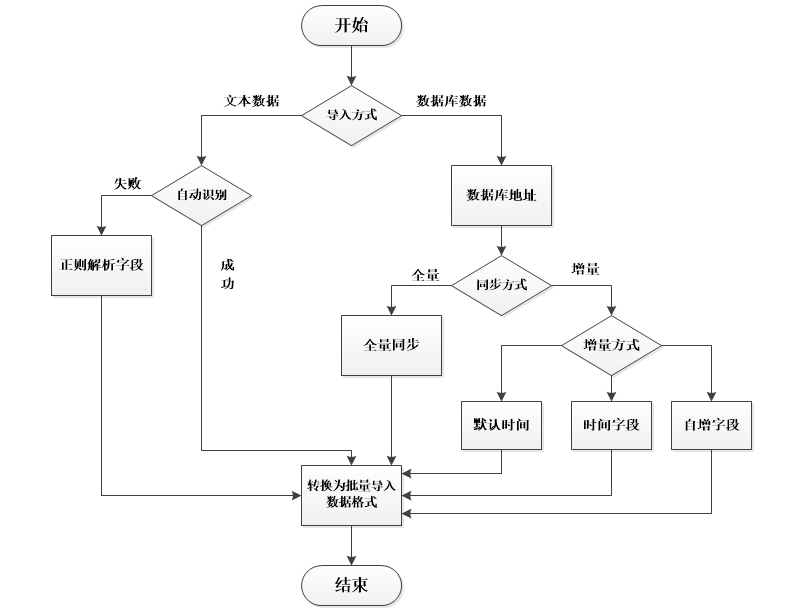
### 数据信息抽取模块

数据信息抽取模型为数据存储与搜索平台提供数据，提供文本数据导入及数据库数据同步等导入数据的方式，并提供相应的字段提取。文本数据导入方式，首先根据内置的算法自动识别文件数据类型，如自动匹配失败，提供正则提取字段的方式来提取数据字段内容；正则提取字段时，通过正则分组的方式来提取字段，默认是一行数据作为一个doc的message，同时，也提供多行匹配的方式来提取字段，可以将正则提交到内置的规则库，下次导入即可自动识别该种格式的文本数据。数据库数据同步方式，提供全量以及增量同步两种方式，全量同步通过数据库设置以及字段配置等来实现数据的全量同步；增量同步方式提供三种增量同步方式，一是基于时间的增量同步，二是基于时间字段的增量同步，三是基于自增字段的增量同步，每种方式都可以配置时间间隔来用定时器完成对数据的增量同步。通过以上的配置，将数据转换成es能够识别的数据格式，通过es的bulkAPI 进行数据导入。



### 索引构建模块

在对数据进行导入之前，需要先确定好数据的mapping。以数据库数据为例，数据库对应es中的索引，表对应es的类型，需要确定es的类型中的数据类型格式，否则，采用自动映射的方式的话，可能会导致es与数据库中的数据类型不一致。同时，需要注意的一点是，es中不同类型中的同名字段，字段类型需要一致，不一致的话数据首先可能会无法导入，其次导入之后操作也会出现问题。