GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 13 • Number 1 • 2019

http://dx.doi.org/10.7494/geom.2019.13.1.51

Małgorzata Pietras-Szewczyk*, Leszek Szewczyk*

Streamlining of Obtaining and Processing Data from Ogimet Service Using Python Language Script

Abstract: This paper contains a study of using Ogimet services as a source of meteorological data and the Python language script to streamline data processing. Meteorological data is important in large number of research projects in different disciplines of sciences and technology. In this case, it was used to analyze cloudiness, but it can also be used for energy, hydrology, and environment analyses.

Attention has been paid to the total cloudiness variability in an area of the Lower Silesia region in Poland during the time period of 2001–2010 using the data from eight synoptic stations (the data was obtained from the Ogimet service). A very important part of the work constituted Ogimet services as a source of free and easily available meteorological data. The biggest advantage of Ogimet is that the process of obtaining data is very easy and helpful in reducing the time needed to collect the data necessary in the research process. The offered data is free and available via the Internet, but it is raw and general. For these reasons, a Python script language application was made for faster and easier data processing. The script applied in this project has been described in detail in the work.

Finally, after processing the data, the daily averages of total cloudiness have been calculated based on the available data for eight meteorological stations. Next, the ten-year average for each day and month have been calculated. The results of the study were compared with works that took a longer data time period of total cloudiness into account.

Keywords: cloudiness, meteorological data, long-term average, obtaining data

Received: 3 April 2017; accepted: 1 September 2018

* University of Lower Silesia, Department of Technical Science, Wrocław, Poland

1. Introduction

The analysis of cloudiness is one of the most important issues in climate and environment research. Clouds are one of the factors that influences the thermal and radiation regime of the Earth. Increases in cloudiness lead to a reduction of solar radiation reaching the Earth's surface [1].

Climatologists treat cloudiness as one of the most interesting among other meteorological elements. Many of them describe the influence of the amount of cloudiness and the cloud types on global radiative forcing [2, 3]. Especially strong feedback on climate driven by clouds can be observed in the Arctic, where approximately 40% of its warming is due to cloudiness variability [4]. Several publications present the relationships between the observed worldwide long-term decrease in daily temperature ranges and the systematic increase of cloud cover; e.g., [5–12]. However, most investigations dealing with the variability of cloudiness concern continental areas.

In Poland, the variability of cloudiness has been analyzed by many authors. The most comprehensive study on this topic is the Atlas of Polish climate [13]. According to it, the differences between the sky coverage of particular regions in Poland are rather small.

Polish studies about cloudiness have a long history. Most studies concern on selected regions of Poland, although there are also studies analizing this parameter for the entire country [12]. The cloudiness variability has been completely analyzed in Krakow, for example. This was the longest series from the city (over 100 years) [14–16].

In the region of southwest Poland, this parameter was described by Dubicka on 35-year-old series data from the Wroclaw observatory. Other studies of cloudiness have been described in the Sudety and Karpaty Mountains [17]. However, during the last 20 years, there have been no published studies about cloudiness variability in the Lower Silesia region. This is the one of the reasons why this study is presented. This paper is dedicated to an analysis of the cloudiness over a ten-year time period (2001–2010) in the Lower Silesia region.

Another issue is connected with the problem of data obtaining and processing. For decades, obtaining and analyzing data describing cloudiness was very difficult and time-consuming. The results of the meteorological observations were noted in the meteorological dailies. Now, each meteorological station notes it into its computer database, thereby making meteorological data easily available for users. The only problem now is data access. A solution to this problem can be Internet services where the data from world meteorological stations is stored and shared with users. An example of this kind of service is Ogimet. Thanks to it, users have access to meteorological data for every daytime period, and they can obtain data in a suitable format. As a result, the process of obtaining data is no longer hard and time-consuming.

2. Objectives, Data, and Methodology

In this paper, the data from eight meteorological stations from southwest Poland has been taken into account. The locations of the synoptic stations are presented in Figure 1.

In these stations, general observations of cloudiness are conducted. The stations are located in Jelenia Góra, Kalisz, Kłodzko, Legnica, Leszno, Opole, Wrocław, and Zielona Góra. The synopses from the synoptic stations from during the time period of 2001–2010 are available online from Ogimet [18].

