

附件 1：有机高分子材料数据的整合 规范

有机高分子材料数据的整合规范

国内的有机高分子材料科学数据资源主要以纸质文本、资料的形式存在，几乎每年都会会有许多此方面的数据手册出版发行。然而数据手册之间的数据同质性较高，且彼此之间存在许多矛盾数据，凸显其数据质量参差不齐的问题，数据可靠性有待检验。另外，手册中不同牌号产品的数据缺乏科学性和系统性，数据条块分割、信息不完整。

然而准确性、可靠性、完整性是数据服务的前提，尤其是工程材料的重要数据和关键数据。因此从文献、手册等媒体介质中获取的原始数据资源需要进行整合和评估才能入库和共享。在数据资源整合和重构时，需要考虑两个维度的整合，即历史数据和新生成数据的纵向整合，和不同分类、不同类型数据的横向整合。此外还要注重解决冗余、冲突、互补拼接的问题。

1 数据资源内容

本节点初期主要搜集和整合一些已有的较为成熟的数据，包括一些专业书籍、手册、企业产品目录，以及本课题组和合作单位在研究过程中产生的一些测试数据和谱图。在节点建设较为完善之后，将进一步考虑逐步在数据库中纳入一些专业期刊文献中报道的测试数据。

表 1列出了本节点重点规划整合的部分数据资源。

2 数据整合的方法

从不同来源获取的数据往往具有不同的数据格式和存在形式，尤其是对于有机高分子材料的各种性能数据，由于大部分数据的提供方

是各个企业，而国内生产塑料、橡胶、纤维、胶黏剂、有机涂料和助剂的大中型生产厂家多达数百个，它们提供的数据往往形式各异，没有统一的规范。因此，国内出版的各种数据手册所收集的数据也往往有类似的问题，整合难度很大。

表 1 初期需要重点整合的部分数据资源

分类	数据资源	备注
自有资 源	高分子材料及助剂的红外光谱、核磁共振谱	基础库
	部分牌号材料的性能测试数据和结构表征数据	牌号库
出版物	高分子材料基础，化学工业出版社，2007	基础库
	通用塑料手册，国防工业出版社，2007	牌号库
	世界通用树脂物性大全，深圳普拉斯公司，2010	牌号库
	塑料材料手册，广东科技出版社，2010	牌号库
	化工产品手册 胶黏剂，化学工业出版社，2005	牌号库
	橡胶制品生产手册，化学工业出版社，2006	基础库+牌号库
	化工产品手册 橡胶及橡胶制品，化学工业出版社，2005	牌号库
	化学纤维概论，中国纺织出版社，2005	基础库
	化工产品手册 纺织纤维，化学工业出版社，2005	牌号库
	新型服用纺织纤维及其产品开发，中国纺织出版社，2006	基础库+牌号库
	现代涂料配方设计，化学工业出版社，2000	牌号库
	高分子材料助剂，化学工业出版社，2008	牌号库
	实用塑料助剂手册，化学工业出版社，2007	牌号库
商业资 源	世界通用树脂物性大全，深圳普拉司公司	牌号库
	中国石油、中国石化及其他企业提供的牌号性能数据	牌号库

网络资源	www.chemyq.com 化工词典	基础库
	www.organchem.csdb.cn 工程塑料数据	牌号库
	chemdatas.com 化工产品数据	基础库+牌号库
外链资源	www.matweb.com 材料牌号数据	基础库外链
	www.cmib.org 材料牌号数据	基础库外链
	www.wikipedia.org 维基百科	基础库外链
文献资源	由 CA 和 SCI 检索到的材料牌号性能数据	牌号库
	由 CA 和 SCI 检索到的质谱、核磁共振谱数据	基础库
	由 Derwent World Patent Index 检索到的专利信息	基础库+牌号库

