

USE OF DESCRIPTOR FINGERPRINTS TO SIMILARITY AND CLUSTER ANALYSIS OF BIODIESEL FUELS ON THE BASIS OF THEIR COMPOSITION AND PROPERTIES.

INTRODUCTION OF FUZZY DESCRIPTOR FINGERPRINTS.

Ivan P. Bangov^{1*} Marina N. Moskovkina¹, Zilya A. Mustafa², Rумыana S. Milina²

¹General Chemistry Chair, Faculty of Natural Sciences, Konstantin Preslavski University of Shumen, 115 Universitetska Str., Shumen 9712, Bulgaria, ²Central Research Laboratory, Faculty of Natural Science, Prof. Assen Zlatarov University, 1 Prof. Yakimov Str., Burgas 8010, Bulgaria.

* Author to whom correspondence should be addressed.

ivan.bangov@gmail.com

ABSTRACT: A novel approach for similarity search of different objects, *fuzzy descriptor fingerprints* has been devised. Each cluster is characterized by a fuzzy descriptor fingerprint which is formed of real values instead of Boolean values 0 =false and 1 =true. A novel similarity index for similarity comparison between Boolean and fuzzy fingerprints, *Fuzzy Factor* (FF) has been introduced. This method was exemplified with similarity and cluster analysis of biodiesel fuels.

Keywords: fingerprints, fuzzy descriptor fingerprints, clustering, Fuzzy Factor (FF), biodiesels.

INTRODUCTION

The present paper introduces a novel chemometric approach of a new type of descriptor fingerprints - *fuzzy descriptor fingerprints* applied to the problem of classification of biodiesel fuels. Descriptor fingerprints approach devised by one of the authors (IB)(1) has been previously successfully applied to the solution of the problem of discrimination between allergens and non-allergens (2) and clustering of biodiesels (3) by using the Butina method (4). With the present introduction of the fuzzy descriptor fingerprints, each cluster is assigned with a fuzzy fingerprint.

Biodiesel fuel is composed by methyl esters of long chain fatty acids (FAME) produced from plant oils, animal fats and other lipids [5-7]. As a "green" fuel biodiesel has a series of advantages over petro-diesel fuels such as their derivation from renewable feed-stock, biodegradability, non-toxic and essentially free of metals, sulfur, carcinogenic aromatics fuel as well as low greenhouse effect and a positive energy balance.

Feed-stock availability for biodiesel production depends on the geography, climate and economics of different countries. At present, the dominant feed-stock (about 80 %) is vegetable oils, namely soy bean oil in USA, rapeseed and sunflower oil in Europe and palm oil in Southeast Asia. Other feed-stock having real or potential commercial interest are animal fats, non-edible and waste oils. Traditional for Bulgaria feed-stock are sunflower and rapeseed oils.

Since biodiesel is a mixture of *Fatty Acid Methyl Esters* (FAME), its properties depend on the chemical structure of the individual FAME and their contents (FAME profile). FAME profiles of biodiesel are influenced by the stocks and origin of the oils used [5,7,8] and can be obtained by chromatographic methods [9-13] providing valuable multi-component information. So, FAME profiles appears to be an instrument for a selection of feed-stock to produce fuels with certain properties [8,14], for investigations [15,16], and for fuel spillage and remedial actions in the environment [17,18].

THE DESCRIPTOR FINGERPRINT APPROACH.

Fingerprints in Chemoinformatics are primarily developed to characterize chemical structures within the various similarity search procedures [19-21]. The latter are usually formed in two ways. In the most popular way an array of structural fragments (structural keys), e.g. 1024 is created. A fingerprint is a string or binary array of 1s and 0s. The elements of the fragment array are juxtaposed to the fingerprint array elements having one-to-one correspondence between fragments and the array elements (see Figure 1). Each studied structure is further fragmented and analyzed for the presences or absence of fragments from the fragment array. In the case of presence of such a fragment a number (bit) 1 is put into the corresponding position (element) of the fingerprint string (bit) array and vice versa in the case of the absence of the corresponding fragment a 0 is put into the corresponding location within the fingerprint. It should be mentioned that the Daylight company has developed

another type of fingerprints - *hashed fingerprints*[22]. However this approach is out the scope of this work and will not be discussed here.

The fingerprints are usually employed in QSAR/QSPR for similarity search and clustering of structures. (Figure 1). Two chemical structures are compared for similarity by comparing their fingerprints by using any similarity measures available in literature [23,24]. For our further investigation we use the well-known Tanimoto [25] similarity index (TI) having the following form:

$$TI = N_C / (N_A + N_B - N_C) \quad (1)$$

where N_A is the number of *I*s in the first structure fingerprint, N_B - is the corresponding number of the second structure fingerprint, and N_C is the number of *I*s common to both structure fingerprints, i.e., *I*s being in the same position of the compared fingerprints. This index takes real values between 0.0 and 1.0. The larger is the value the most similar are the two structures. A comprehensive description of Tanimoto index is given in [23,24]. It is clear that the structural fingerprints incorporate Boolean logic, $I = \text{true}$ and $0 = \text{false}$.

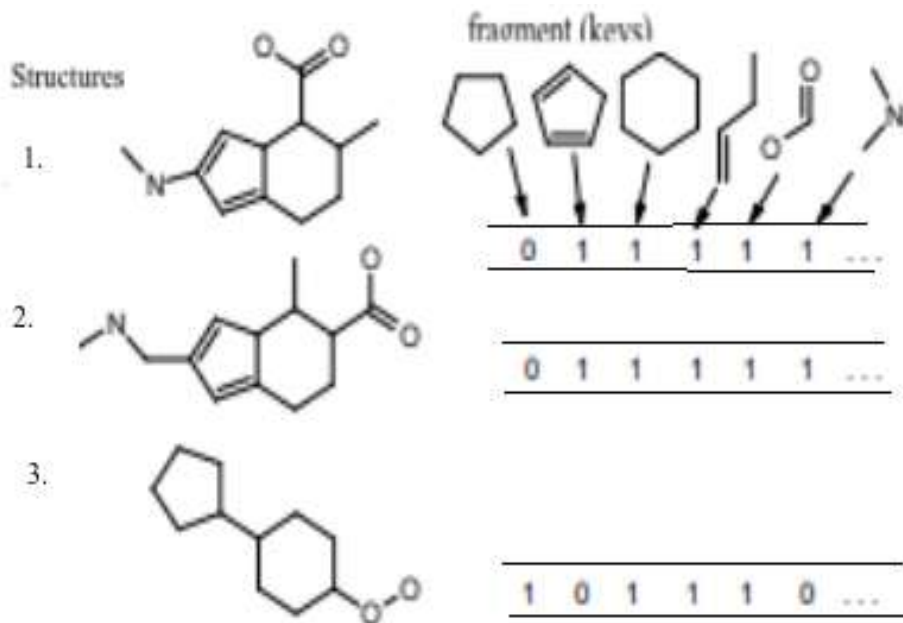


Figure 1. Formation of structural fingerprints for three structures.

Our idea was to employ descriptors instead of structure fragments in the formation of the fingerprints. A *descriptor fingerprint* is created by determination an interval for each descriptor within the fingerprint array and precision of its possible

values, e.g. an initial value (let us call it *initValue*) and ending value (*endValue*) and a precision step (resolution) *resValue*. Hence, for each descriptor its interval is divided into $N=(initValue-endValue)/resValue$ discrete fingerprint array elements (sub-intervals). The concatenation of all the descriptor elements forms the *descriptor fingerprint* itself. Further, in the course of descriptor creation for an object the program determines in which element (sub-interval) the current descriptor value falls by putting *1* in this element, the other descriptor elements remaining zeros. The formation of a descriptor fingerprint is illustrated in **Figure 2**.

This approach can be applied both to chemical structures and to various other chemical and non-chemical objects. It has been applied to discrimination between allergen/non-allergen food proteins [2], and to biodiesel fuels in our case. It obviously extends the area of application of the fingerprint method outside the chemical structure description, even outside chemistry, by using any physical, chemical, and biological as well as other user defined properties for descriptors of forming the descriptor fingerprint. In cases of chemical structures the structural fingerprints can also be concatenated to the descriptor fingerprints, thus both the chemical structure and its properties to be characterized by a common fingerprint. Accordingly, it allows the inclusion of all information available to the similarity perception process.

