

数据加工中对“忠于原文”原则的正确运用

专利检索咨询中心 张雨春



摘要: 本文结合一些原文中存在问题的实例,对非专利文献数据加工工作中如何正确运用“忠于原文”的基本原则进行了讨论。认为对于主题和整体技术方案的判断应严格忠于原文,而对于原文中明显的错误、矛盾和不明确的技术点宜进行自行判断查证后进行修正。

关键词: 非专利文献加工 忠于原文

“忠于原文”是文献数据加工的基本原则。数据加工人员应根据原文给出的具体信息,对文章中的同义词、方剂信息、化合物结构、化合物职能符(职能符的说明见文后附录)等信息进行判断和标引。随着数据加工工作和内部质量控制工作的进行,笔者

发现,有时对于“忠于原文”这一基本原则需要灵活对待。本文将结合实例对此进行讨论。

例 1. 《软珊瑚 Sarcophyton crassocaule 中 Cembrane 型二萜的分离与鉴定》(《高等学校化学学报》2003 年第 24 卷第 6 期 1023-1025 页,

徐效华等)

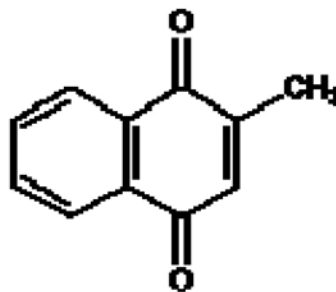
原文开头写道“我们对采自中国南海西沙群岛的软珊瑚 *Sarcophyton crassocaule* 的化学成分进行了生物活性跟踪研究, 从对癌细胞 P388 具有明显抑制活性的 CHC13 提取物中分离得到了 1 个已知结构和 2 个新西松烷型的大环二萜化合物。前文报道了其中一个结构新颖的过氧西松烷型的大环二萜 *Sarcocrassolide A*^[6] (1) 的分离和结构鉴定, 本文报道新化合物 *Sarcocrassolide B* 和 2 个已知结构的新西松烷型的大环二萜的分离和鉴定...”。之后在文中“1.2 提取与分离”部分对最终产物的叙述则为: “得到体积比为 10 : 2 的 *Sarcocrasslide A* (35 mg) 和 *B* (23 mg)。”

从上文可以看出, 文章前后对产物的描述并不一致。根据文章开头部分的描述产物应有 3 个, 即“新化合物 *Sarcocrassolide B*”和另外 2 个“已知结构的新西松烷型的大环二萜”, 但在其后的技术描述部分只得到了 2 个化合物即“*Sarcocrasslide A* 和 *B*”。而仅根据“2 个已知结构的新西松烷型的大环二萜”无法判断这两个化合物的具体信息, 故在标引时应按照技术部分的描述提交化合物 *Sarcocrasslide A* 和 *Sarcocrasslide B*。同时根据原文的描述, 两种化合物对癌细胞 P388

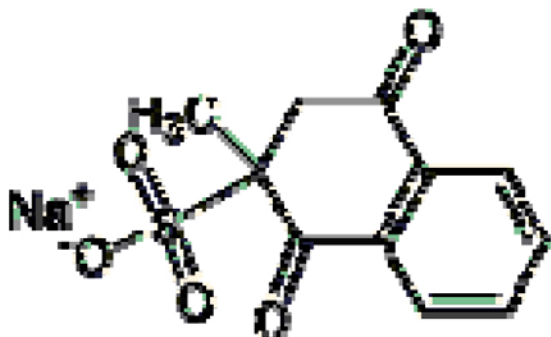
都有细胞毒性, 因此可以认为都具有“抗癌”的生物活性, 应给予职能符“T”; *Sarcocrasslide B* 依原文所述为“新化合物”, 还应给予职能符“N”; *Sarcocrasslide A* 则不是新化合物, 尽管原文对其描述有“结构新颖”一词, 但亦有“前文报道”一词, 而且在引用文献“[6]”中报道过, 因此不能认为它是新化合物。故 *Sarcocrassolide A* 的职能符应对应“K; P; T”, 即“对已知有活性的化合物的制备”。