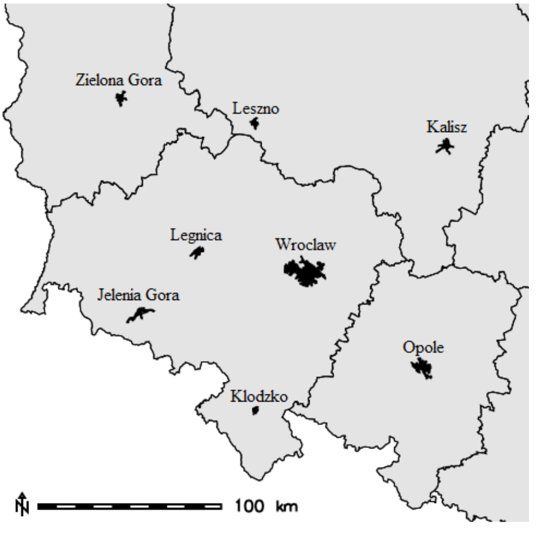

Fig. 1. Location of synoptic stations

2.1. About Ogimet

Ogimet is special weather information internet service that is available for users online at www.ogimet.com. Data access is free and available from the net, mainly from NOAA (the National Oceanic And Atmospheric Administration). The goal of this site is to provide data on current weather conditions in a fast and professional way.

Through Ogimet's website, users can obtain a great deal of meteorological data; for example, a daily weather summary that is based on synoptic report data, Meteosat-9 Infrared Image Satellite composition, and upper air soundingdata. The data is available for all of the meteorological stations that are part of WMO (the World Meteorological Organization).

The data-obtaining process is very simple: users must specify geographic locations and time ranges of the data. To make it simpler, Ogimet provides users with a special interface (Fig. 2) where they can indicate locales (for example, a city name) and a data time range. The next step is to choose the data format – HTML or text files.

Fig. 2. Ogimet online interface

Source: [18]

The advantage of the TXT data format is its easy data storage – the TXT format does not take much disc space, and it is suitable for many programs and databases. Such a format can be processed in MS Excel, for example.

2.2. Obtaining Data from Ogimet Service

The first step of obtaining synoptic data from Ogimet was downloading it in the TXT format. The next step is selecting the data describing the total cloudiness from the synoptic code. In the synoptic code, this data is concluded in a string of numbers. The first numbers inform the user about the station, date, and time of observation. The remainder of the numbers provide information about the weather.

On the basis of the available materials from the synoptic code, the averages of total cloudiness have been calculated for selected stations through the use of an authorial script, which was used to select the data describing the total cloudiness from a text file (TXT format). Next, the average values of the total cloudiness for each station are calculated.

2.3. The Applied Script

The data received from Ogimet was in a raw and general form. It required pre-processing like filtering and grouping (calculating the average value). Because of the data size (a few thousands lines), these were almost impossible to do manually. For this reason, a Python script language application was made for faster and easier data processing. The script used in this project has been described here in detail. Scripts written in Python are simple, fast-working, and easy to modify.

A dedicated Python script was introduced [19]. Python (https://www.python.org/) is a scripting language that provides a variety of functionalities. Because of its relatively simple syntax and clarity, it is widely used to process data in many research projects.

The created script reads all of the input files from Ogimet, processes their data, and outputs to .xls result files.

The Ogimet input data is in the TXT format and was analyzed in the following form:

200112311700 AAXX 31171 12400 42970 43004 11066 21088 39999 40248 51027 81038=
200112311600 AAXX 31161 12400 42675 73004 11061 21093 39992 40241 52030 8255/=
200112311500 AAXX 31151 12400 41575 73004 11061 21090 39983 40232 53027 70183 82438

After being processed, this data has the following form:

Date	Average
31-01-2001	7.92
30-01-2001	7.57
29-01-2001	7.76

The script applied in this project is easy to modify by users.

