

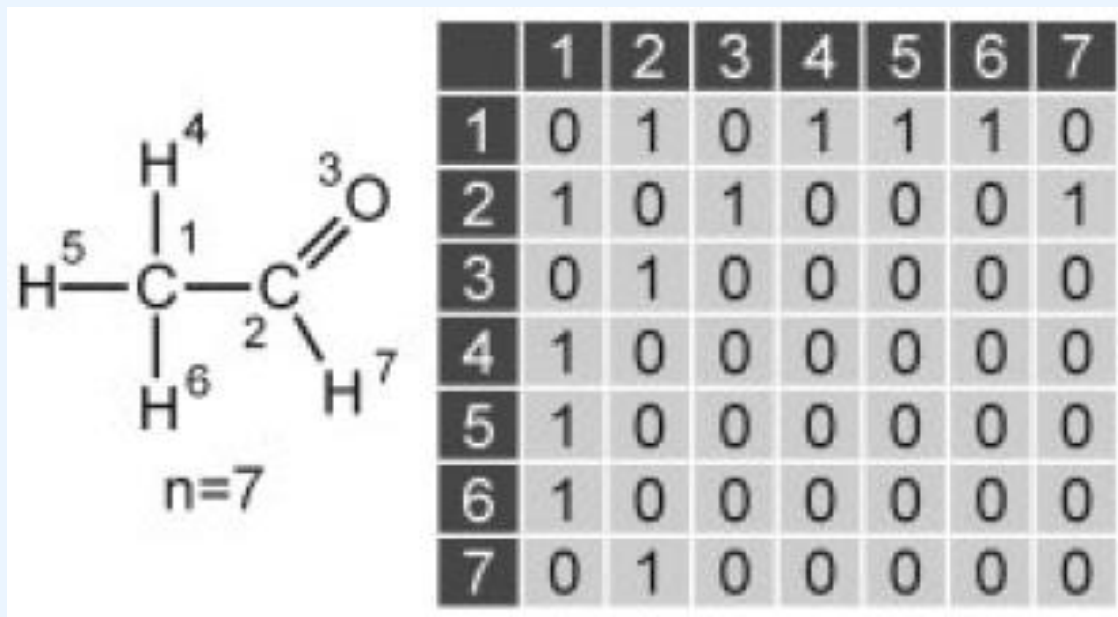
Двумерная форма  
хранения и обработки информации

# Матричная форма представления молекулярного графа

## Матрица смежности

Атомы нумеруются произвольно.

$n$  атомов – матрица размерности  $n \times n$

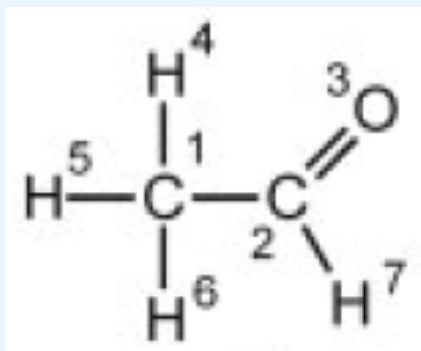


Элементы матрицы:

$a_{ij} = 1$ , если между атомами  $i$  и  $j$  имеется химическая связь,

$a_{ij} = 0$ , если между атомами  $i$  и  $j$  нет химической связи.

# Избыточная – неизбыточная матрица



	1	2	3	4	5	6	7
1	0	1	0	1	1	1	0
2	1	0	1	0	0	0	1
3	0	1	0	0	0	0	0
4	1	0	0	0	0	0	0
5	1	0	0	0	0	0	0
6	1	0	0	0	0	0	0
7	0	1	0	0	0	0	0

	1	2	3	4	5	6	7
1		1		1	1	1	
2	1		1				1
3		1					
4	1						
5	1						
6	1						
7		1					

Этапы упрощения матрицы:

удаление нулей,

устранение дублей,

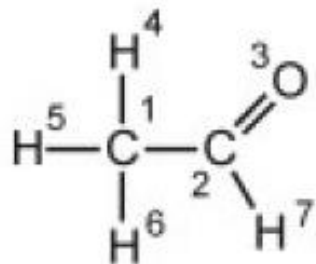
удаление информации об атомах водорода.

