A Novel Algorithm for analysis of Environmental database.

Mazhar Mahmoud Mohmed Hefnawi
Lecturer: Siting and Environmental. Egyptian Nuclear & Radiological Regulatory Authority
Egyptian Nuclear & Radiological Regulatory Authority, ENRRA Cairo, Egypt
mmaazz_2222@yahoo.co.uk

Abstract — Any environmental database system contains one or many environmental measurements such as temperature degree, humidity percentage, CO concentration, NOx concentration, etc. Any environmental measurement is recorded in the environmental database with its date and time. This paper introduces a new algorithm which represents any environmental database in the new form. This new form is considered as cognitive computation by storing all measurements values of one type with their number of duplicated and date and time for each reduplicated measurement value. All these measurements values are sorted in ascending order in the new form. With cognitive form the user can deal with this new form as dealing with the dictionary. The method for Looking for a ward in the dictionary is the same as the method used for looking for a measurement value in this new form. The output result from this cognitive computation is the number of happening of the measurement value and date and time for each happening. This new algorithm is a good tool for building an environmental reference for many sites around the world. This environmental reference can be used in siting evaluation processes, risk assessment processes, and environmental decision making processes.

Keywords - Algorithm; Data structure; Environmental database; advanced programing.

I. Introduction

Several ways exist to describe environmental databases. Environmental databases can be specialized to a specific type of environmental information like concentration data in environmental media. The terms environmental data, environmental information, environmental database, and environmental information system should be defined [1].

Environmental data are technical, spatial, and temporal data for the environmental media air, water, and soil. They pertain to questions of waste, noise, dangerous substances, fauna and lora, landscape, nature, and species conservancy. With the help of the analysis and interpretation of those data environmental information can be created [2].

An environmental database is a particular type of database that stores mainly environmental data. According to environmental informatics experts, a database can be called an ‘environmental database’ if the following three conditions are fulfilled:

1. The majority of data are environmental data.
2. A database system is used for the storage of these data.
3. The database is established as the basis for environmental uses and inquiries [2].

The proposed algorithm in this work is designed for dealing with the environmental database system which stores the temporal data such as meteorological data, seismic data, radiation data, air quality data, etc. This kind of databases has a huge amount of environmental data through several years. Therefore the probability for exiting the duplicated environmental data becomes very high and the needs for querying this data become necessary to assist many topics in the environmental studies as a reference material. All of database tools supported with any database system do not support this kind of cognitive [8, 9]. Therefore the introduced algorithm in this paper is designed to accomplish this kind of cognitive computation [8, 10]. The main jobs for this algorithm is as following

1. Storing the value for each temporal environmental data only one time.
2. The record which includes this value includes also the number of times for reduplicated of this value with date and time.

II. METHOD OF THE ALGORITHM

The algorithm's input data is from the environmental database which was described from the previous section. This data is expressed as an array of records [3]. Every record include date, time and value of the environmental measurement. The first process of this algorithm is generating a data structure object. This object is considered as a container which includes and stores the input data and final result [4, 10]. The algorithm for creating this object is as shown in Fig 1.

```
"N is assumed as a number of the records in the environmental data base"
Type Environmental_Database
Place_Name As TEXT
Environmental_Measurement As REAL
Date_of_Measurement [N] As DATE
Time_of_Measurement [N] As TIME
Num_of_rep As INTEGER
End Type
```

Figure 1.Data structure which is used by the algorithm.

The main idea for this algorithm is as following

1) Reading the first value of the environmental data from the created object.

2) Create an array of the record Environmental_Database and put the first value from step one into this array.

3) Looking for the values of the reset of the environmental data which are equal to the current value

4) Calculate the number of reduplicated data and register this number with date and time for each reduplicated data in the created array.

5) Repeat the first step for reading the next value and if the next value is the last value then go to step 9

6) Check the existing of the new value in the created array.

7) If it is exit then go to step 5.

