrorism incidents starting from 1970 to the end of 2012.

III. METHODOLOGY

A. Crawler

We designed and implemented a web crawler based on the architecture on Figure 1 to periodically download news data from reliable sources such as New York Time, Washington Post, and Al Jazeera. The number of news sources that can be fetched is virtually unlimited. Moreover, the execution time to check for new news can be set to as short as one second. Some settings criteria are followed to assure that all news of interest are fetched on a regular basis. A web crawler can be scheduled to fetch one or many news sources. Similarly, the same news source can be fetched by many crawlers which is significantly beneficial as when a crawler fails, another one can pick up the task. In addition, the more crawlers placed on a news source, the more reliable is the information retrieval. Our system is distributed which allows the crawlers and other components of the system to communicate and exchange messages.

B. Safety Level Formula

Terrorism incidents occur due to many factors. Using real time terrorism incidents data, we developed a formula to calculate the safety level of countries and cities to estimate their risks. A country is unsafe when it undergoes many criminal and terrorism incidents. For our research purpose, only incidents that qualify as terrorism incidents [1] will be considered in the safety calculation. The safety of countries or cities partly depends on the frequency of incidents that occur in them. Moreover, time is an important factor in determining
the safety of a location as it allows to represent the current situation. For instance, a location that undergoes frequent criminal incidents in 2000 is considered unsafe in that year. However, if no incidents occur at the same location in 2014, its considered safer. Based on these assumptions, we developed the theorems below to derive a formula to estimate the current safety level of countries and cities.

C. Theorems

**Theorem 1:** - News Filtering and Incident Identification.
Terrorism incidents news include terms like: bomb, explosion, suicide, kill, coerce to name a few.

**Theorem 2:** - Time Factor.
Recent terrorism incidents have higher impact on the safety than old incidents.

**Theorem 3:** - Frequency Factor.
The safety estimate of countries and cities is based on the frequency of incidents occurring in those particular locations.

As mentioned above in theorem 1, not all the news downloaded by our system will be included in the safety calculation. We developed three main inclusion criteria list below.

D. Inclusion Criteria

**Criteria 1:** As illustrated on Figure 2, only news that contain our predetermined terms such as suicide, explosion will be included.

**Criteria 2:** Time is an important determinant of our formula and it is represented by $x$. Where $x$ is the number of days past the incident. The safety will only include incidents that occurred in the pre-set $x$ period (e.g. last 30 days, 3 months, etc.). According to our theorem 2, recent incidents have higher impact on the safety than older incidents. We use the reciprocal of $x$ to attribute higher values to recent incidents and lower values to older ones. Figure 3 shows how the values will be attributed based on the time the incidents occurred.

**Criteria 3:** The incidents $i$ are captured in real time for each country and city. Each terrorism incident is counted as one (1) regardless of the weapon, target, casualties and perpetrators. The incidents are summed on a daily basis to $\Gamma$:

$$\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3 + \ldots + \Gamma_n$$

where $\lambda_1$ is Today, $\lambda_2$ Yesterday, etc.

Thus, the incidents of 5 days is as follow:

$$\Gamma_5 = \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4 + \Gamma_5$$

Therefore, the incidents of $n$ days is:

$$\Gamma_n = \Gamma_1 + \Gamma_2 + \Gamma_3 + \ldots + \Gamma_n$$

The value of $n$ can be set to include the desired range of days or years.

Now that we have the incidents portion set, we apply the time factor using the reciprocal of $x$, where $x$ is the number of days past the incident. We multiply the daily incidents by the $x$ days of the incidents. For instance, using reciprocal of
The safety based on 5 days incidents is as follows:

\[
\text{Safety} = \Gamma \lambda_1 \left( \frac{1}{x} \right) + \Gamma \lambda_2 \left( \frac{1}{2} \right) + \Gamma \lambda_3 \left( \frac{1}{3} \right) + \Gamma \lambda_4 \left( \frac{1}{4} \right) + \Gamma \lambda_5 \left( \frac{1}{5} \right)
\]

Therefore, the safety based on \( n \) days incidents is as follows:

\[
\text{Safety} = \Gamma \lambda_1 \left( \frac{1}{x} \right) + \Gamma \lambda_2 \left( \frac{1}{2} \right) + \Gamma \lambda_3 \left( \frac{1}{3} \right) + \ldots + \Gamma \lambda_n \left( \frac{1}{x} \right)
\]

Finally, the time based \textbf{Safety} formula is:

\[
\text{Safety} = \sum_{i=1}^{n} \sum_{j=1}^{x} \left( \Gamma \lambda_i \left( \frac{1}{x} \right) \right)
\]

It is worth noting that the values of \( x \) are equivalent to the values of \( n \) in incidents.