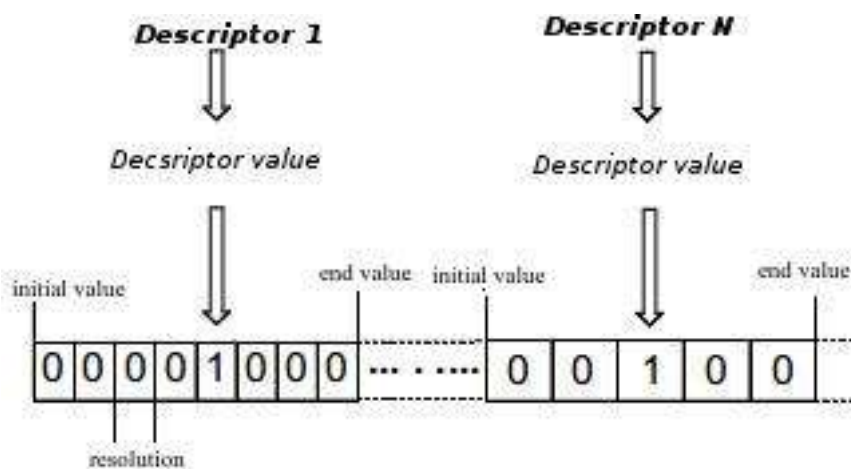


Figure 2. Formation of a descriptor fingerprint.

The descriptors forming a descriptor fingerprint can be both real values or some Boolean values indicating the presence or absence of a feature (any color, presence or absence of a property, chemical group, etc.). These discrete descriptors take one element of the fingerprint array, being either *1* or *0*. A general requirement is the quality of the descriptors to describe uniquely the studied objects and being well discriminating. Here the values of the individual FAME profiles for each biodiesel probe (case)

have been used as descriptors. Two types of descriptors have been used ones having realnumerical values and indicator ones indicating the presence (1) or absence (0) of a property.

INTRODUCTION TO FUZZY DESCRIPTOR FINGERPRINTS.

As mentioned above the fingerprint approach reflects the classical Boolean logic ($I=true$ and $0=false$). We shall call these fingerprints Boolean fingerprints. However, in the case of any type of clustering we have in many cases the I s falling in different positions within a descriptor interval for different objects of the same cluster. Hence, our aim was to reflect this fact by using fuzzy instead of Boolean logic. Fuzzy logic was developed by Zadeh [26]. In place of discrete values I and 0 , here we have real values between 0.0 and 1.0 , say, we can have a value of 0.17 (17%) and respectively 0.83 (83%) instead 0 and 1 . Accordingly, at each element of any descriptor interval (as mentioned above the descriptor interval is divided into sub-intervals) we will have real values between 0.0 and 1.0 .

For the formation of the fuzzy fingerprints, in our case, we have created clusters as described in the Experimental section. For each such a cluster the program sums all the I s of all cluster fingerprints falling in each fingerprint array element (descriptor sub-interval). Let the sum being denoted as S_{el} . In as much as, each descriptor must have value I we delete the S_{el} sums of each sub-interval element to the number n_d of all I s falling in the descriptor interval. Hence, the fuzzy fingerprint array element values are as follows (see **Figure 3**):

$$S_d = \frac{\sum I_s}{n_d} \quad 2$$

Thus, each separate cluster consisting of a number of *Boolean fingerprints* is represented by one *fuzzy fingerprint*. This approach has an additional advantage in the case of large databases. By using cluster fuzzy descriptor fingerprints the number of the database fingerprints is reduced to the number of the clusters instead to the number of the objects/Boolean fingerprints. Accordingly, the similarity comparison is carried out between one or more query object Boolean fingerprints and the fuzzy fingerprint representing the corresponding clusters. Here, we cannot use the Tanimoto index as a measure of similarity. Hence, we introduce a new *Fuzzy Factor (FF)* given with relation (3):

$$FF = 1.0 - \frac{\sum (1.0 - S_d)}{N_d} \quad 3.$$

Here S_d are the sums (2) which positions correspond to the positions of the I s in the query Boolean fingerprint, divided by the number of the descriptors N_d . Thus, for each query fingerprint we obtain N_{clstr} solutions, where N_{clstr} is the number of clusters.

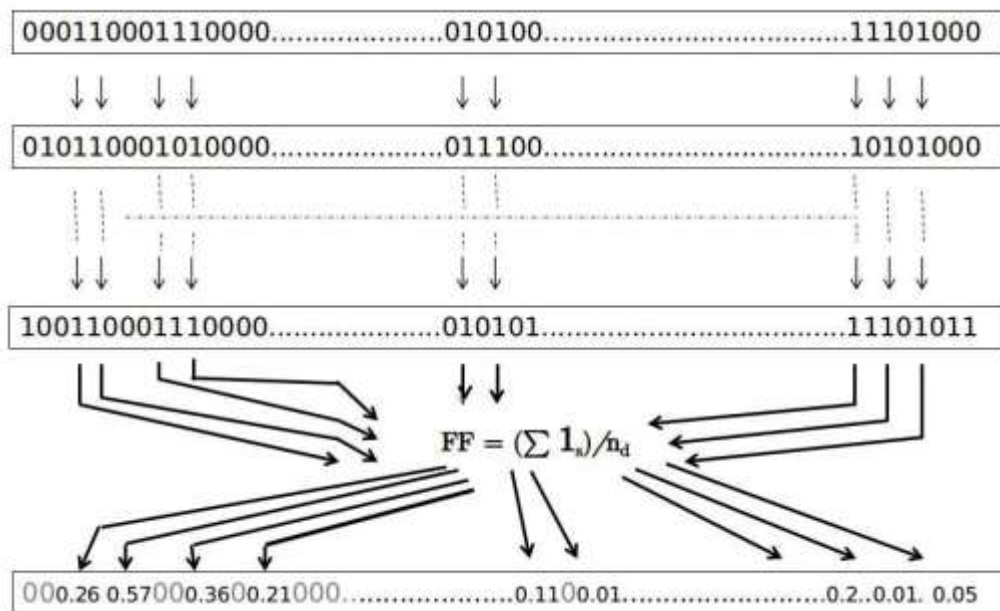


Figure 3. Creation of fuzzy fingerprints.

On can see from the relation 3 that in case of no coincidence of the query fingerprints of value I with fuzzy fingerprint elements of value different from zero all S_d sums will be zero then the FF will be also zero. In case of coincidence of the query fingerprints of value I with fuzzy fingerprint elements of value different from zero all S_d sums will be 1.0 then the FF will be also 1.0. The most probable query object falls in the cluster having the maximal FF value and it is considered the most similar to the associated object having the most similar properties to the corresponding cluster. The formation of fuzzy fingerprints is depicted in **Figure 3** and similarity search in fuzzy clusters is illustrated in **Figure 4**.

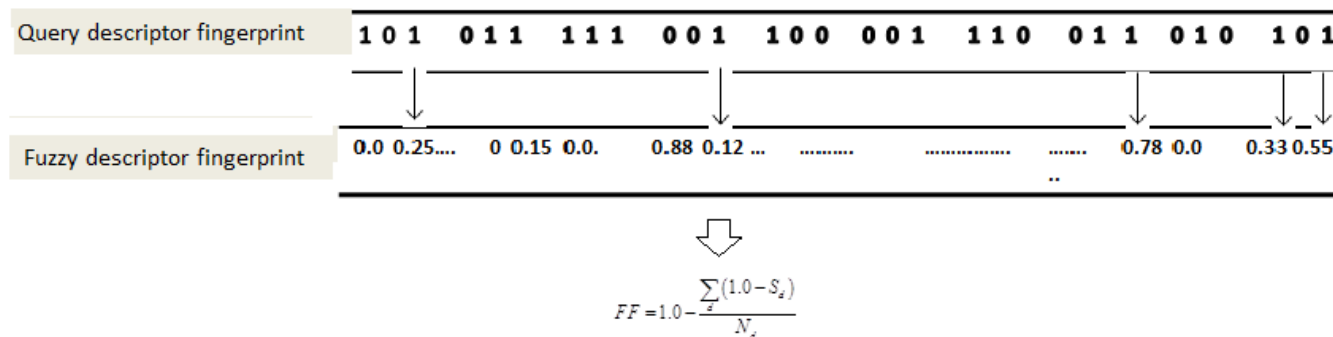


Figure 4. Depiction of the similarity search procedure between a query and a cluster fuzzy fingerprints.