8) If it is not exit then insert the new value in the array and go to step 3

9) Sort the contents of the created array in ascending order according to the value of the environmental data and End the algorithm.

The pseudocode of the proposed algorithm is as shown in Fig. 2

```
Input: A sequence of n records of kind of Environmental_Database (a1,a2…….an).
Output: A permutation (reordering) of m records of kind of Environmental_Database (b1,b2…….bm)
            of the input sequence such that  b1. Environmental_Measurement
            <b2.Environmental_Measurement<…….. bm.Environmental_Measurement.
Var
Size : Integer
Begin
        Size=0
        For i = 1 To N
           Search (A[i]. Environmental_Measurement,i)
        Next i
       Quicksort(b)
End.
Procedure Search(x :Real,index: Integer)
Begin
       If Size > 0 Then
          Ch = Check_value(x)
       Else
           Ch = True
      End If
     If Ch = False Then
         Exit Sub
  Else
        Size = Size + 1
        b[Size]. Environmental_Measurement = x
        b[Size].Num_of_rep = 1
        b[Size]. Date_of_Measurement[1] =   a[index]. Date_of_Measurement[1]
        b[Size]. Time_of_Measurement[1] =   a[index]. Time_of_Measurement[1]
       For i = index To N_end
         If a[i]. Environmental_Measurement = x Then
         Begin
                b[Size].Num_of_rep = b[Size].Num_of_rep + 1
                b[Size]. Date_of_Measurement [Num_of_rep]=   a[i]. Date_of_Measurement[1]
                b[Size]. Time_of_Measurement[Num_of_rep] =   a[i]. Time_of_Measurement[1]
        End
     Next i
  End If
End Sub
Function Check_value(value1 :Integer) As Boolean
Begin
       For i = 1 To Size
         If b[i]. Environmental_Measurement = value1 Then
         Begin
                Check_value = False
                Exit Function
         End
       Next i
      Check_value = True
End
```

Figure 2. The pseudocode of the proposed algorithm.

III. RESULTS AND DISSICUSIONS

The proposed algorithm in this paper is implemented for the environmental database system which stores the temporal data of Gamma radiation levels in Cairo city at year 2009.the structure of this environmental database system is considered as a matrix. Every column of this matrix represents a certain time in the day and every row represents a certain day in year 2007[5]. The gamma radiation levels are recorded every 15 minutes in this database. Therefore the dimension of this matrix is 365 row x 96 columns. Figure 3 shows a small part of this matrix.

