# A Web Scraping of Chemical Compounds with an Anti-Drug Feature Using IoT

**Nadia Mahmood Hussien[1], Yasmin Makki Mohialden[1], Rehab I. Ajel[1]**

*[1]Department of Computer Science, Collage of Science, Mustansiriyah University, Baghdad, Iraq*

*\*Corresponding Author: Nadia Mahmood Hussien*

**Abstract**

*As a result of the COVID-19 epidemic, there has been an increase in the demand for electronic education apps in schools and colleges in recent years. The purpose of this paper is to develop an educational application for chemistry students at different levels of study that will allow them to obtain precise information on chemical substances in a timely and safe manner. This program uses a web scraping technique by applied a RESRful API to extract information from websites, which is then sent to the student's account. Furthermore, due to the use of the Internet of Things, the application has an anti-explosives and narcotics property using (IOT). The application can retrieve and save information entered by the user on chemical compounds with a high level of security. The medication's chemical formula, as well as the covalent and ionic bonding of compounds, can be displayed. It also has a database that lists all of the hazardous substances. If a user enters a dangerous compound containing narcotics more than four times, an alarm message is sent to the administrator via the Internet of Things.*

## Introduction

When you use web scraping, you can get a lot of information from a website without having to go to the site. This can save you a lot of time and effort in the long run. When you scrape web pages, some of the interface shrouds are removed. This isn't just the browser; it can also be at the level of the network connection. You can use a web browser to create and send packets of data, for example: Pretty pictures and sounds can also be read from the data you get back. Videos and text can also be read from the data that you get. Also, you can break down web browsers and use the code to make them do what you want. This paper talked about scraping web pages (Mitchell, 2018). I think it can help first-year students who are having a hard time with school. People can use it to look up information about chemicals and how they affect the body in public places. PubChem is a site which used in this paper for scrapping. When the National Institutes of Health started making it available to the public in 2004, it was part of the Molecular Libraries Roadmap Initiatives. It was made available to the public in 2008. (NIH). Public Chemistry has more than 1000 bioassays, more than 28 million test results, more than 70 depositing organizations, and more than 19 million unique compound structures (Richard et al., 2020). People can use PubChem to look at the biological information of small molecules. It's a big, open-source platform (Dallakyan & Olson, 2015). Find, compare, and integrate biological screening results from these web-based services quickly. You can also look at how well a drug works for a specific target, look at structure-activity relationships, and look at how well a drug works (Bolton et al., 2008; Kim et al., 2019; Wang et al., 2009). In this paper, the IOT (Internet of Things) is used to send a text message in a certain way. Devices and sensors that connect and exchange data are called "sensors." IoT is the term for this. As IoT

technologies have become more popular, there are more and more practical applications in a wide range of fields. These include asset tracking, agriculture, smart metering, smart cities, and smart homes (Mekki et al., 2019).

After this introduction, the rest of the study is divided into five sections: Section Two represents related work; Section Three describes the proposed system; Section Four describes the results; and Section Five, the conclusion.

## Related Work

As a result, numerous of reseachs which used web scrapping in diffrante forms. A few examples are as follows: In Haddaway (2015), the researchers search for gray literature by web scraping. They employ a pattern of data extracted from websites in this study to do business in the private sector. Additionally, they develop and share protocols that facilitate the extraction of search results and other data from web sites. In Hernandez-Suarez et al. (2018) they obtained information from Twitter by web-scanning. A large number of people on this system make use of the Twitter API. It is a public space where individuals can access information that is less than three weeks old. They devised a new method for obtaining historical tweets from any date range in this study by circumventing Twitter API constraints via site scraping techniques. In El Asikri et al. (2020) they employed web scraping to collect data and track down a certain piece of information for later retrieval or analysis. In this study, they examined scarping approaches for extracting material from Web pages. Additionally, they recommended using Python and scrappy to figure out how to leverage web scraping to obtain additional information.

