# Design and Methodology

A web extension is a plug-in that extends the functionality of a web browser. The researchers will create a web extension that can identify fake news and will only run if the link is clicked. The extension will be uploaded to Google Web Store and it will be accessible by Chrome users.

1. **Common types of Fake News**

* **Satire/comedy sites** - Which can offer important critical commentary on politics and society, but have the potential to be shared as actual/literal news
* **Clickbait-** Headlines that can easily attract attention and encourage visitors to click on a link.
* **Full Fabrication –** Everything in this type of story is fake and designed with intent to do harm.
* **Bias** - Sources that come from a particular point of view and may rely on propaganda, decontextualized information, and opinions distorted as facts.

1. **Gathering of Data**

The data for fake news list will be gathered from (kaggle.com, Opensources.com). Kaggle.com is a data science website. Opensources.com provide a continuously updated database of information sources for researchers to leverage in the fight against fake, false, and misleading news. It is also maintained by professionals who have analyzed each source, looking for overall inaccuracy extreme biases, lack of transparency, and other kinds of misinformation (Opensources, n.d.). The researchers will download the file from the source manually.

1. **Analyzing**

There are three phases for detecting an article if its fake. First the web extension would compare the link from the list of fake news sites. If the link is in the list of fake news sites the extension will warn the user that the page may contain false information. Second if the link was not in the list of fake news sites in the database then the web extension will analyze the news articles using Fakebox to assess whether they are fake or not by looking at a range of available aspects of an article (title, content and URL). Lastly if the link was not found and if fakebox doesn’t recognize the content of a news article then the validation of an article will depend on the vote of online users. An online user can vote whether it is “Fake News” or “Legit”. A sample of how voting works below.

Using this formula:  
(A ÷ B) × 100 = C

Where A is the number of users who voted for either fake news or legit.

Where B is the total number of users who voted.

Where C is the percentage of the total votes.



This headline is clearly fake since it’s a clickbait. Assuming that 100 are the total voters and 63% of them voted fake news and 37% of them voted for legit. Since the majority of voters voted for fake news then it will be marked as fake.

The team will be using indexedDB for the client-side storage. The data from the clients will be synced to MySQL. If the link is confirmed fake the system will notify the user with a dialog box popped up. A captcha would also be used to prevent and spammers and bots. The data from the crowd will be gathered from the database and it will be sent to opensources for expert verification and fakebox for training. The team set that the reports will only be counted if the total voters is greater than 100.

*Figure SEQ Figure \\* ARABIC 1*

*Figure 2*

*Figure 3*

1. **Updating**

The extension is automatically updated without user intervention through Google developer’s dashboard which automatically uploads the file through Google’s server. The researchers will update the data of fake new list monthly.

# Conclusions and Recommendation

Fake news can be lessened, since people keep falling prey for such posts; it needs to be stopped. This paper is meant to do just that; the program we are proposing is meant to inform users about fake news. There are already existing applications that identifies fake news, but this project has different approach on how it detects fake news and warn users. Linguistic and network approach are the methods that will be used to` identify fake news since it was used by many researchers. This program will be a plugin for the users’ internet browsers. If this program could be implemented, there will be a big change in how news would be spread, since users will no longer fall for faulty news articles and posts. The team contribution for this project was to expand and increase the data of fake news links for Fakebox and Opensources.co.

# References

Arias, J. (2017, October 4). *Inquirer.net*. Retrieved from Preen: http://preen.inquirer.net/58185/a-list-of-mocha-usons-fake-news-posts

Bisson, D. (2015, September 13). *Most Suspicious TLDs Revealed by Blue Coat Systems*. Retrieved from The State of Security: https://www.tripwire.com/state-of-security/security-data-protection/cyber-security/most-suspicious-tlds-revealed-by-blue-coat-systems/

Conroy, N. J., Rubin, V. L., & Chen, Y. (n.d.). Automatic Deception Detection: Methods for Finding Fake. 1-4.