Station ID : 11

Gamma Levels Measurements from 01/ 01/ 2007 to 31/ 12 /2007

Measurements unit For Gamma Levels is micro Sv/h

time	0:00	0:15	0:30	0:45	1:00	1:15	1:30	1:45	2:00	2:15	2:30	2:45	3:00	3:15	3:30	3:45	4:00	4:15	4:30	4:45	5:00	5:15	5:30	5:45	6:00	6:15	6:30	6:45	7:00	7:15	7:30	7:45	8:0
01/01/07	0.03	0.03	0.05	0.05	0.03	0.03	0.04	0.05	0.03	0.04	0.03	0.04	0.04	0.04	0.05	0.04	0.06	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.03	0.04	0.03	0.04	0.04	0.05	0.04	0.03	0.0
01/02/07	0.04	0.04	0.04	0.04	0.03	0.03	0.05	0.05	0.04	0.03	0.02	0.03	0.03	0.04	0.04	0.05	0.03	0.03	0.04	0.03	0.03	0.04	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.03	0.04	0.0
01/03/07	0.03	0.04	0.04	0.04	0.04	0.06	0.04	0.03	0.05	0.04	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.06	0.03	0.03	0.04	0.03	0.04	0.05	0.03	0.04	0.03	0.03	0.05	0.04	0.0
01/04/07	0.04	0.03	0.04	0.03	0.04	0.03	0.05	0.04	0.03	0.04	0.03	0.04	0.05	0.05	0.04	0.03	0.04	0.04	0.04	0.03	0.05	0.03	0.03	0.03	0.05	0.04	0.03	0.04	0.03	0.05	0.03	0.04	0.0
01/05/07	0.05	0.03	0.03	0.03	0.04	0.05	0.04	0.05	0.04	0.03	0.04	0.03	0.06	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.03	0.06	0.04	0.03	0.05	0.03	0.04	0.04	0.04	0.0
01/06/07	0.02	0.03	0.03	0.05	0.03	0.04	0.04	0.03	0.05	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.03	0.04	0.04	0.05	0.05	0.0
01/07/07	0.04	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.04	0.05	0.04	0.04	0.04	0.04	0.05	0.03	0.04	0.03	0.03	0.03	0.03	0.04	0.04	0.05	0.05	0.04	0.03	0.03	0.05	0.05	0.05	0.03	0.0
01/08/07	0.03	0.03	0.04	0.04	0.04	0.05	0.03	0.04	0.03	0.03	0.04	0.03	0.04	0.04	0.04	0.05	0.03	0.04	0.04	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.0
01/09/07	0.04	0.05	0.04	0.04	0.04	0.05	0.04	0.05	0.04	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.05	0.04	0.04	0.03	0.06	0.05	0.04	0.03	0.04	0.03	0.05	0.03	0.04	0.04	0.03	0.0
01/10/07	0.04	0.05	0.04	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.03	0.05	0.04	0.05	0.04	0.04	0.03	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.03	0.05	0.0
01/11/07	0.03	0.03	0.04	0.04	0.05	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.04	0.03	0.05	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.03	0.04	0.04	0.05	0.04	0.05	0.04	0.04	0.0
01/12/07	0.04	0.05	0.03	0.05	0.03	0.04	0.03	0.04	0.06	0.02	0.04	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.03	0.05	0.04	0.04	0.04	0.03	0.03	0.05	0.04	0.04	0.0
01/13/07	0.04	0.04	0.05	0.05	0.03	0.04	0.04	0.04	0.04	0.06	0.05	0.03	0.03	0.05	0.04	0.04	0.05	0.04	0.03	0.05	0.04	0.03	0.04	0.04	0.03	0.05	0.03	0.03	0.04	0.03	0.03	0.04	0.0
01/14/07	0.05	0.03	0.03	0.05	0.04	0.05	0.03	0.04	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.04	0.04	0.03	0.03	0.05	0.06	0.02	0.03	0.04	0.05	0.03	0.03	0.04	0.0
01/15/07	0.03	0.04	0.04	0.05	0.04	0.04	0.03	0.04	0.03	0.05	0.03	0.04	0.04	0.03	0.04	0.03	0.04	0.05	0.04	0.05	0.02	0.04	0.05	0.04	0.06	0.05	0.05	0.04	0.05	0.04	0.05	0.02	0.0
01/16/07	0.04	0.03	0.04	0.04	0.04	0.04	0.03	0.05	0.04	0.05	0.04	0.04	0.05	0.03	0.04	0.05	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.03	0.03	0.06	0.04	0.0
01/17/07	0.05	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.02	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.03	0.04	0.04	0.03	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.04	0.03	0.04	0.03	0.05	0.0
01/18/07	0.04	0.04	0.03	0.05	0.05	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.05	0.03	0.04	0.02	0.03	0.05	0.04	0.04	0.02	0.03	0.04	0.05	0.04	0.04	0.03	0.04	0.03	0.0
01/19/07	0.02	0.05	0.05	0.05	0.04	0.05	0.03	0.03	0.04	0.03	0.04	0.03	0.05	0.03	0.03	0.05	0.03	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.03	0.04	0.04	0.04	0.0
01/20/07	0.05	0.03	0.04	0.03	0.03	0.05	0.03	0.04	0.04	0.04	0.04	0.03	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.03	0.04	0.03	0.03	0.02	0.05	0.04	0.05	0.03	0.04	0.04	0.03	0.05	0.0
01/21/07	0.03	0.03	0.04	0.05	0.04	0.03	0.04	0.05	0.05	0.04	0.05	0.03	0.03	0.04	0.03	0.04	0.04	0.04	0.03	0.04	0.03	0.03	0.04	0.04	0.05	0.03	0.03	0.03	0.03	0.04	0.03	0.04	0.0
01/22/07	0.06	0.04	0.04	0.04	0.04	0.05	0.04	0.03	0.04	0.03	0.03	0.03	0.03	0.05	0.04	0.05	0.03	0.04	0.05	0.03	0.03	0.04	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.04	0.0
01/23/07	0.03	0.03	0.05	0.04	0.04	0.03	0.04	0.03	0.04	0.06	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.05	0.04	0.04	0.06	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.04	0.04	0.0
01/24/07	0.04	0.04	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.02	0.05	0.04	0.04	0.04	0.04	0.04	0.05	0.03	0.05	0.04	0.04	0.03	0.04	0.03	0.04	0.0
01/25/07	0.03	0.04	0.03	0.04	0.04	0.05	0.03	0.02	0.04	0.05	0.05	0.03	0.04	0.04	0.04	0.03	0.04	0.05	0.05	0.05	0.03	0.04	0.04	0.03	0.06	0.04	0.04	0.05	0.03	0.03	0.05	0.04	0.0
01/26/07	0.03	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.04	0.05	0.05	0.04	0.05	0.04	0.05	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.04	0.05	0.03	0.05	0.03	0.04	0.04	0.04	0.0
01/27/07	0.04	0.04	0.03	0.03	0.04	0.05	0.03	0.03	0.04	0.04	0.05	0.04	0.05	0.04	0.04	0.04	0.05	0.04	0.03	0.04	0.04	0.04	0.05	0.04	0.03	0.04	0.05	0.04	0.02	0.03	0.04	0.03	0.0
01/28/07	0.03	0.04	0.04	0.03	0.04	0.04	0.05	0.04	0.03	0.03	0.04	0.04	0.04	0.03	0.04	0.05	0.04	0.04	0.03	0.05	0.04	0.04	0.03	0.04	0.03	0.04	0.02	0.03	0.04	0.05	0.05	0.04	0.0
01/29/07	0.04	0.04	0.04	0.03	0.04	0.04	0.03	0.05	0.04	0.03	0.05	0.04	0.04	0.03	0.04	0.03	0.05	0.03	0.03	0.03	0.04	0.05	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.0