The most important script attributes are as follows:

– Pattern for searching data in subsequent rows expressed in regular expression syntax (for regex specification, see: [20]).
– In this case, we retrieve the following data: year, month, day, and measurement (in hours)

```
m=re.search(‘^(\d\d\d\d)(\d\d)(\d\d)\d{4}\s.{4}\s\d{5}\s\d{5}\s. {5}\s(\d{1})\d{4}\s.+$’, str(text))
```

– Date concatenation

```
cur_date = “{}-{}-{}”.format(groups[2], groups[1], groups[0])
```

– Measurement

```
value = float(groups[3])
```

– Grouping measurements by date

```
if(not date or cur_date == date):
    date = cur_date
    sum += value
    counter += 1
```

– Writing measurements to resulting .xls file:

```
output_sheet.write(output_row_counter, 0, date)
output_sheet.write(output_row_counter, 1, format(sum/counter, ‘.2f’))
```

– The script can be started by the following command:

```
python files_processor.py <dir_path_with_xls_files>
```

The Python language was applied in this project for the reason that it is very simple in use. Users using the instructions given here can make their own scripts.

3. Results

3.1. Total Cloudiness in Lower Silesia

For the region of Lower Silesia, data had previously been taken into consideration from Wrocław and Mount Szrenica in Karkonosze. This data has been comprehensively described in the work of Prof. Maria Dubicka [1]. The values given in the aforementioned references are similar to the results obtained in this study. Highly consistent is both the annual course of the total cloudiness and annual volume.

As described by Dubicka, the average value of the total cloudiness in Wrocław during the period of 1946–1980 was 70% (1994), while the value of the total cloudiness was 68% during the period of 1901–1980 [17]. The value calculated for Wrocław during the period of 2001–2010 was 68%.

The results of the study were compared with the works that have taken into account longer data time periods of the total cloudiness. In his work [21], Prof. Alojzy Woś developed data from the time period of 1951–2000. The resulting figures do not differ significantly from the studies taking into account the series of measurements. The greatest differences are visible in the case of Kalisz, Legnica, and Wrocław; however, these differences only reach a maximum of 5% of the average annual cloudiness.

3.2. Variability of Annual and Seasonal Total Cloudiness (in Octane) in Lower Silesia Region

Owing to the small number of measurement stations, the number of available studies about the cloud cover in Poland is very small. In the case of the province of Lower Silesia, such studies exist only for individual localizations, including Wrocław and Śnieżka Mount [1].

In this paper, data from eight stations is taken into account where general observations of the cloudiness are conducted. This includes data from Wrocław, Jelenia Góra, Zielona Góra, Kłodzko, Opole, Kalisz, Leszno, and Legnica. Based on the available data for each station, the daily averages of the total cloudiness have been calculated. Next, the ten-year average for each day and month have been calculated (Figs. 3, 4).

The mean annual cloudiness in southwest Poland ranges from 5.1 to almost 6 octane (Fig. 4). The areas with the least amounts of cloudiness are located in the southwestern and western parts of the Lower Silesia region (Jelenia Góra, Kłodzko, Zielona Góra). The cloudiest areas of that region is in the northern part (Legnica).

Generally, the northern parts of the analyzed region are slightly cloudier in autumn than they are in the south; however, the difference in the cloudiness between these areas is not large. A cloudier sky during the winter season is one the characteristic features of the amount of changes of the total cloudiness in Poland [22]. The months with the greatest amounts of cloudiness in southwest Poland are November, December, and January. In December, only Jelenia Góra is characterized by a total cloudiness that is below 6 octane, while the cloudiness observed in the remaining areas of the region is greater than this value.

The cloudiness is slightly above 5 octane in March in the majority of the analyzed area. In April, we observe the minimum cloud cover within the whole year (slightly above 4 octane).