品名: MPs 树脂

测试项目	单位	分析方法	测试条件	SL-801G	SL-803G	SL-805F	SL-838F
密度	g/cm ³	ASTM D792	23°C	1.01-1.03	1.01-1.03	1.01-1.03	1.01-1.03
熔体流动速率	g/10min	ASTM D1238	5kg-200°C	8.00±2.00	8.00±2.00	8.00±2.00	8.50±2.00
硬度 Shore D		ASTM D2240	Shore D	71.0±4.0	71.0±4.0	69.0±4.0	59.0±4.0
屈服拉伸强度	MPa	ASTM D638	50mm/min	≥282	≥23.5	≥24.5	≥13.0
断裂伸长率	%	ASTM D638	50mm/min	≥202	≥160	≥160	≥180
弯曲强度	MPa	ASTM D790	1.7mm/min	≥40	≥34	≥36.0	≥18.0
弯曲模量	MPa	ASTM D790	1.7mm/min	≥1700	≥1600	≥1500	≥1000
悬臂梁冲击强度	J/m	ASTM D256	Notched 23°C	≥27.0	≥27.0	≥40.0	NB3
维卡软化点	°C	ASTM D1525	10N, 50°C/h	≥84	≥80	≥83	≥70
热变形温度	°C	ASTM D648	0.46MPa, 120°C/h	≥75	≥70	≥74	≥62
透光率	%	ASTM D1003	2mm	≥90.0	≥90.0	≥90.0	≥89.0
雾度	%	ASTM D1003	2mm	≤5.0	≤5.0	≤5.0	-
光泽	%	ASTM D523	2mm, 60°	-	-	≥150	-
特点				抗渗性能相当低, 清澈而坚韧, 硬而韧, 对温度变化敏感, 优良的电学性质。			
用途				家庭日常用, 儿童玩具, 文化用品, 医疗器械, 透明展示器, 透明透明鞋材			

(a) 原始表格

Grade_id	Grade_name	Property_name	Value	Unit	Temp.	Standard condition	
3038	SL-838F	拉伸断裂伸长率	>180	%		ASTM D638, 50mm/min	
3038	SL-838F	透光率	>89	%		ASTM D1003, 2mm	
3038	SL-838F	熔体流动速率	8.5±2	g/10min	200°C	ASTM D1238, 5Kg	
3038	SL-838F	密度	1.01-1.03	g/cm ³	23°C	ASTM D792	
3038	SL-838F	悬臂梁冲击强度	不断裂	J/m	23°C	ASTM D256, notched	
3038	SL-838F	拉伸屈服强度	>13	MPa		ASTM D638, 50mm/min	
3038	SL-838F	弯曲弹性模量	>1000	MPa		ASTM D790, 1.7mm/min	
3038	SL-838F	弯曲强度	>18	MPa		ASTM D790, 1.7mm/min	
3038	SL-838F	热变形温度	>62	°C		ASTM D648, 0.46MPa, 120°C/h	
3038	SL-838F	维卡软化点	>70	°C		ASTM D1525, 10N, 50°C/h	
3038	SL-838F	硬度Shore D	59±4.0			ASTM D2240	
3038	SL-838F	特点和用途	抗渗性能相当低...				
3039	PB-5903	透光率	90.5	%		ASTM D1003, 1/8"	
3039	PB-5903	浊度	1.5	%		ASTM D1003, 1/8"	
3039	PB-5903	熔体流动速率	10	g/10min	200°C	D1238, 5Kg	
3039	PB-5903	密度	1.02	g/cm ³	23°C	ASTM D792	
3039	PB-5903	拉伸断裂强度	250	kg/cm ² sup2		ASTM D638, 1/8", 6 mm/min	
3039	PB-5903	弯曲弹性模量	14000	kg/cm ² sup2		ASTM D790, 1/4", 2.8 mm/min	
3039	PB-5903	弯曲强度	320	kg/cm ² sup2		ASTM D790, 1/4", 2.8 mm/min	
3039	PB-5903	Izod无缺口冲击强度	3	kg·cm/ca	23°C	ASTM D256, 1/4"	
3039	PB-5903	热变形温度	71	°C		ASTM D648, 1/4", 120°C/hr	
3039	PB-5903	维卡软化点	87	°C		ASTM D1525, 1/8", 50°C/hr	
3039	PB-5903	硬度Shore D	67				
3039	PB-5903	特点和用途	一般射出级。运...				

