

3.4.2.5 生成子结构索引码

化合物子结构索引服务用于生成化合物结构的子结构索引码，建设子结构检索索引。子结构索引码是一种基于 3/6 原子结构片断的编码¹，支持指定连接位置或任意连接位置的子结构检索功能。建议在完成化合物建库后使用本服务，使用化合物结构建库服务输出作为数据源。

化合物子结构索引服务使用 Access 数据库 STR_Frag 表单的 F1 格式化合物结构数据作为输入数据，生成化合物子结构索引数据，以 Access 数据库格式输出。由于生成化合物子结构索引码需要较长的时间，并且产生的子结构索引码空间较大，建议每个化合物子结构索引服务的数量不超过 5000 个化合物，必要时可以分段处理。化合物子结构索引服务不生成新的 Access 数据库，只在作为输入数据的 Access 数据库中增加 STR_Sub_Molecu_Code1 和 STR_Sub_Molecu_Code2 两个表单用于存储输出的子结构索引码。

可以选择 "单个 F1 格式文件生成子结构索引码" 或者 "多个 F1 格式文件批处理生成子结构索引码"。

使用此服务必须是存储了化合物 F1 格式数据的 Access 数据库。在数据表中，每个化合物占 1 条记录。

使用方法与示例：

多个化合物 F1 格式文件生成子结构索引 用户在图 3.4.2.16 中点击[多个 F1 格式文件批处理生成子结构索引码](#)，即可打开转换页面，如图 3.4.2.17。

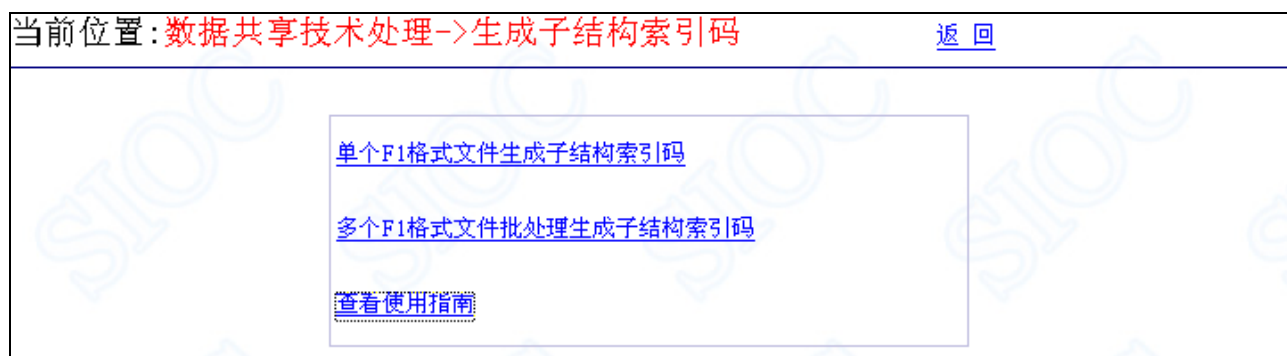


图 3.4.2.16 生成子结构索引服务的首页

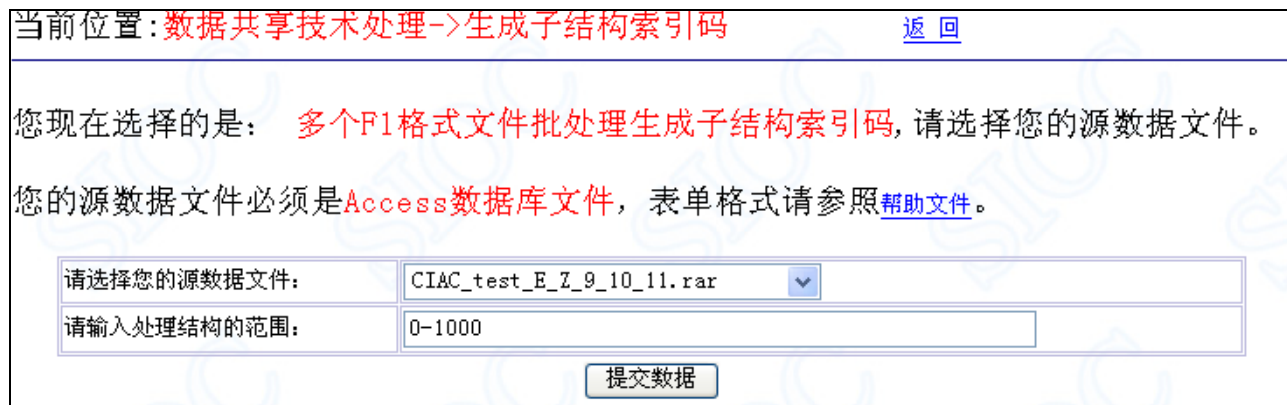


图 3.4.2.17 生成子结构索引

化合物建库作业完成后，生成 Access 数据库，并已获得结果，那么就可以直接选择此结果文件作为源文件，本例中就采用上一节 3.4.2.4 结构登录的结果 CIAC_test_E_Z_9_10_11.rar 中作为源数据文件，如图 3.4.2.17 和图 3.4.2.18。

如果用户有自己的结构数据库，使用此服务前须将其转换成 Access 数据库，然后提交作业。

由于此作业需要较长时间，如果结构太多，可分批次处理，每次输入处理结构的 ID 范围。

程序会将结果数据文件压缩成一个新的 rar 文件，如 CIAC_test_E_Z_9_10_11_12.rar...

子结构索引结果保存在用户上传的 Access 数据库，表 STR_Sub_Molecu_Code1 和 STR_Sub_Molecu_Code2 中，如图 3.4.2.18。

¹ 李英勇，陈维明。一种有效的子结构索引方法。《科学数据库与信息技术论文集》，第八集，中国科学院科学数据库办公室编，中国环境科学出版社，2006，北京，116-120

Microsoft Access

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 记录(R) 工具(T) 窗口(W) 帮助(H)

Adobe PDF

CompStruct : 表

ID	Source	Mol	F1	SRN
1	10015		DI0100000004c1	148198042
2	10016		DI0100000004c1	148198051
3	10177		DI01000000057f	148198069
4	1030		DI01000000063b	148198077
5	10371		DI010000000650	148198085
6	10374		DI010000000996	148198093
7	10375		DI010000000a8b	148198106

“数据表”视图

图 3.4.2.18 生成子结构索引的源数据

Microsoft Access

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 记录(R) 工具(T) 窗口(W) 帮助(H)

t_E_Z_9 : 数据库 (Acces...)

设计(D) 新建(N) | X |

- STR_Sub_Molecu_Code1
- STR_Sub_Molecu_Code2

STR_Sub_Molecu_Code1 : 表

ID	Code	nAtom	nBond	nMAtom	nXAtom	nZAtom
1	ww b4'21'021 b4	6	5	0	0	2
2	ww b4'21'021 b4	6	5	0	0	2
3	ww bCIC'12'J2 b	8	7	0	2	0
4	ww bCOC'15'n bC	9	8	0	0	3
5	ww bCOC'e'101 b	9	8	0	0	2
6	ww bCOC'e'101 b	14	14	0	0	2
7	ww bCOC'15'n bC	16	16	0	0	2
8	ww bCNC'21'012	11	10	0	0	3
9	ww bCNC'21'o0 b	12	11	0	0	3

化学片段登录号

图 3.4.2.19 子结构索引的结果