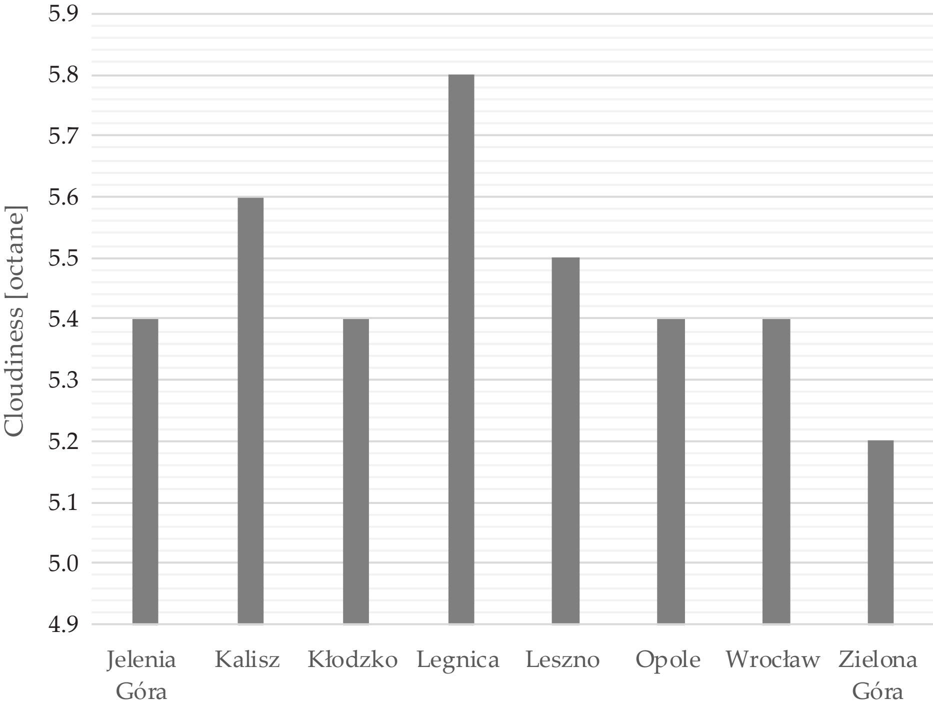

Fig. 3. Mean annual cloudiness in southwest Poland for 2001–2010

The monthly averages have a maximum value during the winter and a minimum in the summer. For all stations, the maximum value of cloudiness was during December – above 6 octane. For Wrocław, the same value occurred in January (6.2). The highest cloudiness during the winter occurred in Kalisz (6.9 octane), while this value is the lowest in Jelenia Góra (5.8 octane). The minimum value of the monthly average of cloudiness occurs in April. In that month, the monthly averages of the cloudiness was less than 5 octane in six of the eight stations. The lowest value was for Jelenia Góra (4.3 octane in April).