(b) 初步整理的牌号性能表

Grade_id	Grade_name	Property_name	Prop_id	Value(numeric)	Min	Max	Unit	Value(text)	Temp.	Standard condition	
3038	SL-838F	拉伸断裂伸长率	303	180			%			ASTM D638, 50mm/min	
3038	SL-838F	透光率	137	89			%			ASTM D1003, 2mm	
3038	SL-838F	熔体流动速率	260	8.5			g/10min	±2	200°C	ASTM D1238, 5Kg	
3038	SL-838F	密度	253	1.01	1.03		g/cm ³ sup3		23°C	ASTM D792	
3038	SL-838F	悬臂梁冲击强度	327				J/m	不断裂	23°C	ASTM D256, notched	
3038	SL-838F	拉伸屈服强度	305	13			MPa			ASTM D638, 50mm/min	
3038	SL-838F	弯曲弹性模量	337	1000			MPa			ASTM D790, 1.7mm/min	
3038	SL-838F	弯曲强度	334	18			MPa			ASTM D790, 1.7mm/min	
3038	SL-838F	热变形温度	374	62			°C			ASTM D648, 0.46MPa, 120°C/h	
3038	SL-838F	维卡软化点	379	70			°C			ASTM D1525, 10N, 50°C/h	
3038	SL-838F	硬度	346	59			Shore D	±4.0		ASTM D2240	
3038	SL-838F	特点和用途	461	抗渗性能相当低...							
3039	PB-5903	透光率	137	90.5			%			ASTM D1003, 1/8"	
3039	PB-5903	浊度	141	1.5			%			ASTM D1003, 1/8"	
3039	PB-5903	熔体流动速率	260	10			g/10min		200°C	D1238, 5Kg	
3039	PB-5903	密度	253	1.02			g/cm ³ sup3		23°C	ASTM D792	
3039	PB-5903	拉伸断裂强度	300	250			kg/cm ² sup2			ASTM D638, 1/8", 6 mm/min	
3039	PB-5903	弯曲弹性模量	337	14000			kg/cm ² sup2			ASTM D790, 1/4", 2.8 mm/min	
3039	PB-5903	弯曲强度	334	320			kg/cm ² sup2			ASTM D790, 1/4", 2.8 mm/min	
3039	PB-5903	Izod无缺口冲击强度	327	3			kg·cm/ca		23°C	ASTM D256, 1/4"	
3039	PB-5903	热变形温度	374	71			°C			ASTM D648, 1/4", 120°C/hr	
3039	PB-5903	维卡软化点	379	87			°C			ASTM D1525, 1/8", 50°C/hr	
3039	PB-5903	硬度	346	67			Shore D				
3039	PB-5903	特点和用途	461	一般射出级。运...							

(c) 牌号性能表整理的最终结果

图 1 牌号性能表的整理过程示例

目前，许多数据库的建库者仍旧是以人工方式向数据库中输入数据，这一方式可以在比较大的程度上保证数据的质量，其缺点是费时费力，在数据库较小时此方法尚可接受，但当数据量达到数以十万甚至百万计的数据时，数据录入将耗费大量的时间和人力。按照保守估计，若每条数据需要一分钟时间进行录入，以正常工作时间计算，十万条数据将耗费一个人近一年的时间才能录入完成，这还没有考虑后期对数据进行整体的规范化和评估过程。因此在本课题中，我们在获得电子化数据之后，尽量采用计算机程序对数据进行自动整理，再通过数据的规范化和评估过程改善数据质量，达到入库要求。

首先对原始资源进行数字化处理，分析其中各种数据(如牌号名称、重要属性)的标志性特征，人工对这些特征进行强化和规范化，同时对待整理的表格等特殊数据进行适当的规范化，以便计算机程序能对不同数据进行区分。然后依据各种数据本身的特点设计程序对其进行识别、分析和整理，对数值和单位进行拆分，并将最终结果插入到数据表合适的位置。