It should be mentioned here one additional advantage of the employment of fuzzy descriptor fingerprints. Their use reduces substantially the representation and search in large databases, as each fuzzy fingerprint representing one and only one

cluster which often consists of a large number of Boolean fingerprints. On the other hand in as much the fuzzy fingerprint values represent the weights of the descriptors (properties) the coincidence of element 1 of the query fingerprints to the corresponding fuzzy descriptor element gives the weight (influence) of this descriptor on the query fingerprint. Each cluster is associated with some query properties, e.g. in the case of biodiesels names of the oils, cetane index, color, etc. (In our case they are given as a string), thus the maximal *FF* with a cluster assumes that the query object has also these properties.

EXPERIMENTAL

A set of 98 biodiesel samples belonging to 6 different classes of biodiesel oils – sunflower, rape, corn, soybean, palm and peanut, as well as samples of oil mixtures of these classes has been used for our fingerprint analysis. The FAME-profile for each sample was created by gas chromatographic (GC) analysis. All GC analyses were performed on a GC system Agilent Technologies 7890A equipped with FID, split/splitless injector and Agilent 7693A automated liquid sampler. The fatty acid methyl esters composition (%) of biodiesels being produced of different vegetable oils (Fatty acid methyl esters palmitate (C16:0), palmitoleate (C16:1), stearate (C18:0), oleate (C18:1), linoleate (C18:2), linolenate (C18:3), arachidate (C20:0), cis-11-eicosenoate (C20:1), behenate (C22:0), cis-13-docosanoate (C22:1), tetracosanoate (C24:0), cis-15-tetracosanoate (C24:1). SAT – total esters of saturated acids, MUNS- total esters of monounsaturated acids, PUNS - total esters of polyunsaturated acids) was determined and used for descriptors for the fingerprint creation.

The program written in Java consists of four tabs. A descriptor configuration file is created in the first tab. This file describes each descriptor with its *initVal*, *endVal* and *resVal* values. By using this configuration file and reading the data from a tab-tabulated text file (converted from an Excel file) *fingerprint* (*.fp) and *names* (*.nam) files are generated in the next tab. Any information about the separate objects, consisting of the oil names, some of their properties, is encoded in the *names* file. The values of the cetane number were additionally introduced in the *names* file. In the *Clustering* tab the *fingerprint* files are further used for clustering (by using Butina method) and creation of fuzzy descriptor fingerprints. The similarity search is carried out in the next *Similarity Search* tab.

Results and Discussion

We have explored two ways of forming clusters, the first one is the clustering to be carried out in a natural way according to the origin of biodiesel, i.e., one cluster of sunflower oil, one cluster of rape, one cluster of soy bean fingerprints, etc, oils. No clustering procedure was used to this end. Then, the number of the clusters will be 11 (see **Table 1**) and the number of the biodiesels forming this type of clusters is the same as these given in **Table 1**.

The second way of clustering is based on the Tanimoto index similarity of the source file Boolean type descriptor fingerprints by using the method of Butina [4] leading to the creation of two files: a cluster (*.clust) file containing the cluster fingerprints, and an output (*.out) text file of the generated clusters. The clustering procedure is based on an analysis through a

pair-wise comparison of the fingerprints and selection by using a threshold Tanimoto value (TV). All generated pairs having TI below the threshold value are discarded, the remaining pairs being further sorted in a decreasing order of the TI values.

Table 1. Number of clusters formed from biodiesels of different vegetable origin. (* X means mixtures of unknown or partially unknown composition).

ClusterNo	Name of the biodiesel oil origin	Number of samples
1	sunflower	33
2	rape	19
3	corn	8
4	soybean	18
5	palm	5
6	peanut	6
7	X-mix-rape*	2
8	X-mix-sunflower*	1
9	X-mix-sunflower-soy*	1
10	X-mix-soy-rape*	1
11	X-mix*	2

Two options concerning the threshold value are encoded in this approach. The first one is a user chosen manual input of the threshold value, the second an automatic generation of it. The automatic generation implies an initial pair-wise comparison of all the objects (data-set Boolean fingerprints) at a threshold value equal to zero, subsequent sorting downwards with respect to TI value, and a further step of comparing the names (including types of oils) of the objects. When a difference between the names is encountered the corresponding TI value plus 0.000001 is accepted for the threshold value. Thus, we found automatically the threshold value of 0.5763 in our case.

At the next step, fuzzy descriptor fingerprints characterizing each cluster are generated for both types of clusters according to the methods discussed above.

An internal validation using the Boolean descriptor fingerprints of the 98 biodiesels (here called query fingerprints) against the cluster fuzzy descriptor fingerprints, generated by the first type of clustering, was carried out in the third *Similarity Search* tab and the results for each Boolean fingerprint are sorted in descending order with respect of the *FF* values generated. In **Listing1** in **Appendix 1** all results for the Boolean query 1, and the maximal and second maximal *FF* results for the others are provided.

Some of the query biodiesel fuels denoted by X are mixtures of partly known composition. One can see from **Listing 1** that except from the first case 98 query **X-mix-soy-rape** being mixture of unknown X and soybean and rape composition which produces low *FF* value, the method accurately recognizes the known oil in all the other cases. The NO MATCH flag is generated by simple string comparison of the names (String type) of biodiesels of the fuzzy cluster and the name of Boolean query fingerprint, e.g., Query name “sunflower” with fuzzy name “sunflower” which implies that the result is correct. Vice versa, comparing the Query name “sunflower” with cluster name “rape” will produce NO MATCH flag. Hence, some of the X-mix-... comparisons, denoted by NO MATCH flag might be correct.

An external validation was performed by separating the test series of 98 biodiesel oils into two groups. The first group (data-set group) was formed with 83 oils belonging to 6 classes of biodiesels (sunflower, rape, corn, soy bean, palm, and peanut) with a known origin. The fuzzy descriptor fingerprints for these clusters were generated. The second query group consists of 15 analytes of both biodiesels of known, partially known and unknown constitution (the two later denoted by X).

The maximal *FF* results of similarity comparison between the 15 Boolean query fingerprints each one with the 6 cluster fuzzy fingerprints are provided in **Figure 5**. One can see that only in one case, the maximal *FF* fingerprint 98 of X-mix-soy-rape produces corn biodiesel to cluster 3, although the full mixture is not known. The mark NO MATCH is assigned if the names of the biodiesels does not coincide (string comparison) with the oil names of the clusters, which is the case of all X-*** mixtures, although the results look correct. Hence this procedure could be applied as an analytic method for elucidation of any mixture composition. The other cases marked as no matching are unknown mixtures other cases marked as no matching are unknown mixtures denoted as X. As discussed above, the program compares the names of the biodiesels as strings and mark them with NO MATCH flag if they are not equal. One can see that in all but mixture 98 the method recognizes at least one of the mixture component. However in case 98 the X component is not known.

```
MAX FF fingerprint Query 23 ---sunflower with CLUSTER 1 sunflower FF=0.52972
MAX FF fingerprint Query 33 ---sunflower with CLUSTER 1 sunflower FF=0.61789
MAX FF fingerprint Query 49 ---rape with CLUSTER 2 rape FF=0.47511
MAX FF fingerprint Query 50 ---rape with CLUSTER 2 rape FF=0.25882
MAX FF fingerprint Query 53 ---corn with CLUSTER 3 corn FF=0.44762
MAX FF fingerprint Query 78 ---soy bean with CLUSTER 4 soy bean FF=0.67407
MAX FF fingerprint Query 80 ---palm with CLUSTER 5 palm FF=0.77667
MAX FF fingerprint Query 88 ---peanut with CLUSTER 6 peanut FF=0.34667
MAX FF fingerprint Query 92 ---X-mix-rape with CLUSTER 2 rape FF=0.24284---> NO MATCH
MAX FF fingerprint Query 93 ---X-mix-sunfl with CLUSTER 1 sunflower FF=0.30344---> NO MATCH
MAX FF fingerprint Query 94 ---X-mix-rape with CLUSTER 5 palm FF=0.25333---> NO MATCH
MAX FF fingerprint Query 95 ---X-mix with CLUSTER 1 sunflower FF=0.19355---> NO MATCH
MAX FF fingerprint Query 96 ---X-mix with CLUSTER 3 corn FF=0.27619---> NO MATCH
MAX FF fingerprint Query 97 ---X-mix-sunfl-soy with CLUSTER 4 soy bean FF=0.27037---> NO MATCH
MAX FF fingerprint Query 98 ---X-mix-soy-rape with CLUSTER 3 corn FF=0.18095---> NO MATCH
```

Figure 5. Similarity comparison of 15 Boolean query fingerprints with the 6 cluster fuzzy fingerprints generated from 83 Boolean fingerprints.