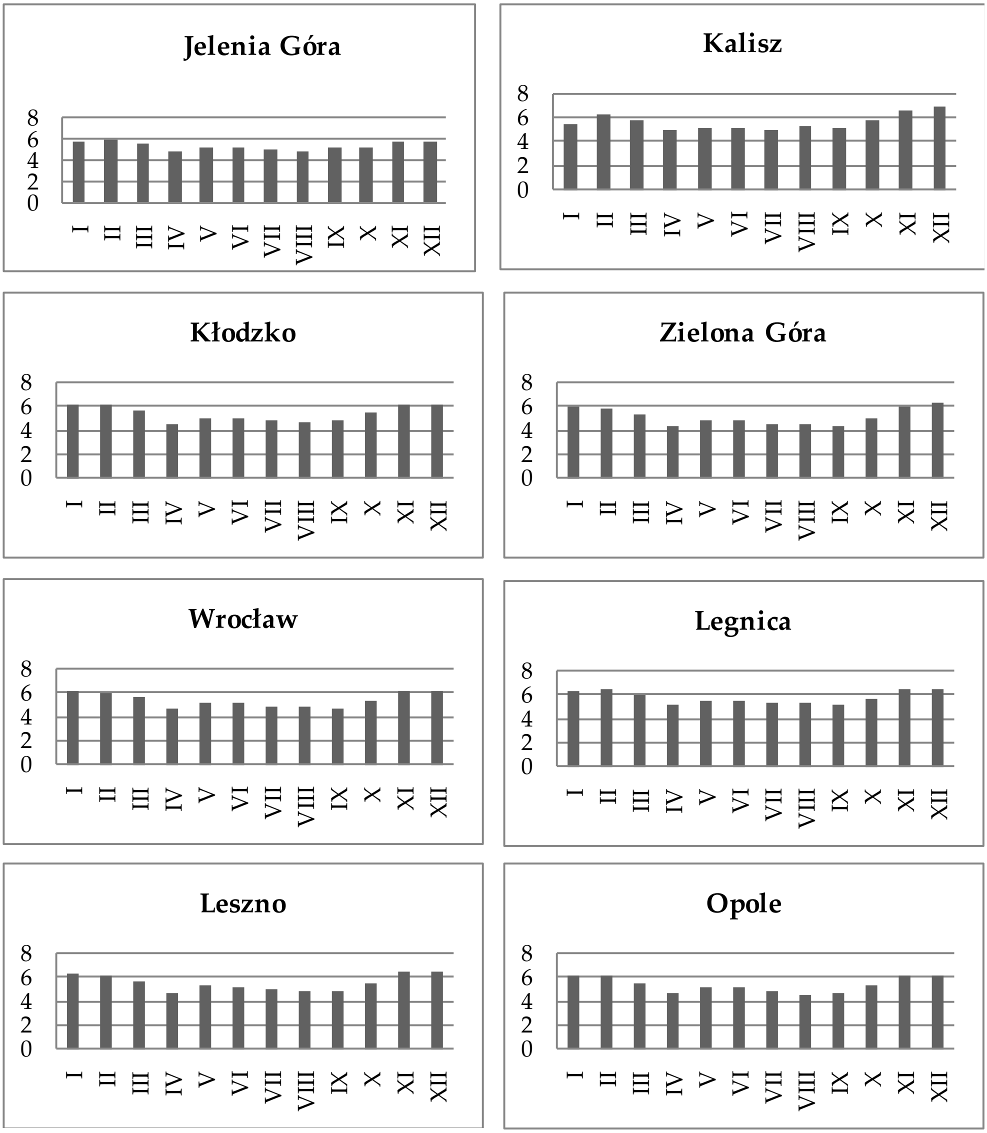

Fig. 4. Monthly average cloudiness for selected cities for 2001–2010

The Table 1 presents the monthly and annual averages of the total cloudiness for each station's analysis in this study compared to the long-term observations made by Prof. Aloysius Woś. In his work from 2010, he developed the data from the time period of 1951–2000 for all stations in Poland. The resulting figures do not differ significantly from the studies taking into account the series of measurements. The greatest differences are visible in the cases of Kalisz, Legnica, and Wrocław; however, the maximum differences only reach 5%.

Table 1. Monthly and annual average cloudiness for 2001–2010

City	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Year
Jelenia Góra	5.7	6.0	5.6	4.8	5.2	5.3	5.0	4.9	5.3	5.3	5.8	5.7	**5.4**
Kalisz	5.5	6.3	5.7	5.0	5.2	5.1	4.9	5.3	5.1	5.7	6.5	6.9	**5.6**
Kłodzko	6.1	6.1	5.6	4.6	5.0	5.0	4.8	4.7	4.8	5.5	6.1	6.2	5.4
Legnica	6.3	6.4	6.0	5.1	5.5	5.5	5.4	5.3	5.1	5.7	6.5	6.5	**5.8**
Leszno	6.3	6.2	5.7	4.7	5.3	5.2	5.0	4.9	4.8	5.5	6.4	6.4	**5.5**
Opole	6.1	6.1	5.5	4.7	5.1	5.2	4.9	4.6	4.7	5.3	6.1	6.2	**5.4**
Wrocław	6.2	6.0	5.6	4.7	5.1	5.2	4.9	4.8	4.7	5.4	6.1	6.2	**5.4**
Zielona Góra	6.0	5.9	5.4	4.3	4.8	4.8	4.6	4.6	4.4	5.0	6.0	6.3	**5.2**

4. Summary

The values of the mean amount of annual total cloudiness in Lower Silesia during the period of 2001–2010 are below 6. This value is higher than the mean from all of Poland, which is below 5 octane in the lowlands to above 6 octane in the mountains [22].