在数据库的材料牌号性能表中，每一条属性数据是作为一条记录存储的，因此原始数据经初步整理的目标是将复杂的数据特别是表格数据拆分转换为“牌号名称——属性名称——属性值”的形式，例如图 1 (a)中的表格共 4 个牌号，各有 15 条属性记录，本表格初步整理后应成为图 1 (b)的形式，成为 4×15 共 60 条记录。而每一条记录除属性值外，还应包括关于此属性值的其他数据，如单位、测试条件等。当此步骤全部或者阶段性完成之后，就需要进一步对数据格式进行处

理，主要目的是将数值数据和文本数据分开存储，将单值数据与区间值数据(如>4，8~9 等)以相应的规范分别记录，将单位、特殊字符等进行规范化处理。数值数据和文本数据两者分开主要是考虑到用户检索时对此两类数据需要输入的检索式不同，对于数值数据是“大于”、“小于”和“等于”，而对文本数据则是“包含”和“不包含”，分开处理后可以减少系统对大量数据类型进行判断的负担。区间值在许多材料属性中存在，比如某材料的电阻率介于 $10^{12} \sim 10^{15} \Omega \cdot \text{cm}$ ，如果不对此类数据设计处理方案，则在用户检索时无法处理。因此在数据整合的最后阶段将数据值分别以数值、最小值、最大值、文本值进行区分，整理完成后成为图 1 (c)中所示的最终结果，之后即可存储入库。

在此部分数据整理时，还会涉及到牌号分类、生产厂家等信息，这些数据因为涉及到数据库记录中“多对多”的关系，需要另表存储，但在初步整理时可以将它们作为牌号的一条属性进行记录，最后在把这些属性单独提取出来，分别合并到牌号分类表(T_Grade)、生产厂家表(T_CompanyName)和牌号生产厂家表(T_Grade_Company)中。

牌号属性数据表是整个数据库的核心内容，包含了所有材料牌号的基本信息、用途、组成和各种使用性能的记录。材料属性数据在材料属性值记录表中以“牌号名称——属性名称——属性值”的形式存储，表格应至少包括如下几个字段(列)：

- **Grade_id**: 为每种材料分配的一个唯一值，以解决某些牌号重名的问题。
- **Grade_name**: 牌号名称。

- **Property_name**: 属性名称, 此值用于页面输出, 不用于检索。
- **Prop_id**: 属性名称所对应的编号, 此值用于检索, 对于某些同义词或者意义接近的词, 比如“杨氏模量”和“拉伸弹性模量”、“挠曲强度”和“弯曲强度”等可使用同样的 **Prop_id** 以便在检索时检索到, 而页面显示时仍以 **Property_name** 为准, 以客观显示原始文献状态, 尽量避免对其涵义的理解出现偏差而造成错误。
- **Value(numeric)**: 确定的数值数据, 以浮点型数据(float)记录。
- **Min**: 数据取值范围的最小值, 以浮点型数据(float)记录。
- **Max**: 数据取值范围的最大值, 以浮点型数据(float)记录, 如果最大值和最小值同时存在, 应保证最大值大于最小值。
- **Unit**: 数值数据和数据取值范围的单位。
- **Value(text)**: 文本数据以及对数值数据的备注(如 \pm 误差值), 文本数据还包括了超链接和表格, 均以 **html** 代码进行标记。
- **Temperature**: 数据的测试时温度。
- **Standard_Condition**: 测试标准以及测试时的其他条件。

附件 2：有机高分子材料数据转换的 规范化

有机高分子材料数据转换的规范化

3 属性名称表数据的整理规范

属性名称表(T_PropertyName)记录了牌号属性数据表中所有引用属性名称的分类、说明和显示顺序信息,表格包含如下几个字段(列):