例 2. 《钙离子对维生素 K3 镇痛作用的影响》(苏州医学院学报 1994 年第 14 卷第 1 期 8-9 页, 袁苏徐等)

该文研究了钙离子对维生素 K3 镇痛作用的影响, 研究结果表明氯化钙能拮抗维生素 K3 的镇痛作用, 而钙离子络合剂 EGTA 和钙离子通道阻滞剂维拉帕米可增强维生素 K3 的镇痛作用。依照标引规则应提交维生素 K3 的结构, 经查, 在 CA 中 Vitamin K3 对应的 CAS 号为 58-27-5, 其结构名称为 2-Methyl-1,4-naphthoquinone (2-甲基-1,4-萘醌), 即如下图所示结构:



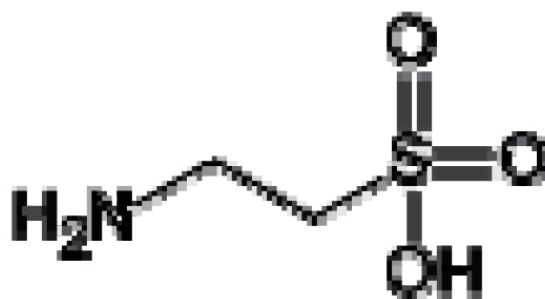
但在原文中的“材料与方法”部分却明确写道“亚硫酸氢钠甲萘醌 (VitK₃)”，即结构应如下图所示：



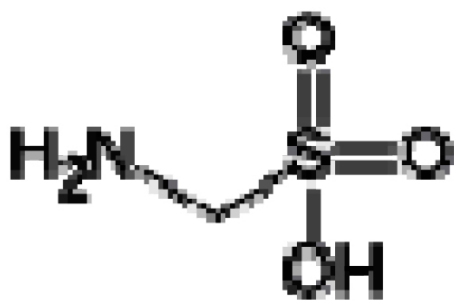
起初有标引员认为这是原文的笔误，应予修改。但是经查证，维生素 K₃ 的结构文献中有两个，即甲萘醌和亚硫酸氢钠甲萘醌。具体原因不详。部分专利文献中亦认为“亚硫酸氢钠甲萘醌”为维生素 K₃，比如《一步法制取维生素 K₃》(专利号：200710012827.4)，其制备工艺中写道：“将氧化工序得到的中间品 β 甲萘醌及焦亚硫酸钠、乙醇水溶液、催化剂按一定比例混合反应”，其中所用到的中间品“β 甲萘醌”就是 2-甲基-1,4-萘醌的简称，即 CA 中的维生素 K₃，而与亚硫酸钠反应后的产物无疑就是“亚硫酸氢钠甲萘醌”。因此尽管 CA 将 β 甲萘醌当作维生素 K₃，但是不能由此判定“亚硫酸氢钠甲萘醌 (VitK₃)”这种描述为原文的明显错误，故此文还是应遵循“忠于原文”的原则以“亚硫酸氢钠甲萘醌”的结构作为维生素 K₃ 进行标引。

例 3. 《中和滴定法测定牛磺酸含量》(《数理医药学杂志》2000 年第 13 卷第 4 期 350 页，卢忠等)

本文用酸碱中和滴定法测定了具有多种生物活性的氨基酸—牛磺酸的含量。依标引规则应提交牛磺酸的结构，但原文写道：“牛磺酸的化学名为氨基乙磺酸，其化学结构式为 NH₂CH₂SO₃H，是一种有机弱酸。”在提交了同义词“牛磺酸；氨基乙磺酸”后，发现原文提供的“化学结构式”信息与系统名不符，经 CA 查证“牛磺酸”的结构应为：



即两者相差一个亚甲基，该案例与例 2 所述的维生素 K₃ 的情况不同。首先是原文自我前后矛盾，其次是经查证市面上并没有两种同名异构的“牛磺酸”存在，且“氨基乙磺酸”名称中的“乙”表明分子中有两个碳。因此应提交正确的牛磺酸结构，而不能根据原文提供的化学结构式信息“NH₂CH₂SO₃H”将牛磺酸的结构绘制为：