The value of the cloudiness in the yearly seasons are similar to the values for Poland. The lowest amount of cloudiness in Lower Silesia occurs during summer (Fig. 4), while the greatest occurs in the winter. The least cloudy months in this region are April and August, whereas November and December are the cloudiest months (Fig. 4). In general, the amount of cloudiness decreases in the spring and summer. In the autumn and winter, the total cloudiness amount increases.

However, the western part of Lower Silesia is cloudier than the eastern part of the country. This is probably caused by the influence of the regional climatic factor related to the effect of the Atlantic Ocean, the primary source of humid masses of polar-maritime air. When this type of air mass flows into Poland, there is a characteristic increase in the amount of cloudiness. An increased cloudiness in waist extended from Legnica to Kalisz is probably related to the predominant type of air circulation – west and southwest in this part of Poland.

The following results of the studies were compared to works that have taken the data from a longer time period of total cloudiness into account. The average long-term general sky cloud amount shown in Figure 5. The resulting figures do not differ significantly from the studies taking into account the series of measurements. The greatest differences are visible in the cases of Kalisz, Legnica, and Wrocław; however, these only reach a maximum of 5% of the average annual cloudiness.

The very important parts of the work are constituted by the Ogimet service as a source of free and easily available meteorological data as well as a Python script to process data in faster and easier way. As it was mentioned, the process of obtaining data is very easy and helps reduce the amount of time needed to collect the data necessary in the research process.

References

[1] Dubicka M.: *Wpływ cyrkulacji atmosfery na kształtowanie warunków klimatu (na przykładzie Wrocławia)*. Acta Universitatis Wratislaviensis. Studia Geograficzne, 60, Wyd. Uniwersytetu Wrocławskiego, Wrocław 1994.

[2] Houghton J.T., Ding Y., Griggs D.J., Noguer M., van der Linden P.J., Dai X., Maskell K., Johnson C.A. (eds.): *Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the IPCC.* Cambridge University Press, Cambridge 2001.

[3] Solomon S., Qin D., Manning M., Chen Z., Marquis M., Averyt K.B., Tignor M., Miller H.L. (eds.): *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the IPCC.* Cambridge University Press, Cambridge 2007.

[4] Vavrus S.: *The impact of cloud feedbacks on arctic climate under greenhouse forcing.* Journal of Climate, vol. 17, 2004, pp. 603–615.

[5] Henderson-Sellers A.: *Increasing cloud in a warming world*. Climatic Change, vol. 9, issue 3, 1986, pp. 267–309.

[6] Dessens J., Buecher A.: *Changes in minimum and maximum temperatures at the Pic du Midi in relation with humidity and cloudiness. 1882–1984*. Atmospheric Research, vol. 37, 1995, pp. 147–162.

[7] Karl T.R., Jones P.D., Knight R.W., Kukla G., Plummer N., Razuvayev V., Gallo K.P., Lindseay J., Charlson R.J., Peterson T.C.: *A new perspective on recent global warming: Asymmetric trends of daily maximum and minimum temperature*. Bulletin of the American Meteorological Society, vol. 74, 1993, pp. 1007–1023.

[8] Miętus M., Owczarek M.: *Variability of mean annual amplitude of air temperature in relation to the mean cloudiness on the Polish coast since 1945*. [in:] Heino R. (ed.), *Climate variations in Europe: proceedings of the European Workshop on Climate Variations held in Kirkkonummi (Majvik), Finland, 15–18 May 1994*, Painatuskeskus, Helsinki 1994, pp. 128–135.

[9] Kaas E., Frich P.: *Diurnal temperature range and cloud cover in the Nordic countries: observed trends and estimates for the future*. Atmospheric Research, vol. 37, 1995, pp. 211–228.

[10] Dai A., Trenberth K.E., Karl T.R.: *Effects of clouds soil moisture precipitation and water vapour on diurnal temperature range*. Journal of Climate, vol. 12, 1999, pp. 2451–2473.

[11] Sun B., Groisman P.Y., Bradley R.S., Keimig F.T.: *Temporal changes in the observed relationship between cloud cover and surface air temperature*. Journal of Climate, vol. 13, 2000, pp. 4341–4357.