We apply this formula to real time data collected by our system to calculate the safety of countries and cities.

\textbf{IV. EXPERIMENTAL RESULTS}

\textbf{A. Data}

Our data is automatically collected in real time from the news sources. The number of news sources is potentially unlimited and is a many to many relationship which means that the same crawler can fetch many news sources and the same news source can be fetched by many crawlers. The data collected is stored in a database which is then analyzed by other components of the system.

\textbf{B. Analysis}

When the data is collected and stored in a database as shown on \textit{Figure 4}, different components of the system scan the database to filter and classify potential terrorism incidents. The safety formula is then applied to the filtered terrorism data. The \textit{GeoChart} on \textit{Figure 5} shows in real time the safety level of all countries undergoing terrorism incidents. The darker the red color the riskier is the country.

At the moment this article is written, the following countries are among the most unsafe in the world: \textit{Syria, South Korea, Nigeria, Ukraine, Pakistan, Israel, Iraq, Kenya, Afghanistan, China, and North Korea}.

Moreover, we included the option to zoom into countries to view cities that are affected by terrorism incidents in that particular country. The map on \textit{Figure 6} represents the estimated safety level of cities of Nigeria in June 2014. This safety estimate is calculated based on 30 days events. A close analysis of the data shows that a total of 18 cities were affected by terrorists and most of the incidents were caused by the group Boko Haram.

The final functionality of the system is the impacts of each terrorist group per country. We calculated the impacts by adding all the incidents of terrorist groups in different countries in a determined period of time (e.g. last 3 months). Each incident of a terrorist group in a country is counted as one (1). Then we summed these numbers to determine the final impact score which is the total number of incidents of each group in each country. Therefore, when estimating the impacts of \textit{Al-Qaeda} on countries based on the last three months incidents, our system returns the result shown on \textit{Figure 7} which is the sum of all incidents related to \textit{Al-Qaeda} by country in that period of time. The darkness of the red color indicates the level of impact of \textit{Al-Qaeda} in different countries and darker red represents higher impact.

Our system is fully automated and does not require any manual intervention starting from news collection, scanning and reporting results. However, not all news reported by the system are terrorism incidents. The system automatically considers any news containing a country name in addition to one or many of the pre-defined terms such as suicide, explosion, and bomb as terrorism incidents and includes them in the country safety level calculation. Furthermore, different
news organizations often report the same incident, and the same news organization can provide updates on the same incident with different titles. Therefore, to easily identify and exclude duplicates, we designed a screen to allow easy access to the details of news included in the safety calculation and terrorist groups events tracking. The details screen allows experts to perform the following tasks: confirm news as terrorism incident, indicate news as false alerts, and indicate news as existing in case it is a terrorism incident already reported by other news. By default, the news get automatically included in the safety calculation, but can be reviewed and marked as false alarm, or existing. False alarm and existing news are excluded from the safety calculation. Only Unconfirmed news and Confirmed news will be automatically included in the safety calculation. Similarly, incidents flagged as existing or false alert will not be included in the terrorist groups events tracking.

V. Conclusion

In this research, we addressed three main issues in terrorism study: availability of live terrorism data from multiple sources, the safety level of countries and cities, and the impacts of terrorist groups in different countries. First, we developed a system consisting of crawlers to automatically collect news data from multiple sources simultaneously. Second, we developed a safety estimation formula by taking into account three main factors: data relevancy, time, and frequency. Then, we applied our safety formula to the live data collected to estimate the safety level of countries and cities. Finally, we analyzed the incidents of different terrorist groups to estimate the impacts of each terrorist group on each country.

In our future research, we plan develop a method to project potential terrorism incidents in different countries by combining real time and historical terrorism incident data. Moreover, we plan to develop another method to track the movements of terrorist groups at different locations based on their attack patterns.

REFERENCES