- Prop_id: 为每种属性分配的一个唯一值。
- Parent_Prop_id: 本属性的上一层属性(父类)所对应的 Prop_id, 如果为空值,则本属性处于分类的最顶层。
- Prop_Layer: 本属性的层级和顺序编号。共 7 位数字,可对三层的属性类别进行控制,其中第一位数字是 1,第二、三位数字是第一层类别(如外观状态、化学组成、物理性质)的序号,第三、四位数字是第二层类别的序号,第五、六位数字是第三层类别的序号。每一层的序号可以有重复也可不重复,如果序号重复则其顺序随机排列,如果采用不重复的序号,每一层可最多有 99 个属性名称。一般来说可仅将属性分两层定义,即第一层和第三层,当第三层属性类别较多时,可在中间定义一个第二层的分类以便使其显示结果更直观和有条理,比如将拉伸强度、断裂伸长率、弹性模量等归入“拉伸性能”,将简支梁冲击强度、悬臂梁冲击强度、落锤式冲击强度等归入“冲击性能”,而“拉伸性能”、“冲击性能”、“弯曲性能”、“硬度”等均处于第二层,它们均包含于第

一层的“力学性能”分类中。Prop_Layer 定义的层级与 Parent_Prop_id 递归所定义的层级是冗余的，两者应一致，在不一致时以 Parent_Prop_id 递归为准，Prop_Layer 主要是控制各类别中成员的显示顺序。

- Prop_Name_cn: 属性的名称，用于检索页面属性名称树状目录的显示，并未包括牌号属性数据表中的所有属性名称，对于一些次要的、未来不用于精确检索的属性，特别是一些罕见的属性类别，可不列入其中以减少数据整合的难度，此类数据可以归入“其他”属性中，以“属性名+’: ’+属性值”的形式作为“其他”属性的属性值存储。
- Comments: 属性的含义解释以及其他备注信息。

3.1 数据优化

从原始数据资源中获取的数据往往不符合上述的这些规范，其存在的问题主要包括：

- 同一属性在不同的资源中使用不同的名称。
- 属性名称往往掺杂该属性的标准、单位等。
- 属性值数据也经常掺杂该属性的单位、测试温度等信息。
- 某些属性值数据是一个范围，如 > 5 ， $\leq 1,8 \sim 9$ 等
- 某些属性值是用科学计数法来表示的。

因此，数据优化的主要任务是解决上述这些问题，对属性名称进行合理的统一，将数据和其他描述信息分离，将数值以符合规范的形式存储，以便未来数据库系统对其内容进行精确的检索。

数据优化的过程可用 Microsoft Excel 实现。对于属性名称的统一可通过筛选、填充的方法进行修改，也可通过添加一个同义词表，通过国家材料科学数据共享平台中心基于本体(ontology)的技术进行处理。

对于复合型数据的拆分，首先可通过筛选、标记、排序将需要优化的数据分类顺序存放，然后依次分别通过查找替换、分列等方法将复合型数据进行拆分。也可依据复合型数据自身的特点，编写 VBA 程序对其中的数据信息进行自动的拆分和提取。在本课题中综合使用了这两种方法。

3.2 数据的初步评估修正和可靠性记录

多途径获得的高分子材料信息和数据资源经由数字化、目标信息提取、复杂数据分解和格式化而成为中间数据表，数据的初步评估建立在中间数据表的基础上，其目的是消除数据中的一些明显错误、粗差或系统误差，提高数据的系统性和完整性。主要方法包括数据的规范化整理、数据格式转换、粗差剔除、多源数据参考修正等。

评估修正的方法主要是以大量数据中同类型数据的筛选比较来完成的，通过筛选发现数据中偏离平均值较多的数据，然后分别在原始数据资源中和其他备查同类数据资源中查询此数据，分析其可靠性。如可靠性存疑则应对其数据进行重新测定，否则应对其进行标记，对于可靠性较差的数据应从数据库中将其剔除。

数据经过初步评估后，按照前文所述材料信息的立体描述方法在数据库中对数据进行质量级别标识，依据数据的生产过程和评估结果

将数据分别标记为以下几类：1. 经过简单处理的原始数据；2. 经标准规范化处理，已进行检查与订正的数据；3. 有数据质量标准，并经规范的质量检验与修正的数据；4. 为了特殊的用途而专门为之整理、加工和生产的科学数据产品。数据质量级别标识记录在牌号属性值表的 **AccuracyStar** 字段。