## The Proposed system

An API called Restful is used in the proposed system to get data from PubChem, which is the world's largest free database of chemical information. Chemicals can be found by their names, molecular formulas, structures, and other things that can help you find them. Among other things, you can find out about chemical and physical properties, biological activities, safety issues, patents, and other things. In Figure 1, you can see the system's block diagram. You can get information about chemical compounds like their names, molecular formulas, and structures by typing in their names in this system. Figure 2 shows how this works.
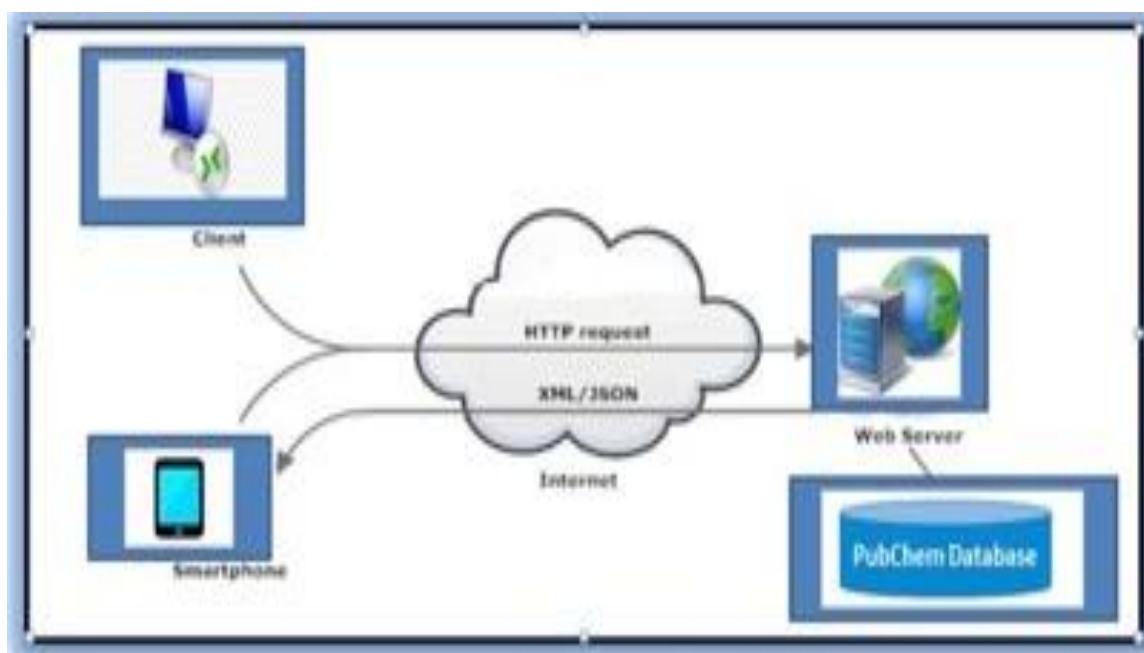


*Figure 1. The system Block Diagram*

*Figure 2. An Exemplification of How the System is Used in*

The system has a very high level of safety. When a user gives it information, it checks it and then compares it to a database that has been set up for that person before In this case, the system sends a real-time alarm message to the administrator. Figure 3 shows how this works. Covalent and ionic bond shapes are also shown in the application's figure 4.



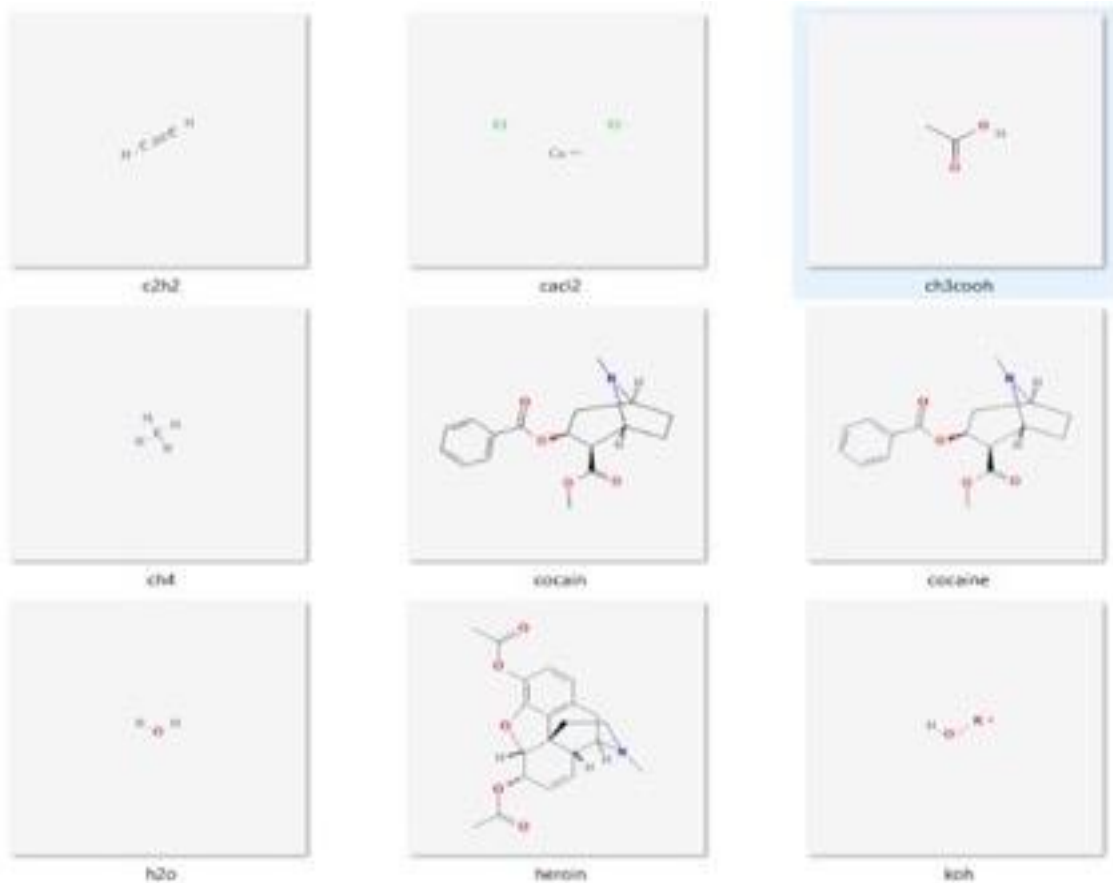*Figure 3. An Illustration of Chemical Substances that are Prohibited*