Next, by using the Butina method for clustering using the Boolean descriptor fingerprints we generated a new set of 18 clusters. They, together with the number of fingerprints in each cluster are depicted in **Table 2**.

Further, from the clusters generated in such a way the fuzzy descriptor fingerprints characterizing each cluster are generated according the method discussed above. Then, an internal validation using the query Boolean descriptor fingerprints of the 98 biodiesels against the cluster fuzzy descriptor fingerprints is carried out and the results for each Boolean fingerprint are sorted in descending order with respect to the FF values generated. Results having maximal FF for each query Boolean descriptor fingerprint, compared to all cluster fuzzy fingerprints are provided in **Listing 2** in **Appendix 1**. The query biodiesel fuels are mixtures of partly know composition. One can see from Figure 5 that except from the first case query 63 soy bean with Cluster 6 the method accurately recognizes the known oil in all the other cases.

CONCLUSIONS

The application of the method of descriptor fingerprints to the problems of clustering of biodiesels indicates that they can be a useful tool to this end. A novel type of *fuzzy descriptor fingerprints* related to each cluster and a new Fuzzy Factor index have been introduced and tested with our set of 98 biodiesels. They were generated both from the 12 clusters shown in Table 1, formed from biodiesels composed of uniform oil types - sunflower, rape, corn, soybean, palm peanut, etc., each cluster being represented by one fuzzy fingerprint.

Additionally, 18 fuzzy clusters fingerprints generated from the 18 clusters (**Table 2**) obtained by employing the method of Butina, have been treated by using 98 query Boolean descriptor fingerprints. The search and subsequent clustering results, based on the highest FF value for each query fingerprint indicate that the fuzzy fingerprint approach leads to very good discrimination between different biodiesel fuels.

Table 2. Clusters generated by Butina method. and number of biodiesels in each cluster.

Cluster No	biodiesel kind	Nu of biodiesels
1	sunflower	31
2	sunflower	2
3	rape	16
4	rape	2
5	rape	1
6	corn	4
7	corn	2
8	corn	2
9	soy bean	19
10	palm	6
11	peanut	5
12	peanut	1
13-18	X-***	each one x1

*Author to whom correspondence should be addressed.

Acknowledgment: This work has been supported by grant DFNI IO1/7 of the National Fund for Scientific Investigation and under project № RD-03-255/09.03.2015, of “Scientific Research” section of Konstantin Preslavsky University of Shumen.

REFERENCES

1. Bangov, I. P.; Naneva, L. C.; Moskovkina, M. N.; Dimitrov, I.; Doytchinova, I.A, Application of Descriptor Fingerprints in QSAR Studies. *Annual of Konstantin Preslavski University of Shumen. Natural Sciences. Chemistry***2013**, 22 B1, 28-36.
2. Dimitrov, I.; Naneva, L.C.; Doytchinova, I.A.; Bangov, I.P. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics***2014**, 30 (6), 846-851.
3. Stoynov, B. P.; Kostadinov, P. R.; Kolev, M. V.; Mustafa, Z. A.; Moskovkina, M. N.; Milina, R. S.; Bangov, I. P. Use of the descriptor fingerprints to clustering of chemical datasets. *Bulg. Chem. Comm.***2014**, 46 (A), 209-214.
4. Butina, D. Unsupervised data base clustering based on Daylight’s fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 747-750.
5. Hoekman, S. K.; Broch A.; Robbins, C.; Cenicerros, E.; Natarajan, M. Review of biodiesel composition, properties, and specifications. *Renewable and Sustainable Energy Reviews***2012**, 16, 143-149.
6. Sharma, Y.C.; Singh, B.; Upadhyay, S.N. Advancements in development and characterization of biodiesel: A review. *Fuel***2008**, 87, 2355–2373.
7. Atabania, A.E.; Silitonga, A.S.; Badruddina, I. A.; Mahlia, T.M.I.; Masjuki, H.H.; Mekhilef, S.A comprehensive review on biodiesel as an alternative energy resource and its characteristics. *Renewable and Sustainable Energy Reviews***2012**, 16, 2070– 2093.

8. Ramos, M. J., Fernandez, C. M., Casas, A., Rodriguez, L., Perez, A. Influence of fatty acid composition of raw materials on biodiesel properties, *Bioresource Technology* 100 (2009) 261–268.
9. Rodriguez, L.;Perez, A. Influence of fatty acid composition of raw materials on biodiesel properties.*Bioresource Technology***2009**,*100*, 261–268.
10. Prados, C. P.; Rezende, D. R.; Batista, L. R.; Alves, N. R.; Filho. Simultaneous gas chromatographic analysis of total esters, mono-, di-, and triacylglycerides and free and total glycerol in methyl or ethyl biodiesel.*Fuel***2011**, *96*, 476 – 481.
11. Dodds, E. D.; McCoy, M. R.; Rea, L. D.; Kennish, M. J. Gas chromatographic quantification of fatty acid methyl esters: flame ionization detection vs election impact mass spectrometry.*Lipids***2005**,*40*, 419 – 428.
12. Mondello, L.; Casilli, A.; Tranchida, P. Q.; Costa.R.; Chiofalo, B.; Dugo, P. Evaluation of fast gas chromatography and gas ghromatography – mass spectrometry in the analysis of lipids.*J. Chromatogr. A***2004**, *1035*, 237 – 247
13. Bravy, E.; Perrety, G.; Montanary, L. Fatty acids by high-performance liquid chromatography and evaporative light-scattering detector.*J. Chromatogr. A***2006**, *1134*, 210 – 214.
14. Milina, R.; Mustafa, Z. Gas chromatographic investigations of compositional profiles of biodiesel from different origin, *Petroleum & Coal***2013**, *55*, 12-19.
15. Moser, B. R.; Vaughn, S. F. Efficacy of fatty acid profile as a tool for screening feedstocks for biodiesel production.*Biomass and bioenergy***2012**, *37*, 31 – 41.
16. Leung, D. Y. C.; Koo, B. C. P.; Guo, Y. Degradation of biodiesel under different storage conditions.*BioresourceTechnology***2006**, *97*, 250 – 256.
17. Knothe, G. Some aspects of biodiesel oxidative stability.*FuelProcessing Technology***2007**,*88*, 669.

18. Khoury, R. R.; Ebrahimi, D.; Hejazy, L.; Bucknall, M. P.; Pickford, R.; Hibbert, D. B. Degradation of fatty acid methyl esters in biodiesels exposed to sunlight and seawater, *Fuel* **2011**, *90*, 2677 – 2683.
19. Barnard, J. M. Representation of Molecular Structures. Overview. In *Handbook of Chemoinformatics*; Gastejger J., Ed.; Willey: VCH; Weinheim, 2003; Vol. 1; p 27.
20. Tomczak, J. DataTypes. In *Handbook of Chemoinformatics*; Gastejger, J., Ed.; Willey-VCH: Weinheim, 2003; Vol 2; p 392.
21. Engel, T. Representation of Chemical Compounds In *Chemoinformatics. A textbook*; Engel, T.; Gastejger, J. Eds.; Willey-VCH: Weinheim, 2003; p 15-168.
22. Willett, P.; Similarity searching in Chemical Databases, In *Handbook of Chemoinformatics*, Gastejger, J. Ed.; Willey-VCH: Weinheim, 2003; Vol. 2; p 904.
23. <http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>
24. Kochev, N.; Monev, V.; Bangov, I.; Searching Chemical Structures. In *Chemoinformatics. A textbook*, Engel T.; Gastejger, J. Eds.; Willey-VCH: Weinheim, 2003; p 291.
25. <https://www.google.bg/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=tanimoto%20coefficient>
26. Zadeh, L. A. Fuzzy sets. *Information and Control* **1965**, *8* [3], 338-353.