Figure 3. A part of Gamma radiation levels in Cairo city at year 2007[5].

As mentioned in pervious section the output result is stored in array. The size of this array is 682 records while the size of input data is equal to 365 multiplied by 96 which are 35040 records. This means that about 2% of input data represent all gamma levels recorded in Cairo city at 2007 without any duplication. Table 1 shows a part of this array. All the gamma levels in this array are stored in ascending order with their number of duplications. Table II shows in details the date and time for their duplication of the Gamma levels who's Id is 12 at table I.

With this array which represents the new form the user can deal with it as dealing with the dictionary. The method for Looking for a ward in the dictionary is the same as the method used for looking for a measurement value in this new form. The output result from this new form is the number of happening of the measurement value and date and time for each happening.

TABLE I. THE MAIN CONSTRUCTION OF THE NEW FORM OF THE INPUT DATA

ID	Gamma level (micro Sv/h)	Number of duplications
1	0	1
2	2.10E-03	2
3	7.05E-03	1
4	1.15E-02	1
5	1.19E-02	1
6	1.20E-02	1
7	1.30E-02	1
8	1.36E-02	1
9	1.36E-02	1
10	1.37E-02	1
11	1.42E-02	3
12	1.48E-02	5
13	1.53E-02	1
14	1.53E-02	1
15	1.59E-02	1
16	1.65E-02	3
17	1.70E-02	1
18	1.70E-02	1
19	1.71E-02	3
20	1.75E-02	1
21	1.76E-02	12