*Figure 4. Example of Covalent and Ionic Bonds*

## Results and Discussion

The system looks at what the user types in. If the user types in forbidden chemical compounds more than or equal to four times, the system sends an alert to the admin. The system compares the user's input to a database that has been set up to store things like heroin, cocaine, and other things that aren't allowed. It shows the SMS message that was sent to the admin in Figure 5.
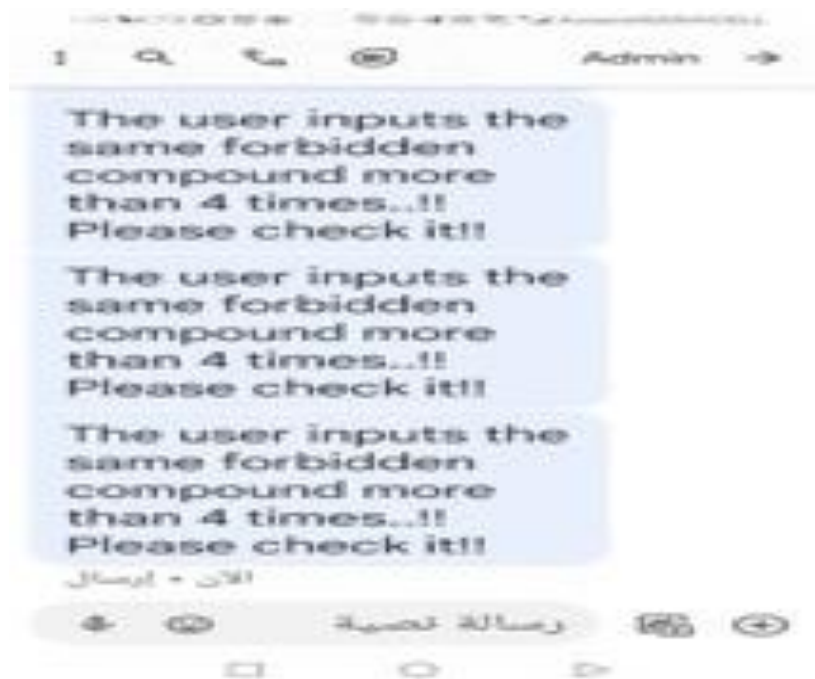


*Figure 5. SMS alarm Message*

## Conclusion

There are a lot of different ways to scrape web data,also known as web data extraction. In this paper, the RESRful API is used to scrape web pages and give the student the information. To protect against forbidden chemical compounds and narcotics, a warning message will be sent when someone enters an unallowed compound using Internet of Things app (IOT). The system serves and monitors the students or users at the same time, which means it does both. PubChem is the website that scrapping.

## References

Bolton, E. E., Wang, Y., Thiessen, P. A., & Bryant, S. H. (2008). PubChem: integrated platform of small molecules and biological activities. In *Annual reports in computational chemistry* (Vol. 4, pp. 217-241). Elsevier.

Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. In *Chemical biology* (pp. 243-250). Humana Press, New York, NY.

El Asikri, M., Krit, S., & Chaib, H. (2020). Using Web Scraping In A Knowledge Environment To Build Ontologies Using Python And Scrapy. *European Journal of Molecular & Clinical Medicine*, *7*(03), 2020.

Haddaway, N. R. (2015). The use of web-scraping software in searching for grey literature. *Grey J*, *11*(3), 186-90.

Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Martinez-Hernandez, V., Sanchez, V., & Perez-Meana, H. (2018). A web scraping methodology for bypassing twitter API restrictions. *arXiv preprint arXiv:1803.09875*.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., ... & Bolton, E. E. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic acids research*, *47*(D1), D1102-D1109.

Mekki, K., Bajic, E., Chaxel, F., & Meyer, F. (2019). A comparative study of LPWAN technologies for large-scale IoT deployment. *ICT express*, *5*(1), 1-7.

Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. " O'Reilly Media, Inc.

Richard, A. M., Huang, R., Waidyanatha, S., Shinn, P., Collins, B. J., Thillainadarajah, I., ... & Tice, R. R. (2020). The Tox21 10K compound library: collaborative chemistry advancing toxicology. *Chemical Research in Toxicology*, *34*(2), 189-216.

Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research*, *37*(suppl_2), W623-W633.