

用 STN 进行高分子通式化合物的结构式检索浅谈

专利审查协作北京中心 李曦 高志纯 刘央央 金柳欣 郑凯 杨芳 李慧¹

摘要：高分子化合物由于其对应的分子结构相对复杂，表征其基本结构和 / 或组成特征的要素繁多，使得在常规的专利库或非专利库进行高分子通式化合物检索的结果并不理想。通过对实际案例的检索和分析，总结了如何制定适当的检索策略，从而运用 STN 数据库准确、高效地检索高分子通式化合物。

关键词：高分子 通式化合物 STN 结构式 检索



李曦：2010 年入局，专利审查协作北京中心化学部，助理研究员，高分子领域。



高志纯：2001 年入局，专利审查协作北京中心化学部，研究员，高分子领域。



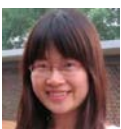
刘央央：2010 年入局，专利审查协作北京中心化学部，助理研究员，高分子领域。



金柳欣：2010 年入局，专利审查协作北京中心化学部，助理研究员，高分子领域。



郑凯：2010 年入局，专利审查协作北京中心化学部，助理研究员，高分子领域。



杨芳：2011 年入局，专利审查协作北京中心化学部，助理研究员，高分子领域。



李慧：2010 年入局，专利审查协作北京中心材料部，助理研究员，陶瓷领域。

一、用 STN 结构式检索高分子通式化合物的动机

高分子化合物（聚合物）是通过一种或多种分子或分子团以共价键结合而成的具有多个重复单元的大分子，其对应的分子结构相对复杂，表征高分子化合物基本结构和 / 或组成的特征有重复单元的名称或结构式、重复单元的排列（如无规共聚、嵌段共聚、交替共聚等）等要素^[1]。在 高分子化合物领域，高分子通式化合物的命名标准并不统一，有的是申请人自定义的命名，或是根本就没有给出具体的命名，而且合成这些高分子通式化合物的原

¹ 等同第一作者

料也是所属领域的一些常规的单体，因此，想要提取这些高分子化合物的名称作为关键词进行检索时往往会遇到难以表达的困难，由于无法准确地用关键词表达重复单元与重复单元之间以及一个重复单元内的结构片段之间的排列方式、单体反应基团的多样性等原因，想要用这些高分子化合物的结构片段或合成他们的原料名称作为关键词在专利库或非专利库进行检索，其结果也未必理想。总的来说，对高分子化合物的检索具体可能会出现以下几种情况：

1. 使用申请人自己定义的名称作为关键词进行检索，基于在所属技术领域使用上述名称表达同一高分子化合物的可能只有申请人自己的局限性，有时只能检索到本申请；

2. 如果只是将高分子化合物的常规的结构单元片段、反应单体或其片段作为检索要素检索，可能会导致噪音过大、检索结果的数量过多以致难以筛选的一个极端；

3. 如果将高分子化合物的结构单元的整体、包含特殊官能团的反应单体作为检索要素检索，可能会造成检索结果的范围过窄，漏掉了有效的对比文件的极端，这可能是由于现有技术中并没有对整个结构单元进行标引或仅使用了与本申请具有不同的官能团结构但具有相同的作用的可替换单体。

针对上述几种情况，可以借助 STN 国际联机系统进行结构式检索。

STN 国际联机系统是世界上最大的

科学技术信息联机检索系统^[2]，其英文全称为 The Scientific & Technical Information Network International，亦称之为“国际科学技术信息网络系统”。它始建于 1983 年，当时由德国卡尔斯鲁厄能源、物理、数学专业信息中心计算机检索系统与美国化学协会的美国化学文摘社（CAS）的计算机联机检索系统互联，并开通联机检索服务；1986 年又与日本科学技术情报信息中心联合形成了目前的国际性联机信息网络检索系统。该系统将位于美国、德国、日本三地的主机系统的数据库进行互联，是最先采用广域网互联方式的网络系统^[3]。STN 国际联机系统包含有 200 多个科技文献数据库，对科学技术文献具有更先进的检索功能，并拥有一些其他任何联机系统中所没有的特殊宝贵信息资源和检索手段，如物质数据库 REGISTRY 和 MARPAT、文摘数据库 CAPLUS。STN 是世界上第一个实现图形检索的系统，特别是化学物质和结构检索功能，是任何系统也无法比拟的。利用 STN 进行检索时，可以用化学结构图形或分子式或化学物质名称等在世界上最庞大，收录数据最完全的化学物质名称数据库 REGISTRY 数据库中准确地找到想要寻找的化学物质，并将检索得到的物质转入 CAPLUS 数据库中作进一步的文献检索。