Appendix 1.

PRINT Clusters for Queries

```

MAX FF fingerprint Query 1 ---sunflower with CLUSTER 1 sunflower FF=0.48737
Query 1 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.23333--> NO MATCH
Query 1 ---sunflower with CLUSTER 3 corn FF=0.22500--> NO MATCH
Query 1 ---sunflower with CLUSTER 5 palm FF=0.22222--> NO MATCH
Query 1 ---sunflower with CLUSTER 9 X-mix FF=0.20000--> NO MATCH
Query 1 ---sunflower with CLUSTER 10 X-mix-sunfl-soy FF=0.20000--> NO MATCH
Query 1 ---sunflower with CLUSTER 11 X-mix-soy-rape FF=0.20000--> NO MATCH
Query 1 ---sunflower with CLUSTER 4 soybean FF=0.14035--> NO MATCH
Query 1 ---sunflower with CLUSTER 7 X-mix-rape FF=0.13333--> NO MATCH
Query 1 ---sunflower with CLUSTER 6 peanut FF=0.10000--> NO MATCH
Query 1 ---sunflower with CLUSTER 2 rape FF=0.09091--> NO MATCH
MAX FF fingerprint Query 2 ---sunflower with CLUSTER 1 sunflower FF=0.49545
Query 2 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.43333--> NO MATCH
MAX FF fingerprint Query 3 ---sunflower with CLUSTER 1 sunflower FF=0.54280
Query 3 ---sunflower with CLUSTER 3 corn FF=0.38611--> NO MATCH
MAX FF fingerprint Query 4 ---sunflower with CLUSTER 1 sunflower FF=0.57917
Query 4 ---sunflower with CLUSTER 3 corn FF=0.47778--> NO MATCH
MAX FF fingerprint Query 5 ---sunflower with CLUSTER 1 sunflower FF=0.62563
Query 5 ---sunflower with CLUSTER 3 corn FF=0.44444--> NO MATCH
MAX FF fingerprint Query 6 ---sunflower with CLUSTER 1 sunflower FF=0.61957
Query 6 ---sunflower with CLUSTER 3 corn FF=0.44444--> NO MATCH
MAX FF fingerprint Query 7 ---sunflower with CLUSTER 1 sunflower FF=0.61957
Query 7 ---sunflower with CLUSTER 3 corn FF=0.44444--> NO MATCH
MAX FF fingerprint Query 8 ---sunflower with CLUSTER 1 sunflower FF=0.55694

```

```

Query 8 ---sunflower with CLUSTER 3 corn FF=0.41944--> NO MATCH
MAX FF fingerprint Query 9 ---sunflower with CLUSTER 1 sunflower FF=0.46199
Query 9 ---sunflower with CLUSTER 3 corn FF=0.41111--> NO MATCH
MAX FF fingerprint Query 10 ---sunflower with CLUSTER 1 sunflower FF=0.45505
Query 10 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.43333--> NO MATCH
MAX FF fingerprint Query 11 ---sunflower with CLUSTER 1 sunflower FF=0.51566
Query 11 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.40000--> NO MATCH
MAX FF fingerprint Query 12 ---sunflower with CLUSTER 1 sunflower FF=0.52866
Query 12 ---sunflower with CLUSTER 10 X-mix-soy-rape FF=0.46667--> NO MATCH
MAX FF fingerprint Query 13 ---sunflower with CLUSTER 1 sunflower FF=0.63081
Query 13 ---sunflower with CLUSTER 3 corn FF=0.27500--> NO MATCH
MAX FF fingerprint Query 14 ---sunflower with CLUSTER 1 sunflower FF=0.63169
Query 14 ---sunflower with CLUSTER 3 corn FF=0.30278--> NO MATCH
MAX FF fingerprint Query 15 ---sunflower with CLUSTER 1 sunflower FF=0.63169
Query 15 ---sunflower with CLUSTER 3 corn FF=0.30278--> NO MATCH
MAX FF fingerprint Query 16 ---sunflower with CLUSTER 1 sunflower FF=0.62765
Query 16 ---sunflower with CLUSTER 3 corn FF=0.33611--> NO MATCH
MAX FF fingerprint Query 17 ---sunflower with CLUSTER 1 sunflower FF=0.62563
Query 17 ---sunflower with CLUSTER 3 corn FF=0.30278--> NO MATCH
MAX FF fingerprint Query 18 ---sunflower with CLUSTER 1 sunflower FF=0.55492
Query 18 ---sunflower with CLUSTER 4 soy bean FF=0.28070--> NO MATCH
MAX FF fingerprint Query 19 ---sunflower with CLUSTER 1 sunflower FF=0.58030
Query 19 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.33333--> NO MATCH
MAX FF fingerprint Query 20 ---sunflower with CLUSTER 1 sunflower FF=0.60947

