```
for ID in users: #forming an entry
    for line in data[ID]:

        ext = ""

        clock = time[ID][data[ID].index(line)]
        USER = vk.method("users.get", {"user_ids":ID})[0] # getting user's ID
        USER = "%s %s" % (USER['first_name'], USER['last_name'])
        line = line.split(" ")

        matter_left = how_much_left[base[line[0]][:base[line[0]].index("(")]]

        line[2] = line[2].upper()
        if len(line[2]) == 1:
            if line[2] == "П": sight = "ПРИХОД"

        else:
            if line[2][0] == "Л":
                if line[2][1:].isdigit():
                    sight = "ЛАБОРАТОРНАЯ №%s" % line[2][1:]
                else:
                    sight = line[2]
```

Fig. 1. *Part of the code, which is forming an entry for logging*

for which this reagent was taken, or adding newly arrived reagents as a result of a purchase, transfer, or exchange.

Implementation of this solution in practice would have a positive effect on the efficiency of chemical accounting for at least two reasons: the information input speed and its correctness.

References

1. *Dokumentaciya VK_api (VK_api documentation). Available at: https://vk-api.readthedocs.io/en/latest/index.html (accessed 23 January 2022).*
2. *Tk documentation and tutorials. Available at: http://tkdocs.com (accessed 22 January 2022).*
3. *Tkinter documentation. Available at: https://docs.python.org/3/library/tkinter.html (accessed 22 January 2022).*

INTERPRETATION OF OLD CHEMICAL IMAGES

А. Н. Ассаад[1], К. Р. М. Мендоса[2], Ц. Ло[3], М. О. Патз[4]
Научный руководитель – к.х.н., доцент Н. Б. Шахова

National Research Tomsk Polytechnic University
634050, Russia, Tomsk, Lenin Avenue, 30, adonisassaad03@gmail.com[1], cristian.machuca.mendoza@gmail.com[2], liangzaihua@icloud.com[3], matheus_osmar@hotmail.com[4]

Introduction

With the continuous development of communication, computer, automation and other technologies, machine learning algorithms represented by deep learning have been widely used in image recognition, speech recognition, natural language processing, and other fields. The chemical information contained in scientific literature and network plays an important role in molecular structure retrieval and new drugs design.

Researchers use ChemDraw, ChemSketch, and other chemical mapping software to manually draw chemical structure images in the literature. On the other side, deep learning algorithms are used to automatically convert chemical structure images in the literature into InChI, SDF, which can be recognized by computers.

Unfortunately, a lot of data sets provided to the public are very small. This does not allow machine models to interpret them. In addition, images from the past do have a lot of corruption that leads

to a negligible performance by machines. Hence, development efforts by researchers will be slowed because of the necessity of manual work. Based on this, this project established an algorithm to interpret old chemical images to help chemists.

Recognition method of chemical structure diagram based on deep learning

The overall framework of the Bahdanau Attention Mechanism[1]method proposed in this paper is shown in Figure 1., which is divided into encoding and decoding parts. The encoder uses a pretrained Convolutional Neural Network (CNN) to encode chemical structure images into fixed-length high-dimensional feature vectors. In order to support CNN, Bahdanau Attention Mechanism is added to the back end of it, which translates InChI to binary, and can then process variable-length information sequences. The decoding side also uses Bahdanau Attention Mechanism to decode it into InChI.

Old images of chemicals in this project were provided and generated by the Bristol-Myers Squibb company. The images were synthetic but different in multiple sections, from distinct angles, various resolutions, and unalike noise levels. Each image was provided with its respective International Chemical Identifier (InChI). Four million pictures were used as data. 2.4 million of them were utilized as a learning data set to train the algorithm while the other 1,6 million were used for testing the data resulting from the algorithm.

The output of the data used as the testing set for the algorithm is considered the results of the experiment right after it is decoded to InChI. These results were compared to original data using a string metric named Levenshtein distance [2]. This string metric is a measurement to identify how different two strings are. The higher the Levenshtein distance number gets, the bigger the difference is between the strings. Consequently, the data results were 5.2 % different than the original data set. Specifically, the algorithm predicted and translated to InChI correctly 94.8 % of the given input.

Conclusion

In this paper, an end-to-end chemical image interpretation method by deep learning methods is proposed. This method does not need manual work; it directly identifies the old images gone through corruption. The algorithm is effective and needs to be developed furthermore.

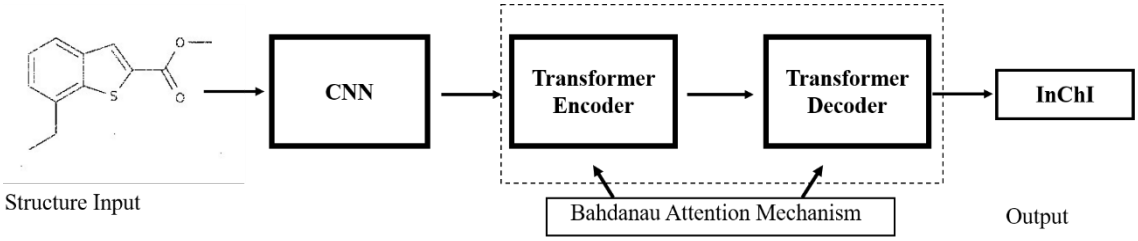

Fig. 1. *Chemical structure interpretation framework based on Bahdanau Attention Mechanism*

References

1. *Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).*
2. *Levenshtein, Vladimir I. "Binary codes capable of correcting deletions, insertions, and reversals." Soviet physics doklady, 1966. – Vol. 10. – № 8.*