此外，在数据库设计时我们还创建了记录数据来源的数据表 (**T_RefSource**)，这样可在牌号属性值表中记录属性数据的同时，在 **RefSource_id** 字段和 **RefSource_detail** 字段分别记录入库数据的来源信息和该源数据中的详细参考信息，便于数据的进一步核对与评估。

3.3 图片、曲线、表格等特殊数据的记录规范

有机高分子材料科学领域的的数据多包含分子结构图、晶体结构、各种测试曲线以及表格等非纯文本的数据，对于这些数据的处理原则是尽量用纯文本数据对其进行描述，如果不能用纯文本数据描述的则以其他文件的形式保存，并在数据库的属性值处记录该文件的相对路径。下面就几种常见数据类型的记录方式进行规范。

3.4 图片文件的处理和记录规范

本部分参考材料共享网数据提交规范：

- 格式：相关图片或复杂图形应采用常见图片格式(如 **jpg**、**bmp**、**gif**、**png**)来进行提交和存储，虽然对于内容足够坚定或动态变化的图片可采用 **gif** 格式，但对于大多数科学数据来说，其图片往往特别注重细节的清晰显示，所以应尽量采用 **jpg** 格

式来提交绝大多数的材料相关图片。

- 压缩：对原始图片的压缩，建议利用 Photoshop 图片处理程序，选择点击“文件”菜单下的“存储为 Web 和设备所用格式”，可以在同等质量下获得更小的图片体积，从而有效节省存储空间、提高页面响应速度。
- 文件大小：如无高度清晰显示的需求，应在不影响图片质量的情况下对图片进行压缩，图片的大小规格应尽量控制在 1MB 以内。
- 命名：图片的命名应与关联表格文件中相应列的内容一致，例如有机高分子材料分类表中橡胶材料所对应的分子式一列数据内容“rubber_format.jpg”与该图片的命名应该一致。如果一个表格含有大量的数据，文件名中可使用该条记录所具有的唯一 id 的编号和列名称，比如表格 T_PropertyValue 中 Grade_id=1234 的 Vaule(text) 一列内容为一张图片，则其名称可命名为“1234_text.jpg”。
- 存储路径：图片存储的路径应与关联表格的名称相对应，如表格 T_PropertyValue 中引用的图片文件应存储在名为“T_PropertyValue”的文件夹中。

3.4.1 曲线的处理和记录规范

本部分描述了对红外光谱、核磁共振谱、DSC 谱等曲线数据的记录方式。

- 曲线可用一系列的点的坐标来描述，但是科学数据相关的曲线

往往具有几百乃至几万个数据点，而数据库中字段的长度并不是无限的，且若设置过于长的字符串字段将浪费大量的数据库空间，因此本数据库中曲线均以类似上述图片记录规范以外部文件的形式存储。

- 格式：记录曲线数据的文件为文本格式，后缀名为“.csv”，此类文件由多行数据组成，每行的末尾为回车换行符；每一行中多个数据用半角的逗号(“,”)分隔。此格式的文件可用文本编辑器打开和修改，也可以用 Microsoft Excel 直接将表格数据存储为“CSV(逗号分隔)”文件。
- 数据文件的内容：第一行记录数据的维度，二维坐标记录为“2”，三维坐标记录为“3”，超出维度的数据将不被读取。第二行记录曲线的类型是开放的还是闭合的，开放的记录为“open”，闭合的记录为“loop”。第三行记录文件的备注信息，可为空白。从第四行起是数据点的数据，每一行为一个数据点坐标，坐标的维度之间用半角的逗号分隔。例如，一个三角形曲线的描述文件可以记录为：

```
2
loop
本例子描述一个三角形
0,0
1,0
1,1
```

