From the previous exercise, take the following table:

Table 1 with fingerprints:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1 |  |  |  | -OH | -Cl | -NO2 | -CH3 | -COOH |
| 2,4,6-trinitrophenol | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 2,3-dinitrophenol | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 3-hydroxybenzaldehyd | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2,4,6-trimethylphenol | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |

1. Count Tanimoto coefficient for pairs:
	* 2,4,6-trinitrofenol and 2,3-dinitrofenol:
	* 2,4,6-trinitrofenol and 3-hydroxybenzaldehyd:
2. Fill values of Tanimoto coefficient into the table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2 | 2,4,6-trinitrophenol | 2,3-dinitrophenol | 3-hydroxy-benzaldehyd | 2,4,6-trimethylphenol |
| 2,4,6-trinitrophenol |  |  |  |  |
| 2,3-dinitrophenol |  |  |  |  |
| 3-hydroxybenzaldehyd |  |  |  |  |
| 2,4,6-trimethylphenol |  |  |  |  |

1. Which pair(s) have the Tanimoto coefficient the lowest and the highest?
2. Count the Cosine coefficient for the pairs:
	* 2,4,6-trimethylphenol and 2,3-dinitrophenol:
	* 2,4,6-trimethylphenol and 3-hydroxybenzaldehyd:
3. Count (binarny) Euclid distance for pairs:
	* 2,4,6-trinitrophenol and 2,3-dinitrophenol:
	* 2,4,6-trinitrophenol and 3-hydroxybenzaldehyd:
4. Do consensus fingerprints for fingerprints from Table 1 (use OR):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1 |  | -NO2 | -OH-COOH | -Cl-CH3 |
| 2,4,6-trinitrophenol |  |  |  |  |
| 2,3-dinitrophenol |  |  |  |  |
| 3-hydroxybenzaldehyd |  |  |  |  |
| 2,4,6-trimethylphenol |  |  |  |  |

1. Use the charges on atoms O and H from the previous exercise:

Table with charges:

|  |  |  |
| --- | --- | --- |
| Table 3 | Charge on O (from OH skupiny) | Charge from H (from OH skupiny) |
| 2,4,6-trinitrophenol | -0,5690 | 0,5670 |
| 2,3-dinitrophenol | -0,6110 | 0,4640 |
| 3-hydroxybenzaldehyd | -0,6910 | 0,4060 |
| 2,4,6-trimethylphenol | -0,7140 | 0,3670 |

1. Create a descriptor space graph in Excel, specifically: Put a charge on the x-axis at O, a charge on the y-axis at H. Show a graph that shows the X and Y points.

Note: If you are not familiar with Excel, do the following: Copy the above table into Excel. Label the values of the hubs. Put "Insert" and select "Charts" and select the X,Y chart.

1. Count the Euclid distance:
	* 2,4,6-trinitrophenol and 2,3-dinitrophenol:
	* 2,4,6-trinitrophenol and 3-hydroxybenzaldehyd:
2. Which pair has the biggest Euclid distance?
3. Which pair has the lowest Euclid distance?
4. Which Euclid distance is more informative? Binary or with real numbers?