例 4. 《盐酸洛美利嗪胶囊含量测定》(《中国药师》2006 年第 9 卷第 2 期 119-120 页, 唐晓霞)

原文的首句为“洛美利嗪 (Lomerizine hydrochloride) 主要通过钙通道阻滞而发挥作用, 本品能显著改善脑血管收缩引起的头痛。”依照标引规则应对文中出现的具有活性的主题化合物的同义词进行标引, 即“洛美利嗪; Lomerizine hydrochloride”。但是“hydrochloride”一词的意思为“氢氯化物”即“盐酸盐”, 因此“Lomerizine hydrochloride”中文应为“盐酸洛美利嗪”, 属于原文的明显错误, 故准确的标引方式应为“盐酸洛美利嗪; Lomerizine hydrochloride”或“洛美利嗪; Lomerizine”。而根据原文的题目和文章结尾的讨论, 笔者更倾向于前者。

例 5. 《亚胺培南治疗老年肺部感染 38 例临床分析》(《西北药学杂志》2002 年 12 月第 17 卷第 6 期 269-270 页, 白加宁等)

原文摘要的描述为“对 38 例经其它抗生素治疗无效的老年肺部感染患

者, 根据病情, 用亚胺培南 (默沙东生产的泰能) 500mg, tid, 或 500mg, qid 进行治疗, 疗程 5 ~ 10 d。”如果仅根据该描述应标引同义词“亚胺培南; 泰能”, 但是在原文最后的讨论部分却写道“泰能含有亚胺培南和西司他丁钠 (1 : 1)”。后经查询得知, “泰能”为复方, 其组成为“亚胺培南 500mg 和西司他丁钠 500mg。辅料为碳酸氢钠”, 因此不能将亚胺培南和泰能等同为同义词。笔者认为对于这种原文前后描述明显不一的情况, 数据加工人员可利用手头现有资源进行查证, 从而得出正确结论, 以正确的形式标引。而且按标引规则, 应将复方药物泰能由亚胺培南和西司他丁钠组成的信息标引入方剂信息, 同时亚胺培南对应的职能符应为“K; T; M”。

此外还有文章中的错别字或通假字。比如部分时间较早的文献中会出现“生物硷”、“葡萄糖甙”等, 其中“硷”字现已改用为“碱”字, 而“甙”则改称“苷”了。另外涉及中药的文献经常对部分中药的名称叙述有误, 比如“桅子”、“黄苓”、“瓜蒌”这三味药, 笔者特意查阅了《中草药与民族药药材图谱》、《实用中草药图典》等相关书籍, 并未发现其中有这三味药的记载, 从字体结构和五笔输入法的字码相似度来判断, 其分别应为“梔子”、“黄芩”、“瓜蒌”的误写, 但是目前

这种错误并非存在于少数文献中，笔者认为对于这种常见的错别字/通假字宜在数据库里生成最终数据时进行直接替换或建立默认同义词，以便在日后的数据检索过程中进行直接关联，从而提高检索的全面性和准确性。

综上所述，由于非专利文献在写作和出版等环节的各种原因，在原文中不时会出现一些明显的错误或技术信息不明确等情况。笔者认为“忠于

原文”是对于文章中所包含的技术信息和作者意图的真实再现，而不是“拘泥于原文”或“愚忠于原文”。因此对于这种情况，数据加工人员宜利用手头现有资源，比如联系上下文、查询网络、查询引用文献、结合自己的专业知识或公知常识进行判断，如果确定了原文中具有“确凿的错误”，则不应再继续“忠于原文”，而是应对其进行修正，以保证数据的质量。

附：化合物的职能符 (Role) 及其说明

Role	说明	Role	说明
A	被分析或测定的物质	Q	由起始原料定义的产物
C	催化剂	R	去除剂或纯化剂
D	检测剂	S	中间体或起始物
E	赋形剂	T	活性成分或治疗剂或前药
K	已知化合物	U	新治疗用途
M	联合药物	V	反应物
N	新化合物	X	去除的物质
P	已知化合物的制备	—	—

(专利检索咨询中心 杜军 杨晓春 审校)

