

SMT数据统计规范

高会军，孙一勇，马志强，李政凯 *

05,03,2017

1 总则

目的：规范，未来方便生成各类报表、图表。统一文件路径、数据命名格式等。

统计内容纷繁庞杂，未来统计目的不甚明确，需要有利与我的报告。由于所统计数据包罗各个行业各个领域，仅作粗略数据采集难以满足后期数据统计和处理需要：以国内贴片机为例，现有数据中至少30+ 企业、10+研究机构和高校正在从事（或从事过）贴片机研发和生产相关的项目，其中仅产品型号一项至少有200 项之多。鉴于此，特制定本规范对数据管理、命名进行规则化处理。数据命名须尽量遵循以下原则：

- 1：利于软件调用、方便生成各类格式的图表，每个厂家(组织)、每个型号的信息作为数据库的一部分，统一格式命名，各自对应其数据，有条不紊，比如:1)统计现有数据库中海外型号贴片机中日本厂家，能够很快地得出一共有N家，包括有松下、JUKI...等数据; 2)统计现有国内贴片机中支持电动飞达的厂家，通过查询数据就可以极快调动出现有数据库中此类厂家名称、支持的型号有哪些；
- 2：方便拓展，能够及时对现有的数据库做出补充，比如：1)某厂家2016年面世一款型号产品，现有数据库中并没有囊括这类型号，一旦发现可以及时添加此类信息，并能够统计这类数据的来源； 2)已知某不完整或不可信数据，那么须对该数据已知部分进行统计，未统计到部分可为空。
- 3：分类尽量完整，数据中可以反映一些联系，比如：1)现有统计数据中仅有贴片机厂家，不包含某研究机构或高校，发现H高校和A企业正在合作(或曾经合作过)，高校和企业属于并列关系，通过查询H高校就可以查找出H高校和A企业的合作关系; 2) A企业和B企业都在

*哈尔滨工业大学智能控制与系统研究所

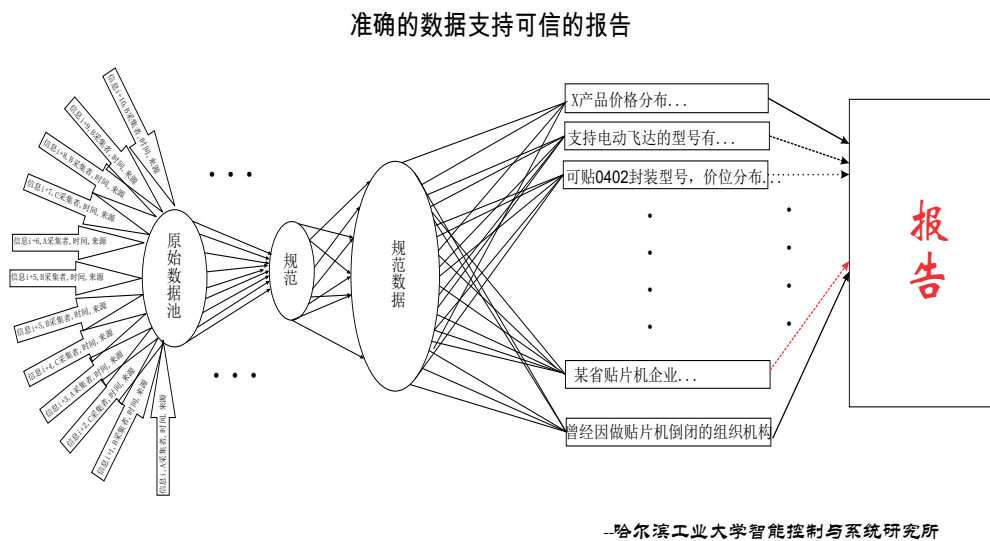


Figure 1: 整体结构

使用某S外企的软件界面、J企业的飞达、可与L企业的优化软件进行联调，但A企业与W企业的后端产品由于机械尺寸、软件环境等原因不能进行联合使用，发现这类信息，分别添加到A、S、J、W企业的相关产品信息下。