*DeZyre.* (2016, January 29). Retrieved from https://www.dezyre.com/article/top-10-machine-learning-algorithms/202

*Facebook, Google testing apps to spot fake news but experts say key is being educated*. (2017, February 7). Retrieved from Japan Times: https://www.japantimes.co.jp/news/2017/02/07/business/facebook-google-testing-apps-spot-fake-news-experts-say-key-educated/#.WlM6JlWWZpg

Google. (n.d.). *Google Recaptcha*. Retrieved from Google: https://www.google.com/recaptcha/intro/android.html

IBM. (n.d.). *How SVM works*. Retrieved from IBM: https://www.ibm.com/support/knowledgecenter/en/SS3RA7\_15.0.0/com.ibm.spss.modeler.help/svm\_howwork.htm

*Indiana University researchers launch tool to understand spread of fake news*. (2016, December 21). Retrieved from IU BLOOMINGTON NEWSROOM: http://archive.news.indiana.edu/releases/iu/2016/12/iu-hoaxy.shtml

Inquirer. (2017, February 7). *Fighting fake news isn’t just up to Facebook and Google*. Retrieved from Inquirer.net: http://technology.inquirer.net/58483/fighting-fake-news-isnt-just-facebook-google

Kiely, E., & Robertson, L. (2016, November 18). *FactCheck.org*. Retrieved from How to Spot Fake News: https://www.factcheck.org/2016/11/how-to-spot-fake-news/

Kinmont, K. (n.d.). Fake News Detection in Twitter. 4-4.

McIntire, G. (n.d.). *Machine Learning Finds “Fake News” with 88% Accuracy*. Retrieved from KDnuggets: http://www.kdnuggets.com/2017/04/machine-learning-fake-news-accuracy.html

*Opensources*. (n.d.). Retrieved from http://www.opensources.co/

Palmisano, S. (2016, December 29). *FAKE NEWS LEADS TO REAL PROBLEMS*. Retrieved from Eagle News: http://eaglenews.org/opinion/fake-news-leads-to-real-problems/

Patel, S. (2017, May 3). *Chapter 2 : SVM (Support Vector Machine) — Theory.* Retrieved from Machine Learning 101: https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

Philstar. (2017, October 4). Retrieved from Philstar: https://www.philstar.com/headlines/2017/10/04/1745491/no-govt-agency-monitoring-fake-news-social-media-nbi-exec-says

Project, M. I. (2016, April 17). *A new understanding: What makes people trust and rely on news*. Retrieved from AmericanPress Institute: https://www.americanpressinstitute.org/publications/reports/survey-research/trust-news/

Saini, B. S., & Bala, A. (n.d.). A Review of Bot Protection using CAPTCHA for Web Security. *Journal of Computer Engineering*, 1-7.

Singh, V., Dasgupta, R., Sonagra, D., Raman, K., & Ghosh, I. (n.d.). Automated Fake News Detection Using Linguistic Analysis. 1-3.

Stecula, D. (2017, July 27). *The real consequences of fake news*. Retrieved from The Conversation: http://theconversation.com/the-real-consequences-of-fake-news-81179

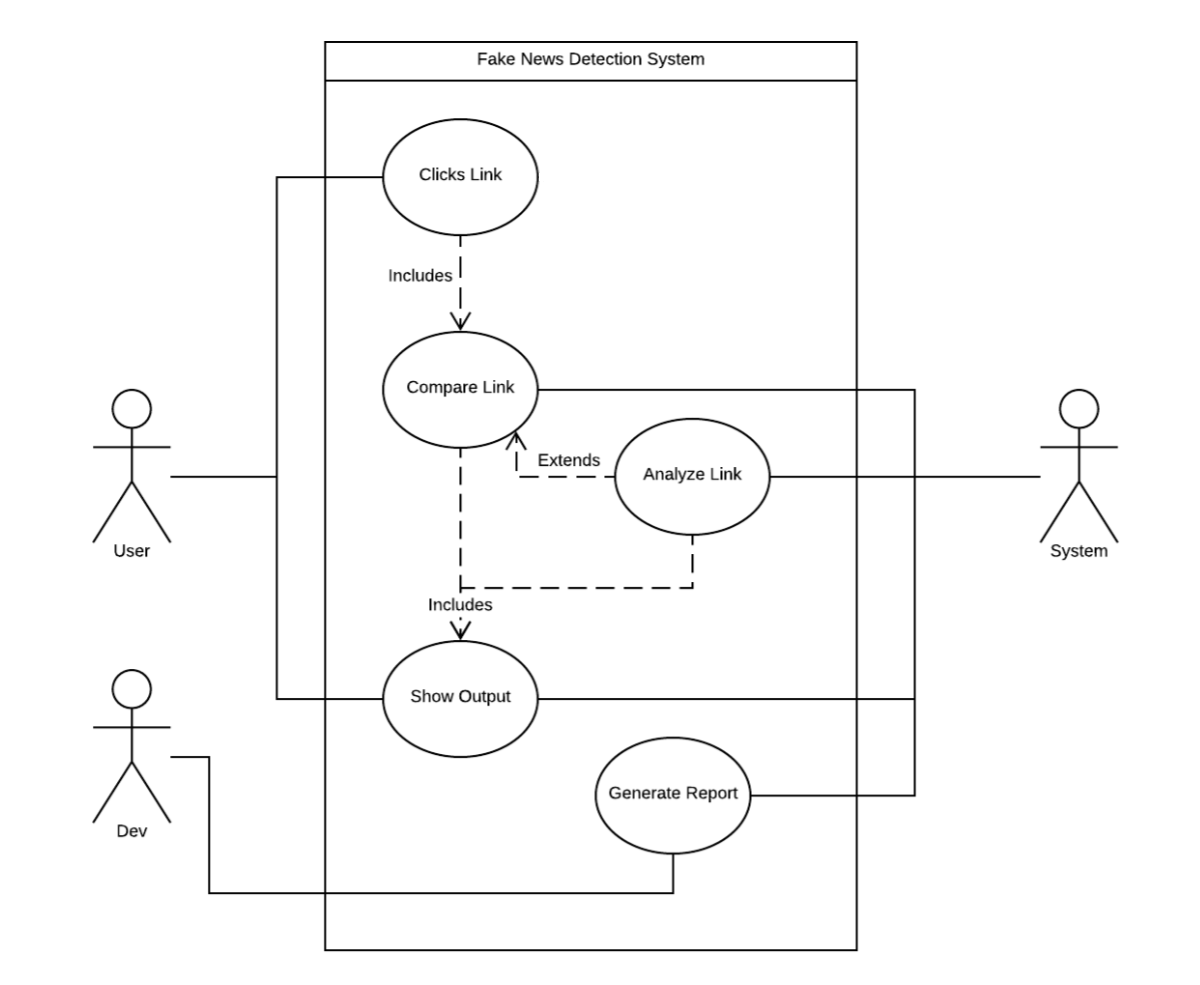
vvv. (n.d.).