	1	2	3	4	5	6	7
1		1		1	1	1	
2			1				1
3							
4							
5							
6							
7							

	1	2	3
1		1	
2			1
3			

# Матрица расстояний

(примеры иных типов матриц)



a)

	C1	C2	O3	H4	H5	H6	H7
C1	0	1.400	2.190	1.022	1.023	1.022	2.106
C2	1.400	0	1.123	1.999	1.982	1.999	1.022
O3	2.190	1.123	0	2.349	2.708	2.995	1.859
H4	1.022	1.999	2.349	0	1.668	1.661	2.895
H5	1.023	1.982	2.708	1.668	0	1.668	2.562
H6	1.022	1.999	2.955	1.661	1.668	0	2.336
H7	2.106	1.022	1.859	2.895	2.566	2.336	0

b)

	C1	C2	O3	H4	H5	H6	H7
C1	0	1	2	1	1	1	2
C2	1	0	1	2	2	2	1
O3	2	1	0	3	3	3	2
H4	1	2	3	0	2	2	3
H5	1	2	3	2	0	2	3
H6	1	2	3	2	2	0	3
H7	2	1	2	3	3	3	0

Chemoinformatics: A Textbook. Ed. J.Gasteiger, T.Engel. 2003

а) геометрическое расстояние (ангстрем);

б) топологическое расстояние (число связей по кратчайшему пути)

# Проблема разрастания объема базы данных

В матрице смежности:

$$\text{Число элементов матрицы} = f(n^2)$$

Нерационально для больших молекул.

Значительно лучше, если:

$$\text{Число элементов} = f(n^1)$$

Это достигается в форме

таблицы соединений ([connection table](#)).

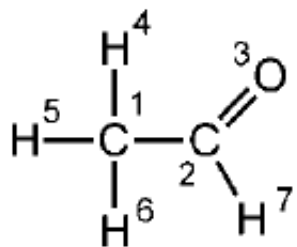
# Таблица соединений -1

Таблица соединений – отображение состава вещества и связей между атомами в табличной форме.

Пример: этаналь.

Пронумеровать атомы в производном порядке.

Один из путей: Заполнить две таблицы.



Список атомов	
1	C
2	C
3	O
4	H
5	H
6	H
7	H

Список связей		
1-й атом	2-й атом	Порядок связи
1	2	1
2	3	2
2	7	1
1	4	1
1	5	1
1	6	1

# Таблица соединений - 2 (избыточная)

Пример: этаналь; второй путь.

Пронумеровать атомы в производном порядке.

Заполнить одну таблицу.



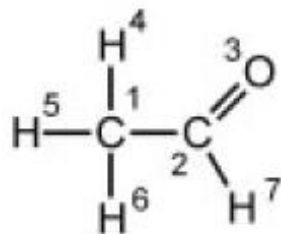
№	Атом	Сосед № 1	Порядок связи	Сосед № 2	Порядок связи	Сосед № 3	Порядок связи	Сосед № 4	Порядок связи
1	C	2	1	4	1	5	1	6	1
2	C	1	1	3	2	7	1		
3	O	2	2						
4	H	1	1						
5	H	1	1						
6	H	1	1						
7	H	2	1						

Информативность избыточна, т.к. каждый атом упоминается дважды, сведения о водороде стандартны.

# Таблица соединений - 2 (неизбыточная)

Если убрать повторы, сжать, получаем:

В случае  
"обычных"  
органических  
соединений  
полезная  
информация  
при этом  
не теряется.



№	Атом	Сосед № 1	Порядок связи	Сосед № 2	Порядок связи
1	C	2	1		
2	C			3	2
3	O				

⇒

№	Атом	Сосед № 1	Порядок связи
1	C	2	1
2	C	3	2
3	O		



Информация о структуре:

- из измерительной аппаратуры,
- из молекулярных редакторов,
- из программ расчета.

Форматы разнообразны,  
необходим стандарт обмена информацией.

Де-факто:

**MOL-файлы**

*abcde.mol*

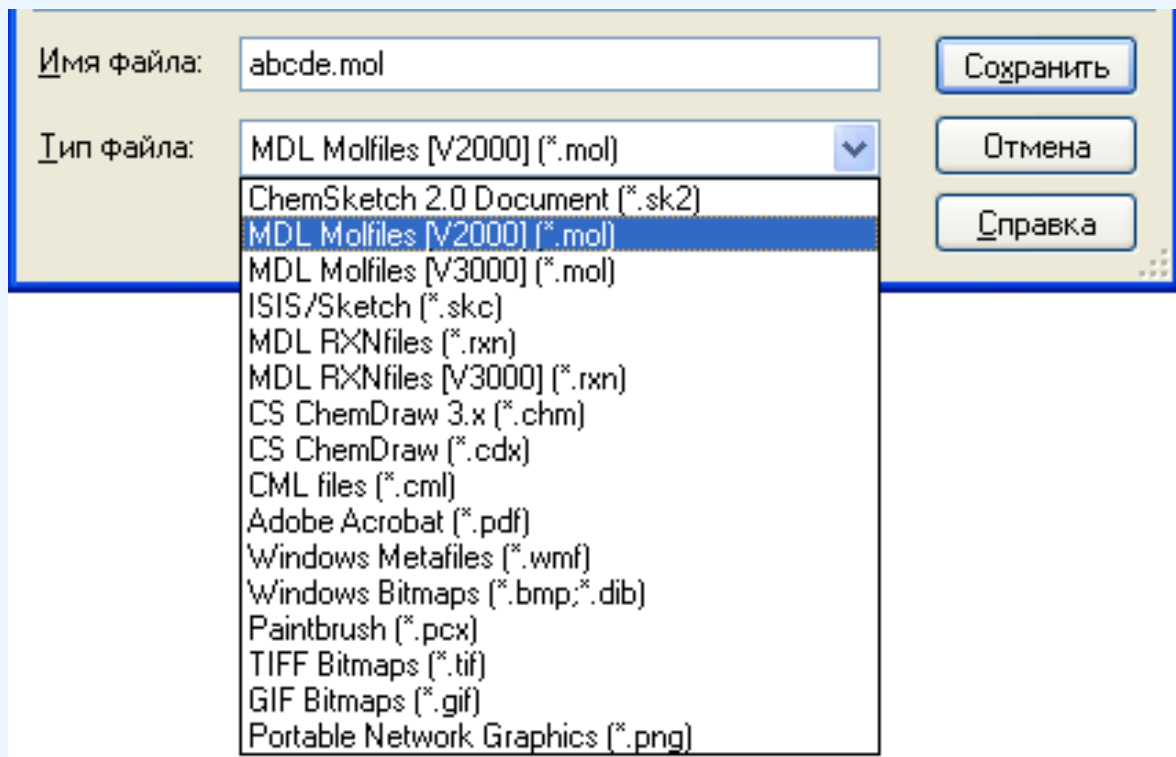
(есть варианты).

В основе

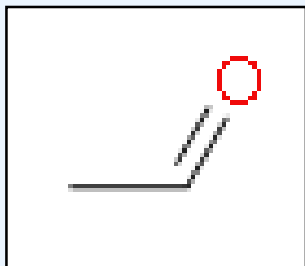
MOL-файла –

таблица

соединений.



# MOL-файл (2D, без атомов H)



3 атома

2 связи

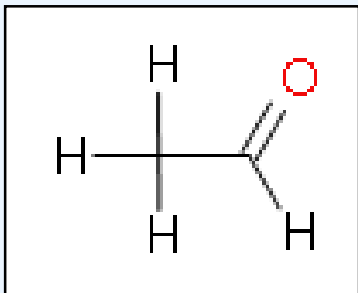
СПИСОК АТОМОВ

ИНЫЕ ПАРАМЕТРЫ

```
SMXDraw1201020342D
3 2 0 0 0 0 0 0 0 0999 V2000
4.7059 -6.9865 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5.5327 -6.9865 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5.9461 -6.2705 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0 0
2 3 2 0 0 0 0
M END
```

координаты  
x, y, z

СПИСОК СВЯЗЕЙ



## MOL-файл (2D, с атомами H)

SMMXDraw01201020342D

```

7  6  0  0  0  0  0  0  0  0999 V2000
  4.9749   -5.2654   0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
  5.7945   -4.4244   0.0000 H   0  0  0  0  0  0  0  0  0  0  0
  5.8108   -5.9984   0.0000 H   0  0  0  0  0  0  0  0  0  0  0
  5.8016   -5.2654   0.0000 C   0  0  3  0  0  0  0  0  0  0  0
  7.0418   -5.9814   0.0000 H   0  0  0  0  0  0  0  0  0  0  0
  7.0418   -4.5494   0.0000 O   0  0  0  0  0  0  0  0  0  0  0
  6.6284   -5.2654   0.0000 C   0  0  0  0  0  0  0  0  0  0  0

4  1  1  0  0  0  0
4  2  1  0  0  0  0
4  3  1  0  0  0  0
7  4  1  0  0  0  0
7  5  1  0  0  0  0
7  6  2  0  0  0  0

```

M END