3.4.2 表格数据的记录规范

本部分描述了对简单表格数据的两种记录方式。

- 拆分为多条数据记录：对于原始数据资源中存在的表格，如果其内容非常重要且对于大量存在的材料类型(或牌号)具有类似的条目，即其属性可以归纳为少数几个，则应以每个数据一条记录的形式对表格内容完全拆分提取后，存储于牌号属性数据表中。
- 以单条无格式文本存储表格数据：如果数据内容较为复杂，或者对其中的数据值没有做精确检索的需求，则可将表格整体作为一个值存储于一条记录中，因此需将表格转换为无格式的文本。例如有机涂料的配方数据即可以此方法处理，在用户检索时可对配方组成进行模糊检索，而不能检索其配方中某一组分的精确值(也无此必要)。
- 无格式文本存储表格的具体方式：原始表格应为完全分隔的表格，其中无合并的单元格。在数据库中以 **html** 代码文本的形式存储其内容，即用 `<table></table>` 标记表格的开始和结束，用 `<tr></tr>` 标记每一行的开始和结束，用 `<td></td>` 标记单元格。例如，对于如下的表格：

1,1	1,2	1,3
2,1	2,2	2,3

可用以下 **html** 文本存储，其中内容可不用换行，此处加了换行符仅为方便阅读。

```
<table>
<tr><td>1,1</td><td>1,2</td><td>1,3</td></tr>
<tr><td>2,1</td><td>2,2</td><td>2,3</td></tr>
</table>
```

- 自动化处理：以下代码是将表格转换为 html 文本的 word 宏，在 Microsoft Word 中启用宏后，将代码输入 VBA 模块中。运行这个模块将替换当前 Word 文件中的所有表格为转换后的 html 文本。

```
Sub 将所有表格转换为 html()
Dim T As Table
Dim i, j As Integer
Dim txt As String
For Each T In ActiveDocument.Tables '遍历当前文件的所有表格
txt = "<table><tr><td>"
For i = 0 To T.Rows.Count - 1
For j = 0 To T.Columns.Count - 1
txt = txt & Left(T.Cell(i + 1, j + 1).Range.Text, Len(T.Cell(i + 1, j + 1).Range.Text) - 2) & "</td><td>"
Next
txt = Left(txt, Len(txt) - 4) & "</tr><tr><td>"
Next
txt = Chr(13) & Chr(10) & Left(txt, Len(txt) - 8) & "</table>" & Chr(13) & Chr(10)
T.Cell(1, 1).Select
T.Delete
Selection.InsertBefore Text:=txt
Next
End Sub
```

3.4.3 超链接的记录规范

- 在数据库字段(列)中已定义的超链接可直接以网络地址的形式存储，本节讨论的是在其他文本数据中包含的超链接。处理方式类似前述表格，即以 html 代码对其进行标记。
- 格式：以 分别作为超链接文本的开始和结束标

[高分子形态与加工课题组主页](http://pmp.iccas.ac.cn)

3.4.4 上下标及特殊字符的记录规范

此处现行《材料共享网数据提交规范》存在一个小矛盾，即依照现行规范， a^e 用 `a⊇` 表示， a_e 用 `a⊆` 表示，但是 `⊇` 和 `⊆` 同时是另外两个字符 \supseteq 和 \subseteq 的代码，目前存在冲突。在更新的提交规范制定之前，上下标暂时使用 `html` 标记进行描述，而特殊字符按照现行《材料共享网数据提交规范》规定的代码输入，或者使用 Unicode 字符直接输入。

- 上标用 `^N` 进行规范，其中的 N 为上标字符。例如 mm^2 可用 `mm²` 表示。
- 下标用代码模式 `_M` 进行规范，其中的 M 为下标字符。例如在文本中使用 `σ_{a}` 则页面将以 σ_a 显示出来。
- 摄氏度的符号使用 Unicode 的 $^{\circ}\text{C}$ (字符代码 2103)，华氏度的符号使用 Unicode 的 $^{\circ}\text{F}$ (字符代码 2109)，不使用角度 ($^{\circ}$) 与 CF 组合出的字符。
- 不使用“Symbol”“Webdings”“Wingdings”等特殊字体定义的符号，因为数据库中无法使用这些特殊字体，入库时会被错误地读取为其他字符。