*Why use SVM?* (n.d.). Retrieved from alteryx community: http://blog.yhat.com/posts/why-support-vector-machine.html

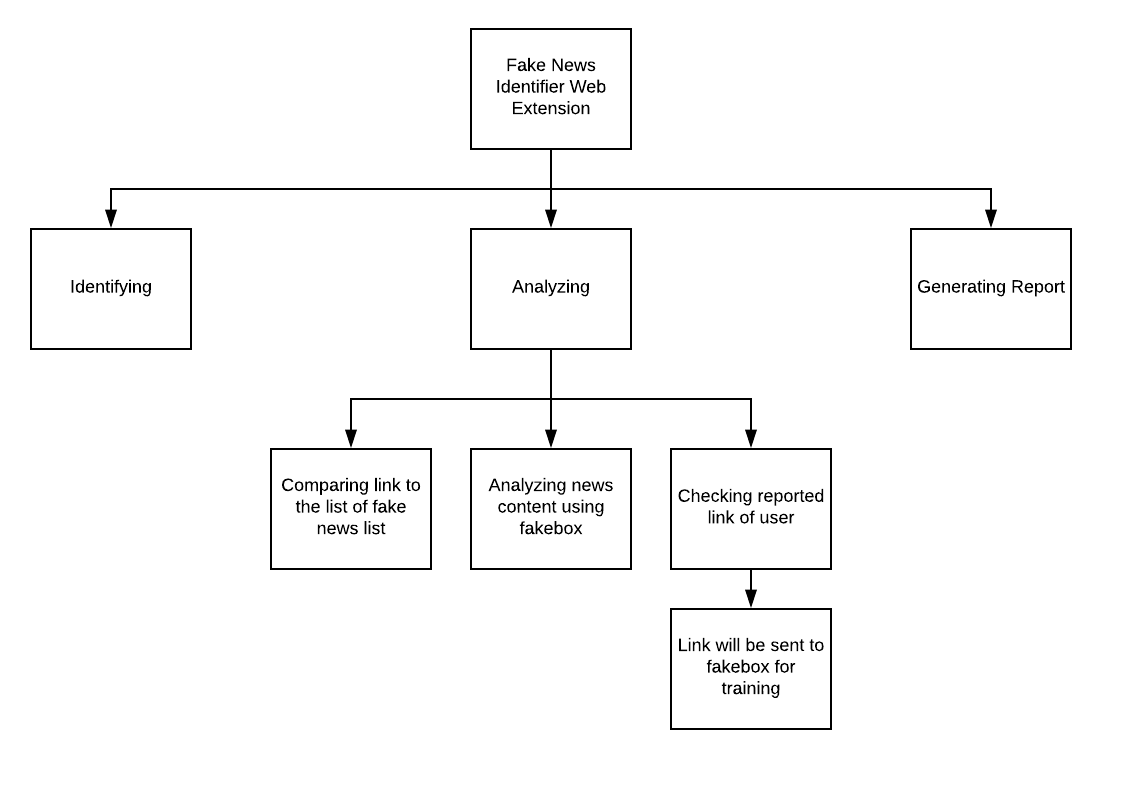
# UML Diagrams

**Use Case**

This diagram represents on how the system perform in collaboration with the other.