TABLE II. DATE AND TIME FOR EVERY DUPLICATION OF THE GAMMA LEVELS WHOSE ID IS 12 AT TABLE I

ID	Gamma level (micro Sv/h)	Number of duplications	Date	Time
12	1.48E-02	5	03/10/07	5:45:00 AM
			04/20/07	4:00:00 PM
			05/22/07	4:45:00 PM
			07/15/07	11:15:00 AM
			09/22/07	2:15:00 PM

Fig. 4 shows the relationship between all of the Gamma radiation levels and numbers of times for their duplication for Cairo city at year 2007. This figure shows that most of the duplicated data is found in range between 2.00 E-2 and 6.00 e-2 micro Sv/h. Fig. 5 shows in more details this range.

IV. CONCLUSIONS

1. The proposed algorithm in this work is designed for dealing with the environmental database system which stores the temporal data such as meteorological data, seismic data, radiation data, air quality data, etc.

2. This kind of databases has a huge amount of environmental data through several years. Therefore the probability for exiting the duplicated environmental data becomes very high and the needs for querying this data become necessary to assist many topics in the environmental studies as a reference material.

3. All of database tools supported with any database system do not support this kind of cognitive computation. Therefore the introduced algorithm in this paper is designed to accomplish this task.

4. This new algorithm is considered as a cognitive computation tool that can be used to build an environmental reference for many sites around the world [1, 6].

5. This environmental reference can be used in siting evaluation processes, risk assessment processes, and environmental decision making processes [2, 7].

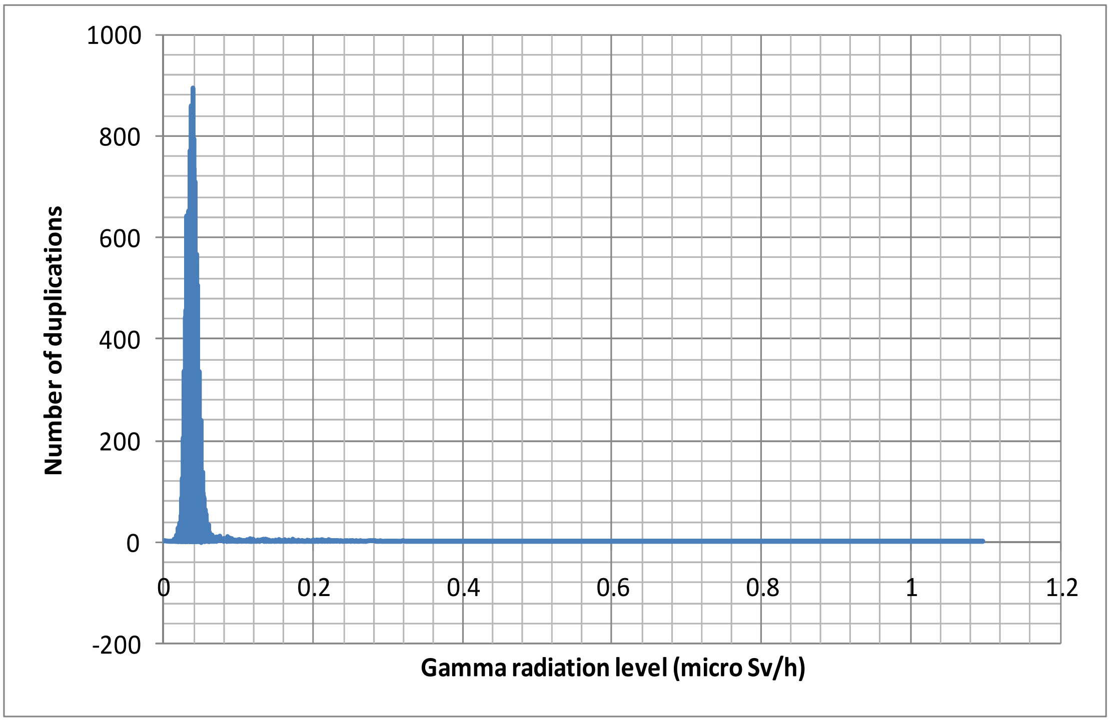

Figure 4. Relationship between all of the Gamma radiation levels and numbers of times for their duplication for Cairo city at year 2007.