```


Query 20 ---sunflower with CLUSTER 4 soy bean FF=0.29474--> NO MATCH
MAX FF fingerprint Query 21 ---sunflower with CLUSTER 1 sunflower FF=0.62159
 Query 21 ---sunflower with CLUSTER 3 corn FF=0.30278--> NO MATCH
MAX FF fingerprint Query 22 ---sunflower with CLUSTER 1 sunflower FF=0.60947
 Query 22 ---sunflower with CLUSTER 4 soy bean FF=0.29123--> NO MATCH
MAX FF fingerprint Query 23 ---sunflower with CLUSTER 1 sunflower FF=0.51856
 Query 23 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.40000--> NO MATCH
MAX FF fingerprint Query 24 ---sunflower with CLUSTER 1 sunflower FF=0.55492
 Query 24 ---sunflower with CLUSTER 10 X-mix-soy-rape FF=0.26667--> NO MATCH
MAX FF fingerprint Query 25 ---sunflower with CLUSTER 1 sunflower FF=0.62563
 Query 25 ---sunflower with CLUSTER 3 corn FF=0.31944--> NO MATCH
MAX FF fingerprint Query 26 ---sunflower with CLUSTER 1 sunflower FF=0.56616
 Query 26 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.26667--> NO MATCH
MAX FF fingerprint Query 27 ---sunflower with CLUSTER 1 sunflower FF=0.46269
 Query 27 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.26667--> NO MATCH
MAX FF fingerprint Query 28 ---sunflower with CLUSTER 1 sunflower FF=0.46269
 Query 28 ---sunflower with CLUSTER 9 X-mix-sunfl-soy FF=0.33333--> NO MATCH
MAX FF fingerprint Query 29 ---sunflower with CLUSTER 1 sunflower FF=0.60947
 Query 29 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.33333--> NO MATCH
MAX FF fingerprint Query 30 ---sunflower with CLUSTER 1 sunflower FF=0.49703
 Query 30 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.33333--> NO MATCH
MAX FF fingerprint Query 31 ---sunflower with CLUSTER 1 sunflower FF=0.41263
 Query 31 ---sunflower with CLUSTER 8 X-mix-sunfl FF=0.30000--> NO MATCH
MAX FF fingerprint Query 32 ---sunflower with CLUSTER 1 sunflower FF=0.57311
 Query 32 ---sunflower with CLUSTER 3 corn FF=0.44444--> NO MATCH
MAX FF fingerprint Query 33 ---sunflower with CLUSTER 1 sunflower FF=0.57715
 Query 33 ---sunflower with CLUSTER 3 corn FF=0.42778--> NO MATCH
MAX FF fingerprint Query 34 ---rape with CLUSTER 2 rape FF=0.41829
 Query 34 ---rape with CLUSTER 7 X-mix-rape FF=0.26667--> NO MATCH
MAX FF fingerprint Query 35 ---rape with CLUSTER 2 rape FF=0.46109
 Query 35 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.33333--> NO MATCH
MAX FF fingerprint Query 36 ---rape with CLUSTER 2 rape FF=0.41526
 Query 36 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 37 ---rape with CLUSTER 2 rape FF=0.53965
 Query 37 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.33333--> NO MATCH
MAX FF fingerprint Query 38 ---rape with CLUSTER 2 rape FF=0.53056
 Query 38 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 39 ---rape with CLUSTER 2 rape FF=0.53647
 Query 39 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 40 ---rape with CLUSTER 2 rape FF=0.47927
 Query 40 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.33333--> NO MATCH
MAX FF fingerprint Query 41 ---rape with CLUSTER 2 rape FF=0.50427
 Query 41 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 42 ---rape with CLUSTER 2 rape FF=0.49140
 Query 42 ---rape with CLUSTER 7 X-mix-rape FF=0.26667--> NO MATCH
MAX FF fingerprint Query 43 ---rape with CLUSTER 2 rape FF=0.48837
 Query 43 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 44 ---rape with CLUSTER 2 rape FF=0.55783
 Query 44 ---rape with CLUSTER 7 X-mix-rape FF=0.30000--> NO MATCH
MAX FF fingerprint Query 45 ---rape with CLUSTER 2 rape FF=0.51526
 Query 45 ---rape with CLUSTER 7 X-mix-rape FF=0.30000--> NO MATCH
MAX FF fingerprint Query 46 ---rape with CLUSTER 2 rape FF=0.50011
 Query 46 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 47 ---rape with CLUSTER 2 rape FF=0.51275
 Query 47 ---rape with CLUSTER 7 X-mix-rape FF=0.30000--> NO MATCH
MAX FF fingerprint Query 48 ---rape with CLUSTER 2 rape FF=0.44670
 Query 48 ---rape with CLUSTER 7 X-mix-rape FF=0.30000--> NO MATCH
MAX FF fingerprint Query 49 ---rape with CLUSTER 2 rape FF=0.44078
 Query 49 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 50 ---rape with CLUSTER 2 rape FF=0.30215
 Query 50 ---rape with CLUSTER 10 X-mix-soy-rape FF=0.26667--> NO MATCH
MAX FF fingerprint Query 51 ---rape with CLUSTER 2 rape FF=0.49419
 Query 51 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.33333--> NO MATCH
MAX FF fingerprint Query 52 ---rape with CLUSTER 2 rape FF=0.46677
 Query 52 ---rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH
MAX FF fingerprint Query 53 ---corn with CLUSTER 3 corn FF=0.45556
 Query 53 ---corn with CLUSTER 4 soy bean FF=0.37544--> NO MATCH
MAX FF fingerprint Query 54 ---corn with CLUSTER 3 corn FF=0.53056
 Query 54 ---corn with CLUSTER 4 soy bean FF=0.41404--> NO MATCH
MAX FF fingerprint Query 55 ---corn with CLUSTER 3 corn FF=0.40278
 Query 55 ---corn with CLUSTER 4 soy bean FF=0.34737--> NO MATCH
MAX FF fingerprint Query 56 ---corn with CLUSTER 3 corn FF=0.47778
 Query 56 ---corn with CLUSTER 4 soy bean FF=0.40702--> NO MATCH
MAX FF fingerprint Query 57 ---corn with CLUSTER 3 corn FF=0.58611
 Query 57 ---corn with CLUSTER 4 soy bean FF=0.46667--> NO MATCH
MAX FF fingerprint Query 58 ---corn with CLUSTER 3 corn FF=0.46944
 Query 58 ---corn with CLUSTER 10 X-mix-soy-rape FF=0.40000--> NO MATCH
MAX FF fingerprint Query 59 ---corn with CLUSTER 3 corn FF=0.51944
 Query 59 ---corn with CLUSTER 4 soy bean FF=0.49474--> NO MATCH
MAX FF fingerprint Query 60 ---corn with CLUSTER 3 corn FF=0.51944