- 4：本规范为所有数据的地图，通过查询本数据规范，可以直接知道如何调用任一数据，比如数据格式为：外企-美国-A企业-A1高速机-0.05mm精度；A2中速机-0.04mm精度；A3低速机-高精度，大片，0.001mm精度等几类型号：A1：仅支持电动飞达...。查询时通过下表快速即可以知道如何获取A企业的A22型号的所有信息。即表可以看出任意一个点上数据的调用方式。
- 5：考虑通性，所有企业共有的一些数据在同类中进行命名时须尽量使用同一命名方式，方便数据读写和更改。比如：A、B、C企业均为外企，D、E、F企业均为国内民企，那么他们的员工结构(-02)、某类产品(-08)在这些企业下的子类须为同一格式，即，员工结构分别为 国外-地域-A（或B或C）-02，国内-地域-D（或E或F）-02，改类产品分别为 国外-地域-A（或B或C）-08，国内-地域-D（或E或F）-08。

2 贴片机数据格式规范

数据格式：一级，二级，三级...

数据编号格式：国内外－地域(所在城市)－公司、研究机构－产品内容、人员结构－LED专用贴片机、泛用机、快速机、桌面机－贴片机数据、了解数据－数据子类

贴片机数据包括：机器型号、重量、价格、精度、速度、机械结构、传动结构、喂料器、气源、电源、贴装尺寸、

贴片机：LED贴片机、泛用机，贴插机，国内、国外，

注解 1 *CC*-当前待确定数据，*XX*-或者数字代表当前已经确立数据。

CC- 国内外：

- 00未知
- 01国外
- 02国内
- ...

地域：（主要是产品对应公司所在地，比如A集团B公司的产品进行调研，以B公司的所在地为准）

01-CC 国外分类

- 00未知
- 01德国
- 02美国
- 03日本
- 04新加坡
- ...

02-CC 国内分类

- 00未知
- 01中国台湾
- 02深圳
- 03广州
- 04北京

- 05苏州
- 06杭州
- 07哈尔滨
- ...

XX-XX-CC组织性质： 主要是指该公司或者团体的性质

- 00未知
- 01上市公司
- 02公司
- 03研究所
- 04高校
- 05学术团体
- 06民间组织
- ...

XX-XX-XX-CC 统计物性质 被统计内容性质：（主要是产品对应的属性）

- 00未知
- 01研究报告
- 02专利和论文
- 03优化软件
- 04贴片机
- 05丝印机
- 06接驳台
- 07外围产品(飞达、焊锡、焊锡膏)
- 08波峰焊台
- 09机器人

- 10机构历史
- 11组织规模
- 12机构人员
- 13机构资金
- ...

XX-XX-XX-12-CC 机构人员

- 00未知
- 01总人数
- 02研究人员总量
- 03研究人员比例(博-硕-本-高中及以下)
- ...

XX-XX-XX-11-CC 机构资金

- 00未知
- 01注册资金
- 02销售额
- 03研发资金
- 04盈利情况
- ...

3 信息来源

每条信息都有其一定的可信程度，在数据统计过程中不对其置信程度做任何评价，但是要对每一条信息的来源进行说明。仅作信息收集，不做信息处理。如此，一方面保证信息数据库的包容性，另一方面对信息后期处理进行一定程度的置信分析。同时，原始数据应当保留在另外一个库中，注明数据收集人姓名。

比如：有关A组织机构现有200条信息，其中100条是通过网站、口述得到的，另外100条是通过该组织机构的宣传海报、专利授权、网站等内容获得的。在整理报告时，可以对两部分数据进行差别统计和处理。

信息来源：

- 00未知
- 01同上一级
- 02公司网站
- 03电话
- 04听人口述
- 05文献资料
- ...

注解 2 所有信息来源的文献、网站、以及可再次获取的信息，须在文末以参考资料的形式给出。尽量做到数据可多次复查、禁得住查。比如，某品牌产品在某论坛口碑不佳、易损，统计此类信息时，则是该品牌的口碑信息之一，在参考资料一项中需给出其评价信息的网页链接。