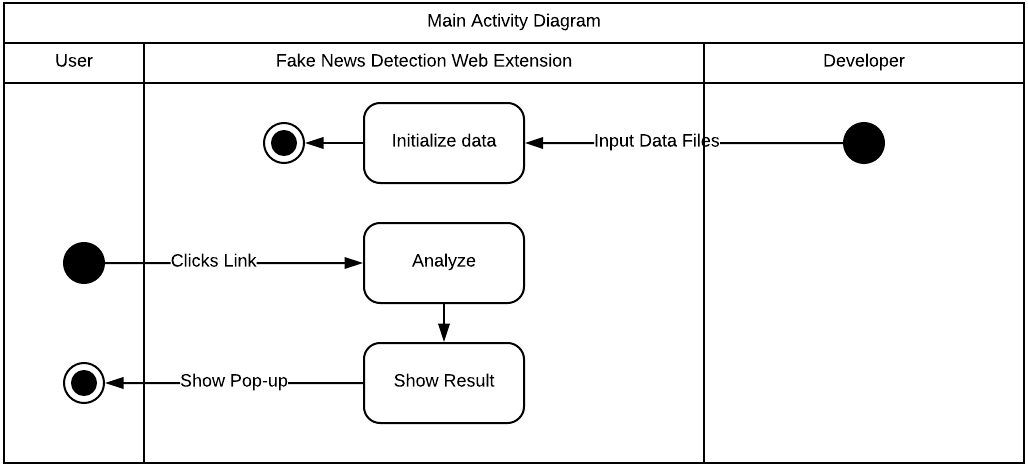


**Functional Decomposition Diagram**

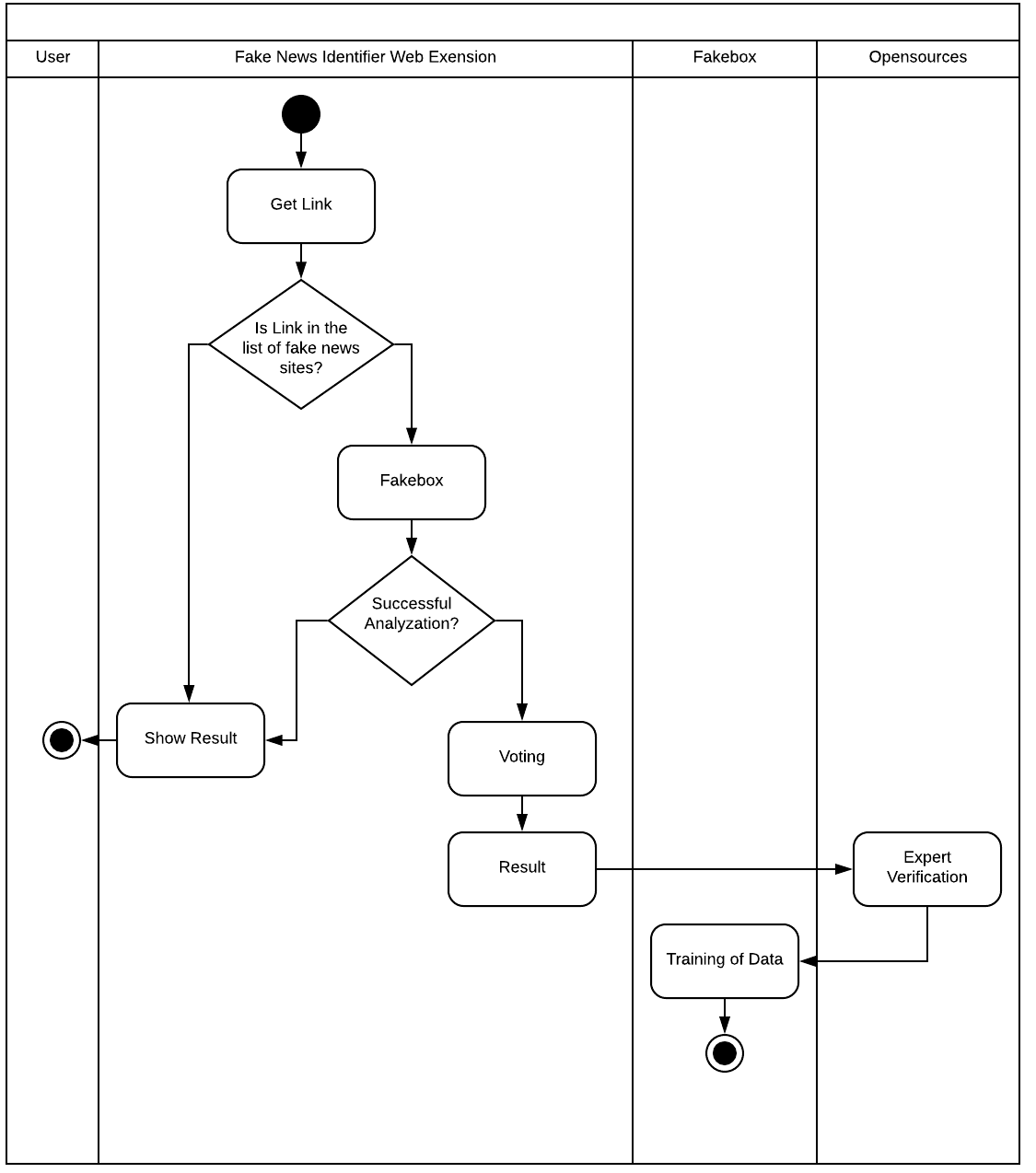
****This diagram represents the process of identifying fake news in hierarchical manner.

