**Наполнение Базы.**

Вся информация хранится в химических базах с определенным форматом (\*.sdf). По сути это текстовый файл разбит по определенному принципу. Сначала идет структура (у нас это поле mdl\_form). За ней следуют теги, что определяют название поля и потом собственно значение этого поля. Каждая запись заканчивается символами - $$$$

ниже привел пример еденичной записи, в добавленых файлах будет пример этой базы.

**Наполнение и обновление базы** что на сайте должно идти с такого файла.

То есть на сайте есть пустая база в которой указаны лишь название полей. Название полей совпадают с тегами химической базы. Затем значения по тегам переписываются в базу сайта.

Операции наполнение и обновление должны **отличаться**.

**Наполнение** – это полная перепись базы что на сайте. (То есть, все поля удаляются, остаются лишь их название. На их место записывается новая база).

**Обновление** – обновление лишь части записей.

Обновление происходит следующим образом. По идентификатору (cat\_namber) находится запись и перезаписываются поля, что есть в базе которую импортируют. Если каких-то полей в базе что импортируют нету, а в базе сайта есть – они остаются без изменений.

**Удаление** – происходит по списку идентификаторов cat\_namber.

**!!!!! хочу отдельно акцентировать**, что база будет большой и стоит вопрос о ее продуктивности (**время наполнения, обновления, удаления, и генерации самой страницы, то есть основного функционала**). Если говорить о нашей \*.sdf – базе которая в себе имеет все то, что и должна иметь база сайта, то речь идет о ~ 20 ГБ. И вот об этом хочу услышать Ваши соображения!!!!

Mrv0541 11131515092D

26 28 0 0 0 0 999 V2000

1.9422 -0.9360 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

2.6007 -2.1013 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0

1.9348 -1.7129 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

1.2134 0.3108 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

3.2888 -0.9471 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

1.9422 0.7288 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

1.2134 -0.5216 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

0.4921 0.7399 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0

-0.2331 0.3218 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

3.2851 -1.7202 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

1.9422 1.5649 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0

-0.2331 -0.5142 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0

-1.6943 0.3218 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0

-0.9323 0.7510 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

2.6636 0.3070 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0

-2.4268 0.7399 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

-3.8807 1.5833 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

-4.5909 2.0162 0.0000 Cl 0 0 0 0 0 0 0 0 0 0 0 0

-3.1815 0.2959 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

-2.4268 1.5649 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

-3.8807 0.7399 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

-3.1371 1.9829 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

3.9769 -0.5771 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

3.9547 -2.1198 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

4.6317 -0.9323 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

4.6428 -1.7351 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0

2 3 1 0 0 0 0

3 1 2 0 0 0 0

4 7 1 0 0 0 0

5 1 1 0 0 0 0

6 4 1 0 0 0 0

7 1 1 0 0 0 0

8 4 1 0 0 0 0

9 8 1 0 0 0 0

10 5 1 0 0 0 0

11 6 2 0 0 0 0

12 9 2 0 0 0 0

13 14 1 0 0 0 0

14 9 1 0 0 0 0

15 6 1 0 0 0 0

16 13 1 0 0 0 0

17 22 1 0 0 0 0

18 17 1 0 0 0 0

19 16 2 0 0 0 0

20 16 1 0 0 0 0

21 19 1 0 0 0 0

22 20 2 0 0 0 0

23 5 2 0 0 0 0

24 10 2 0 0 0 0

25 23 1 0 0 0 0

26 25 2 0 0 0 0

2 10 1 0 0 0 0

24 26 1 0 0 0 0

21 17 2 0 0 0 0

M END

> <CdId>

1

> <Mol Weight>

372.802

> <Formula>

C19H17ClN2O4

> <id>

3387

> <Catalog\_namber>

1.1042024E8

> <Purity>

96

> <Molecular\_Formula>

C19H17ClN2O4

> <Available\_from\_stock>

35

> <CAS\_number>

13333-86-3

> <MDL\_number>

MFCD00179413

> <SMILES>

OC(=O)C(Cc1c[nH]c2ccccc12)NC(=O)COc3ccc(Cl)cc3

> <Status>

in stock

> <price1mg>

23

> <price2mg>

16

> <price3mg>

57

> <price4mg>

23

> <price5mg>

22

> <price10mg>

16

> <price15mg>

57

> <price20mg>

16

> <price25mg>

17

> <price5umol>

22

> <price10umol>

17

> <price20umol>

53

$$$$