 Query 60 ---corn with CLUSTER 4 soy bean FF=0.40702--> NO MATCH
MAX FF fingerprint Query 61 ---soy bean with CLUSTER 4 soy bean FF=0.70526
 Query 61 ---soy bean with CLUSTER 3 corn FF=0.46944--> NO MATCH
MAX FF fingerprint Query 62 ---soy bean with CLUSTER 4 soy bean FF=0.66316
 Query 62 ---soy bean with CLUSTER 3 corn FF=0.41389--> NO MATCH
MAX FF fingerprint Query 63 ---soy bean with CLUSTER 4 soy bean FF=0.53333
 Query 63 ---soy bean with CLUSTER 3 corn FF=0.41389--> NO MATCH
MAX FF fingerprint Query 64 ---soy bean with CLUSTER 4 soy bean FF=0.65263
 Query 64 ---soy bean with CLUSTER 3 corn FF=0.41111--> NO MATCH
MAX FF fingerprint Query 65 ---soy bean with CLUSTER 4 soy bean FF=0.64211
 Query 65 ---soy bean with CLUSTER 3 corn FF=0.37778--> NO MATCH
MAX FF fingerprint Query 66 ---soy bean with CLUSTER 4 soy bean FF=0.63509
 Query 66 ---soy bean with CLUSTER 3 corn FF=0.33889--> NO MATCH
MAX FF fingerprint Query 67 ---soy bean with CLUSTER 4 soy bean FF=0.68772
 Query 67 ---soy bean with CLUSTER 3 corn FF=0.39444--> NO MATCH
MAX FF fingerprint Query 68 ---soy bean with CLUSTER 4 soy bean FF=0.66667
 Query 68 ---soy bean with CLUSTER 3 corn FF=0.39444--> NO MATCH
MAX FF fingerprint Query 69 ---soy bean with CLUSTER 4 soy bean FF=0.62105
 Query 69 ---soy bean with CLUSTER 3 corn FF=0.45278--> NO MATCH
MAX FF fingerprint Query 70 ---soy bean with CLUSTER 4 soy bean FF=0.51579
 Query 70 ---soy bean with CLUSTER 3 corn FF=0.35556--> NO MATCH
MAX FF fingerprint Query 71 ---soy bean with CLUSTER 4 soy bean FF=0.61404
 Query 71 ---soy bean with CLUSTER 3 corn FF=0.38056--> NO MATCH
MAX FF fingerprint Query 72 ---soy bean with CLUSTER 4 soy bean FF=0.47368
 Query 72 ---soy bean with CLUSTER 1 sunflower FF=0.42563--> NO MATCH
MAX FF fingerprint Query 73 ---soy bean with CLUSTER 4 soy bean FF=0.59298
 Query 73 ---soy bean with CLUSTER 3 corn FF=0.40278--> NO MATCH
MAX FF fingerprint Query 74 ---soy bean with CLUSTER 4 soy bean FF=0.65965
 Query 74 ---soy bean with CLUSTER 3 corn FF=0.43056--> NO MATCH
MAX FF fingerprint Query 75 ---soy bean with CLUSTER 4 soy bean FF=0.67018
 Query 75 ---soy bean with CLUSTER 3 corn FF=0.45278--> NO MATCH
MAX FF fingerprint Query 76 ---soy bean with CLUSTER 4 soy bean FF=0.71930
 Query 76 ---soy bean with CLUSTER 3 corn FF=0.46944--> NO MATCH
MAX FF fingerprint Query 77 ---soy bean with CLUSTER 4 soy bean FF=0.69474
 Query 77 ---soy bean with CLUSTER 3 corn FF=0.43611--> NO MATCH
MAX FF fingerprint Query 78 ---soy bean with CLUSTER 4 soy bean FF=0.68070
 Query 78 ---soy bean with CLUSTER 3 corn FF=0.48611--> NO MATCH
MAX FF fingerprint Query 79 ---soy bean with CLUSTER 4 soy bean FF=0.63158
 Query 79 ---soy bean with CLUSTER 3 corn FF=0.41944--> NO MATCH
MAX FF fingerprint Query 80 ---palm with CLUSTER 5 palm FF=0.76667
 Query 80 ---palm with CLUSTER 4 soy bean FF=0.37895--> NO MATCH
MAX FF fingerprint Query 81 ---palm with CLUSTER 5 palm FF=0.73333
 Query 81 ---palm with CLUSTER 4 soy bean FF=0.31228--> NO MATCH
MAX FF fingerprint Query 82 ---palm with CLUSTER 5 palm FF=0.71111
 Query 82 ---palm with CLUSTER 1 sunflower FF=0.26667--> NO MATCH
MAX FF fingerprint Query 83 ---palm with CLUSTER 5 palm FF=0.58889
 Query 83 ---palm with CLUSTER 4 soy bean FF=0.29825--> NO MATCH
MAX FF fingerprint Query 84 ---palm with CLUSTER 5 palm FF=0.73333
 Query 84 ---palm with CLUSTER 3 corn FF=0.31667--> NO MATCH
MAX FF fingerprint Query 85 ---palm with CLUSTER 5 palm FF=0.73333
 Query 85 ---palm with CLUSTER 3 corn FF=0.31667--> NO MATCH
MAX FF fingerprint Query 86 ---peanut with CLUSTER 6 peanut FF=0.48889
 Query 86 ---peanut with CLUSTER 3 corn FF=0.20000--> NO MATCH
MAX FF fingerprint Query 87 ---peanut with CLUSTER 6 peanut FF=0.62222
 Query 87 ---peanut with CLUSTER 3 corn FF=0.25833--> NO MATCH
MAX FF fingerprint Query 88 ---peanut with CLUSTER 6 peanut FF=0.43889
 Query 88 ---peanut with CLUSTER 8 X-mix-sunfl FF=0.16667--> NO MATCH
MAX FF fingerprint Query 89 ---peanut with CLUSTER 6 peanut FF=0.68333
 Query 89 ---peanut with CLUSTER 3 corn FF=0.25833--> NO MATCH
MAX FF fingerprint Query 90 ---peanut with CLUSTER 6 peanut FF=0.62778
 Query 90 ---peanut with CLUSTER 3 corn FF=0.21667--> NO MATCH
MAX FF fingerprint Query 91 ---peanut with CLUSTER 6 peanut FF=0.69444
 Query 91 ---peanut with CLUSTER 3 corn FF=0.22500--> NO MATCH
MAX FF fingerprint Query 92 ---X-mix-rape with CLUSTER 7 X-mix-rape FF=0.63333
 Query 92 ---X-mix-rape with CLUSTER 2 rape FF=0.26795--> NO MATCH
MAX FF fingerprint Query 93 ---X-mix-sunfl with CLUSTER 8 X-mix-sunfl FF=0.63333
 Query 93 ---X-mix-sunfl with CLUSTER 1 sunflower FF=0.36774--> NO MATCH
MAX FF fingerprint Query 94 ---X-mix-rape with CLUSTER 7 X-mix-rape FF=0.66667
 Query 94 ---X-mix-rape with CLUSTER 8 X-mix-sunfl FF=0.30000--> NO MATCH
MAX FF fingerprint Query 95 ---X-mix with CLUSTER 9 X-mix-sunfl-soy FF=0.40000--> NO MATCH
 Query 95 ---X-mix with CLUSTER 8 X-mix-sunfl FF=0.36667--> NO MATCH
MAX FF fingerprint Query 96 ---X-mix with CLUSTER 10 X-mix-soy-rape FF=0.46667--> NO MATCH
 Query 96 ---X-mix with CLUSTER 3 corn FF=0.35278--> NO MATCH
MAX FF fingerprint Query 97 ---X-mix-sunfl-soy with CLUSTER 9 X-mix-sunfl-soy FF=1.00000
 Query 97 ---X-mix-sunfl-soy with CLUSTER 8 X-mix-sunfl FF=0.63333--> NO MATCH
MAX FF fingerprint Query 98 ---X-mix-soy-rape with CLUSTER 10 X-mix-soy-rape FF=1.00000
 Query 98 ---X-mix-soy-rape with CLUSTER 9 X-mix-sunfl-soy FF=0.26667--> NO MATCH