[12] Żmudzka E.: *Wielkość zachmurzenia w Polsce w drugiej połowie XX wieku.* Przegląd Geofizyczny, t. 3–4, nr 3–4, 2003, pp. 159–185.

[13] Lorenc H. (red.): *Atlas klimatu Polski*. Instytut Meteorologii i Gospodarki Wodnej, Warszawa 2005.

[14] Matuszko D.: *Wieloletni przebieg częstości rodzajów chmur w Krakowie (1906–1995)*. [in:] Dubicki A. (red.), *Zmiany i zmienność klimatu Polski: ich wpływ na gospodarkę, ekosystemy i człowieka: ogólnopolska konferencja naukowa, Łódź 4–6 listopada 1999*, "BGW", Warszawa 1999, pp. 175–179.

[15] Miętus M., Filipiak J., Owczarek M.: *Czasowo-przestrzenna struktura opadów w rejonie Zatoki Gdańskiej i jej możliwe zmiany w skali XXI wieku.* [in:] Cyberski J. (red.), *Powódź w Gdańsku 2001*, Wyd. Gdańskie, Gdańsk 2003, pp. 35–55.

[16] Morawska-Horawska M.: *Cloudiness and sunshine in Cracow, 1861–1980, and its contemporary tendencies*. International Journal of Climatology, vol. 5, issue 6, 1985, pp. 633–642.

[17] Dubicka M., Limanówka D.: *Zmienność zachmurzenia i usłonecznienia w Sudetach i Karpatach oraz na ich przedpolu*. [in:] Pyka J.L. (red.), *Klimat Dolnego Śląska i krain polarnych*, Acta Universitatis Wratislaviensis, 1590, Prace Instytutu Geograficznego. Meteorologia i Klimatologia. Seria C, t. 1, Wyd. UW, Wrocław 1994, pp. 45–60.

[18] Ogimet, http://www.ogimet.com/ [access: 23.01.2017].

[19] Python, https://www.python.org/ [access: 23.01.2017].

[20] Regular expression, https://en.wikipedia.org/wiki/Regular_expression [access: 23.01.2017].

[21] Woś A.: *Klimat Polski w drugiej połowie XX wieku.* Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań 2010.

[22] Filipiak J., Miętus M.: *Spatial and temporal variability of cloudiness in Poland, 1971–2000*. International Journal of Climatology, vol. 29, 2009, pp. 1294–1311.

Ułatwienie procesu pozyskiwania i przetwarzania danych z serwisu Ogimet

Streszczenie: Artykuł poświęcony jest wykorzystaniu usługi Ogimet jako źródła danych meteorologicznych opisujących zachmurzenie oraz skryptu w języku Python do optymalizacji procesu przetwarzania pozyskanych danych.

Dane meteorologiczne są istotne w wielu zagadnieniach badawczych z różnych dyscyplin nauki i techniki. W tym przypadku dane zostały wykorzystane do analizy wielkości zachmurzenia. Z równym powodzeniem opisane narzędzia mogą być wykorzystane w innych dziadzinach, takich jak hydrologia, ochrona środowiska czy energetyka.

Zasadniczym elementem pracy jest opis usługi Ogimet jako źródła wolnych i łatwo dostępnych danych meteorologicznych. Największą zaletą serwisu jest prostote i szybkie pozyskiwanie danych. Oferowane dane są bezpłatne i dostępne przez Internet, ale są one surowe i ogólne. Z tego powodu zaproponowano użycie języka skryptowego Python do przetwarzania danych. Skrypt zastosowany w tym projekcie został szczegółowo opisany w pracy.
Po przetworzeniu danych, na podstawie dostępnych informacji z ośmiu stacji meteorologicznych, obliczono wartości średnich dobowych całkowitego zachmurzenia. Następnie obliczono średnie dziesięcioletnie dla każdej ze stacji. Wyniki zostały porównane danymi zawartymi w pracach, w których analizowano zachmurzenie w dłuższym okresie.

Słowa kluczowe: zachmurzenie, dane meteorologiczne, średnie długoterminowe, pozyskiwanie danych