**Main Activity Diagram**

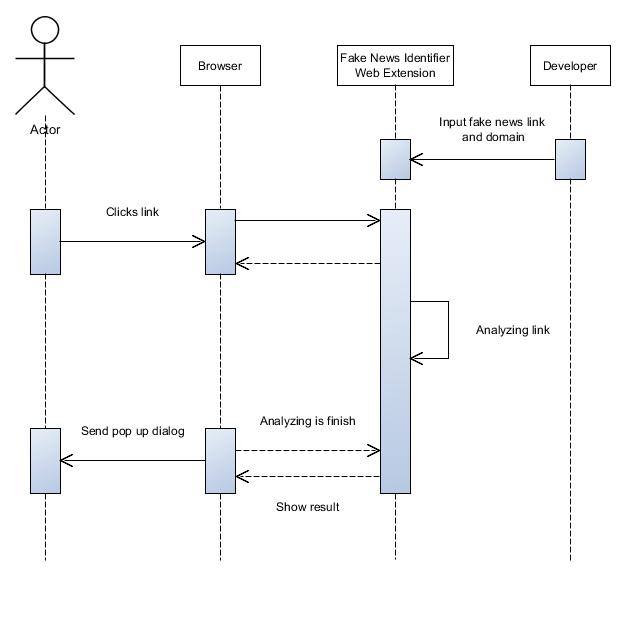
This diagram represents the flow from one activity to another activity.

****

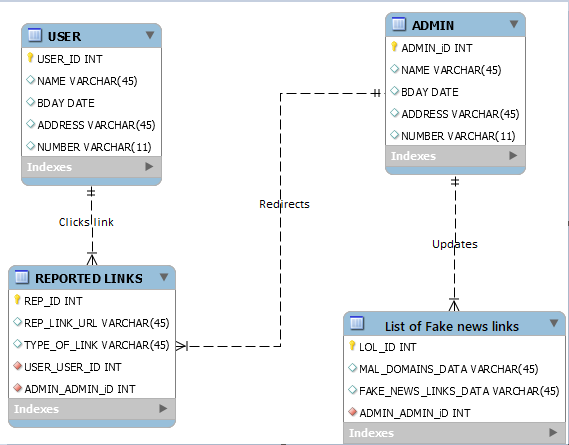
**Analyzing Activity Diagram**

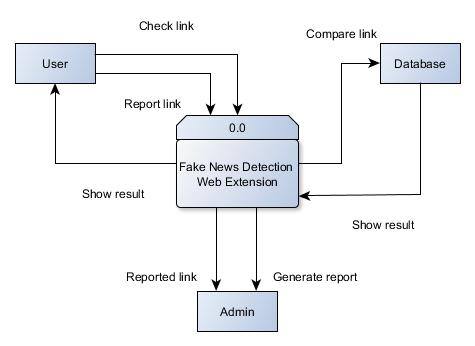
****

**Main Sequence Diagram**

This represents the interactions logic between the objects in the system in the time order that the interactions take place.

**Entity Relationship Diagram**

****This diagram illustrates the entities and their relationship to each other.

**Context Diagram**

This diagram shows the relationship that system has with other external entities

**Data flow Diagram**

This represents the graphical visualization of the movement of data through an information system.