Listing 1. Results from internal validation of the 98 fingerprints versus all 11 fuzzy cluster descriptor fingerprints, generated by the 11 clusters. The full results are presented only for the 1st query fingerprint, the maximal and second maximal FF results provided for the other query fingerprints.

PRINT Clusters for Queries

MAX FF fingerprint Query 1 ---sunflower with CLUSTER 1 sunflower FF=0.49735
 Query 1 ---sunflower with CLUSTER 2 sunflower FF=0.33333
 Query 1 ---sunflower with CLUSTER 14 X-mix-sunfl FF=0.26667--> NO MATCH
 Query 1 ---sunflower with CLUSTER 16 X-mix FF=0.26667--> NO MATCH
 Query 1 ---sunflower with CLUSTER 6 corn FF=0.23333--> NO MATCH
 Query 1 ---sunflower with CLUSTER 7 corn FF=0.23333--> NO MATCH
 Query 1 ---sunflower with CLUSTER 10 palm FF=0.22222--> NO MATCH
 Query 1 ---sunflower with CLUSTER 8 corn FF=0.20000--> NO MATCH

Query 1	--sunflower with CLUSTER 13	X-mix-rape FF=0.20000-->	NO MATCH		
Query 1	--sunflower with CLUSTER 17	X-mix FF=0.20000-->	NO MATCH		
Query 1	--sunflower with CLUSTER 18	X-mix-sunfl-soy FF=0.20000-->	NO MATCH		
Query 1	--sunflower with CLUSTER 9	soy bean FF=0.14035-->	NO MATCH		
Query 1	--sunflower with CLUSTER 12	peanut FF=0.13333-->	NO MATCH		
Query 1	--sunflower with CLUSTER 11	peanut FF=0.09333-->	NO MATCH		
Query 1	--sunflower with CLUSTER 3	rape FF=0.08333-->	NO MATCH		
Query 1	--sunflower with CLUSTER 4	rape FF=0.06667-->	NO MATCH		
Query 1	--sunflower with CLUSTER 5	rape FF=0.06667-->	NO MATCH		
Query 1	--sunflower with CLUSTER 15	X-mix-rape FF=0.06667-->	NO MATCH		
=====					
MAX FF fingerprint Query 2	--sunflower with CLUSTER 1	sunflower FF=0.49735			
MAX FF fingerprint Query 3	--sunflower with CLUSTER 2	sunflower FF=0.63333			
MAX FF fingerprint Query 4	--sunflower with CLUSTER 2	sunflower FF=0.63333			
MAX FF fingerprint Query 5	--sunflower with CLUSTER 1	sunflower FF=0.62724			
MAX FF fingerprint Query 6	--sunflower with CLUSTER 1	sunflower FF=0.62079			
MAX FF fingerprint Query 7	--sunflower with CLUSTER 2	sunflower FF=0.63333			
MAX FF fingerprint Query 8	--sunflower with CLUSTER 1	sunflower FF=0.55842			
MAX FF fingerprint Query 9	--sunflower with CLUSTER 2	sunflower FF=0.73333			
MAX FF fingerprint Query 10	--sunflower with CLUSTER 14	X-mix-sunfl FF=0.53333-->	NO MATCH		
=====					
MAX FF fingerprint Query 11	--sunflower with CLUSTER 2	sunflower FF=0.56667			
MAX FF fingerprint Query 12	--sunflower with CLUSTER 2	sunflower FF=0.53333			
MAX FF fingerprint Query 13	--sunflower with CLUSTER 1	sunflower FF=0.64573			
MAX FF fingerprint Query 14	--sunflower with CLUSTER 1	sunflower FF=0.64229			
MAX FF fingerprint Query 15	--sunflower with CLUSTER 1	sunflower FF=0.64229			
MAX FF fingerprint Query 16	--sunflower with CLUSTER 1	sunflower FF=0.63584			
MAX FF fingerprint Query 17	--sunflower with CLUSTER 1	sunflower FF=0.63584			
MAX FF fingerprint Query 18	--sunflower with CLUSTER 1	sunflower FF=0.56057			
MAX FF fingerprint Query 19	--sunflower with CLUSTER 1	sunflower FF=0.59627			
MAX FF fingerprint Query 20	--sunflower with CLUSTER 1	sunflower FF=0.61864			
MAX FF fingerprint Query 21	--sunflower with CLUSTER 1	sunflower FF=0.63154			
MAX FF fingerprint Query 22	--sunflower with CLUSTER 1	sunflower FF=0.61864			
MAX FF fingerprint Query 23	--sunflower with CLUSTER 2	sunflower FF=0.73333			
MAX FF fingerprint Query 24	--sunflower with CLUSTER 1	sunflower FF=0.56057			
MAX FF fingerprint Query 25	--sunflower with CLUSTER 1	sunflower FF=0.63584			
MAX FF fingerprint Query 26	--sunflower with CLUSTER 1	sunflower FF=0.57907			
MAX FF fingerprint Query 27	--sunflower with CLUSTER 1	sunflower FF=0.47749			
MAX FF fingerprint Query 28	--sunflower with CLUSTER 1	sunflower FF=0.47749			
MAX FF fingerprint Query 29	--sunflower with CLUSTER 1	sunflower FF=0.62294			
MAX FF fingerprint Query 30	--sunflower with CLUSTER 1	sunflower FF=0.50760			
MAX FF fingerprint Query 31	--sunflower with CLUSTER 2	sunflower FF=0.50000			
MAX FF fingerprint Query 32	--sunflower with CLUSTER 2	sunflower FF=0.66667			
MAX FF fingerprint Query 33	--sunflower with CLUSTER 2	sunflower FF=0.60000			
MAX FF fingerprint Query 34	--rape with CLUSTER 3	rape FF=0.48889			
MAX FF fingerprint Query 35	--rape with CLUSTER 4	rape FF=0.83333			
MAX FF fingerprint Query 36	--rape with CLUSTER 3	rape FF=0.47639			
MAX FF fingerprint Query 37	--rape with CLUSTER 3	rape FF=0.60972			
MAX FF fingerprint Query 38	--rape with CLUSTER 3	rape FF=0.60972			
MAX FF fingerprint Query 39	--rape with CLUSTER 3	rape FF=0.62222			
MAX FF fingerprint Query 40	--rape with CLUSTER 4	rape FF=0.83333			
MAX FF fingerprint Query 41	--rape with CLUSTER 3	rape FF=0.57917			
MAX FF fingerprint Query 42	--rape with CLUSTER 4	rape FF=0.70000			
MAX FF fingerprint Query 43	--rape with CLUSTER 4	rape FF=0.70000			
MAX FF fingerprint Query 44	--rape with CLUSTER 3	rape FF=0.63889			
MAX FF fingerprint Query 45	--rape with CLUSTER 3	rape FF=0.59306			
MAX FF fingerprint Query 46	--rape with CLUSTER 3	rape FF=0.56806			
MAX FF fingerprint Query 47	--rape with CLUSTER 3	rape FF=0.56250			
MAX FF fingerprint Query 48	--rape with CLUSTER 4	rape FF=0.56667			
MAX FF fingerprint Query 49	--rape with CLUSTER 3	rape FF=0.49583			
MAX FF fingerprint Query 50	--rape with CLUSTER 5	rape FF=0.73333			
MAX FF fingerprint Query 51	--rape with CLUSTER 3	rape FF=0.56389			
MAX FF fingerprint Query 52	--rape with CLUSTER 3	rape FF=0.54306			
MAX FF fingerprint Query 53	--corn with CLUSTER 6	corn FF=0.65000			
MAX FF fingerprint Query 54	--corn with CLUSTER 6	corn FF=0.73333			
MAX FF fingerprint Query 55	--corn with CLUSTER 7	corn FF=0.70000			
MAX FF fingerprint Query 56	--corn with CLUSTER 8	corn FF=0.76667			
MAX FF fingerprint Query 57	--corn with CLUSTER 6	corn FF=0.68333			
MAX FF fingerprint Query 58	--corn with CLUSTER 6	corn FF=0.63333			
MAX FF fingerprint Query 59	--corn with CLUSTER 7	corn FF=0.70000			
MAX FF fingerprint Query 60	--corn with CLUSTER 8	corn FF=0.76667			
MAX FF fingerprint Query 61	--soy bean with CLUSTER 9	soy bean FF=0.70526			
MAX FF fingerprint Query 62	--soy bean with CLUSTER 9	soy bean FF=0.66316			
MAX FF fingerprint Query 63	--soy bean with CLUSTER 6	corn FF=0.55000-->	NO MATCH		
MAX FF fingerprint Query 64	--soy bean with CLUSTER 9	soy bean FF=0.65263			
MAX FF fingerprint Query 65	--soy bean with CLUSTER 9	soy bean FF=0.64211			
MAX FF fingerprint Query 66	--soy bean with CLUSTER 9	soy bean FF=0.63509			
MAX FF fingerprint Query 67	--soy bean with CLUSTER 9	soy bean FF=0.68772			
MAX FF fingerprint Query 68	--soy bean with CLUSTER 9	soy bean FF=0.66667			
MAX FF fingerprint Query 69	--soy bean with CLUSTER 9	soy bean FF=0.62105			
MAX FF fingerprint Query 70	--soy bean with CLUSTER 9	soy bean FF=0.51579			
MAX FF fingerprint Query 71	--soy bean with CLUSTER 9	soy bean FF=0.61404			
MAX FF fingerprint Query 72	--soy bean with CLUSTER 9	soy bean FF=0.47368			
MAX FF fingerprint Query 73	--soy bean with CLUSTER 9	soy bean FF=0.59298			
MAX FF fingerprint Query 74	--soy bean with CLUSTER 9	soy bean FF=0.65965			
MAX FF fingerprint Query 75	--soy bean with CLUSTER 9	soy bean FF=0.67018			
MAX FF fingerprint Query 76	--soy bean with CLUSTER 9	soy bean FF=0.71930			
MAX FF fingerprint Query 77	--soy bean with CLUSTER 9	soy bean FF=0.69474			
MAX FF fingerprint Query 78	--soy bean with CLUSTER 9	soy bean FF=0.68070			
MAX FF fingerprint Query 79	--soy bean with CLUSTER 9	soy bean FF=0.63158			
MAX FF fingerprint Query 80	--palm with CLUSTER 10	palm FF=0.76667			
MAX FF fingerprint Query 81	--palm with CLUSTER 10	palm FF=0.73333			
MAX FF fingerprint Query 82	--palm with CLUSTER 10	palm FF=0.71111			
MAX FF fingerprint Query 83	--palm with CLUSTER 10	palm FF=0.58889			
MAX FF fingerprint Query 84	--palm with CLUSTER 10	palm FF=0.73333			
MAX FF fingerprint Query 85	--palm with CLUSTER 10	palm FF=0.73333			
MAX FF fingerprint Query 86	--peanut with CLUSTER 11	peanut FF=0.52889			
MAX FF fingerprint Query 87	--peanut with CLUSTER 11	peanut FF=0.67556			
MAX FF fingerprint Query 88	--peanut with CLUSTER 12	peanut FF=0.86667			
MAX FF fingerprint Query 89	--peanut with CLUSTER 11	peanut FF=0.74667			
MAX FF fingerprint Query 90	--peanut with CLUSTER 11	peanut FF=0.68000			
MAX FF fingerprint Query 91	--peanut with CLUSTER 11	peanut FF=0.76000			
MAX FF fingerprint Query 92	--X-mix-rape with CLUSTER 13	X-mix-rape FF=0.86667			
MAX FF fingerprint Query 93	--X-mix-sunfl with CLUSTER 14	X-mix-sunfl FF=1.00000			
MAX FF fingerprint Query 94	--X-mix-rape with CLUSTER 15	X-mix-rape FF=1.00000			
MAX FF fingerprint Query 95	--X-mix with CLUSTER 16	X-mix FF=1.00000			
MAX FF fingerprint Query 96	--X-mix with CLUSTER 17	X-mix FF=1.00000			
MAX FF fingerprint Query 97	--X-mix-sunfl-soy with CLUSTER 18	X-mix-sunfl-soy FF=1.00000			
=====					
MAX FF fingerprint Query 98	--X-mix-soy-rape with CLUSTER 17	X-mix FF=0.46667-->	NO MATCH		

Listing 2. Internal validation by comparing the 98 Boolean query fingerprints with fuzzy fingerprints of the 18 clusters generated by the Butina method. The first query is presented to all 18 fuzzy fingerprints, the other queries with maximal **FF**.