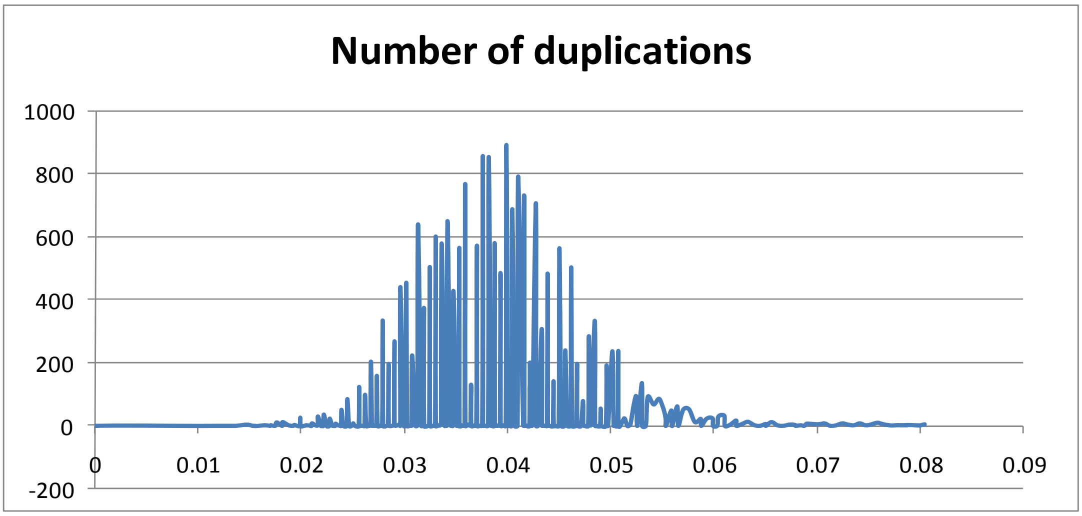

Figure 5. Most of duplicated data from figure 2 in more details.

References

[1] N. M. Avouris And B. Page, ‘Environmental Informatics, Methodology And Applications Of Environmental Information Processing, Introduction’, Eds. N. M. Avouris And B. Page, Euro Courses, Computer And Information Science, Vol. 6, Kluwer, Dordrecht, 1995

[2] D. Orton, ‘Database Reviews: Environmental’, Ed. D. Orton, Online Searching In Science And Technology, The British Library, London, 1995, Pp. 75 82.

[3] Milton Abramowitz And Irene A. Stegun, Editors. Handbook Of Mathematical Functions. Dover, 1965

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] Abou-Bakr M. Ramadan , and Mazhar M. Hefnawi “Adaptation of a data Acquisition System For Monitoring Air Quality And Radioactivity,” Arab journal of Nuclear Sciencesb and application 2006, 39(1).

[6] Spielman, Seth E., and Jean-Claude Thill. "Social area analysis, data mining, and GIS." Computers, Environment and Urban Systems 32.2 (2008): 110-122.

[7] Li, Deren, and Shuliang Wang. "Spatial data mining and knowledge discovery."Advances in Spatio-Temporal Analysis 5 (2007): 173.

[8] Huang, Kaizhu, et al. "Special Issue Editorial: Cognitively-Inspired Computing for Knowledge Discovery." Cognitive Computation 10.1 (2018): 1-2.

[9] Fazzolari, Michela, et al. "A study on text-score disagreement in online reviews." Cognitive Computation 9.5 (2017): 689-701.

[10] Cambria, E., Chattopadhyay, A., Linn, E., Mandal, B., & White, B. (2017). Storages Are Not Forever. Cognitive Computation, 9(5), 646-658.