二、在 REGISTRY 和 CAPLUS 中进行高分子化合物结构式检索的基本步骤

如何利用已知的结构式在 STN 中进行高分子通式化合物的结构式检索呢？一般来说，在 STN 中对聚合物的检索类型分为一个单体的均聚物 EXACT search，一个确定单体的所有共聚物 FAMILY search，一个 substructure 的所有均聚物 / 一个 substructure 的所有共聚物 FAMILY search，STN 中的聚合物结构检索通常分为以下步骤：

- Step 1. 画结构式
- Step 2. 上传结构到 REGISTRY
- Step 3. 用 AND 运算符联合 screen 和结构
- Step 4. 使用 EXACT/ FAMILY/ SUBSTRUCTURE search
- Step 5. 将检索得到的结果转入 CAPLUS 数据库中作进一步的文献检索。

其中 screen 相当于结构过滤器，其由一串数字构成，不同的数字串代表不同的结构检索类型，它可以在结构检索时用于进一步优化检索结果。对于聚合物的检索，screen 都是手工添加的，STN 会自动产生一些普通的 screen，聚合物检索时常用的 screen 有：

- 2043 聚合物（普通分类）
- 2067 均聚物和共聚物
- 2068 用结构重复单元定义的聚合物
- 2069 带有端基的重复单元结构
- 2070 无端基的重复单元结构
- 多组分物质 Screens
- 2127 2 或更多组分
- 2077 3 或更多组分
- 2078 4 或更多组分

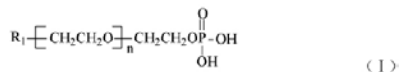
合理地选择结构式在 STN 中进行检索是高分子通式化合物结构式检索的关键，这是因为同一高分子通式化合物的结构式的画法并不唯一，而 REGISTRY 在收录高分子通式化合物时，不一定会对高分子通式化合物的结构式作归一化处理，也可能根本没有对高分子通式化合物及其对应的具体的高分子化合物进行标引，而只是标引了制备该高分子通式化合物的原料。如何在 STN 中对高分子通式化合物进行有效、精确的检索取决于是否具有合适的检索策略，从而能够快速筛选出合适的结构式以提供检索。下面通过以下的检索实例来说明在 STN 中如何选择结构式准确、高效地检索高分子通式化合物。

三、实际检索案例

（一）【案例 1】

1. 案情介绍

案例 1 涉及一种含磷酸二氢根的聚乙二醇磷酸二氢酯及其制备方法。权利要求 1 为：一种聚乙二醇磷酸二氢酯，其结构通式为：



其中 R_1 为 $\text{CH}_3\text{CH}_2\text{O}-$ 、 $\text{HO}-$ 、 $\text{H}_2\text{N}-$ 或 $\text{HOOC}-$ 中的任一种； n 为 2 ~ 200 的正整数。

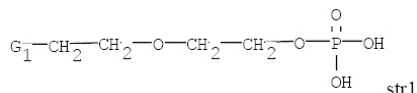
2. 在 STN 以外的专利数据库和非专利数据库中的检索

首先以聚乙二醇，磷酸酯为基本检索要素，在专利数据库和非专利数据库中进行常规的检索。由于以上述检索要素为基础扩展的关键词所表示的结构在所属领域

均是非常常见的结构，因此出现的频率较高，检索的噪音较大，难以提取合适的对比文件；而 CA 中只对本申请具体的原料、中间体和产物进行标引，通过使用这些原料、中间体和产物的 CAS 登记号和关键词并没有检索到合适的对比文件，只检到了本申请。最终在 ISI WEB of KNOWLEDGE 中以发明人和主题为入口只检索到一篇可以评价权利要求 1 结构通式 (I) 所示的化合物（其中 R_1 为 $\text{CH}_3\text{CH}_2\text{O}-$ 时）的创造性的非专利文献：“聚乙二醇/羟基磷灰石纳米杂化材料的制备及表征”，张鹏等，《高等学校化学学报》，第 33 卷，第 1 期，22-25 页，2012 年 1 月 31 日。

3. 在 STN 中的初步检索

继而在 STN 中进行结构式检索。按照权利要求 1 的公式 (I) 所示的化合物绘制检索结构式 str1，所述检索结构式 str1 如下所示：



其中， G_1 的定义为 H_0 、 H_2N 或 H_0OC 。由于公式 (I) 所示的聚乙二醇磷酸二氢酯是通过单端羟基聚乙二醇与磷酰化试剂反应得到的，而单端羟基聚乙二醇中的氧化乙烯基的重复单元数 ≥ 2 ，因此将 str1 中氧化乙烯的重复单元数绘制为 2。使用 str1 检索的步骤如下：

=> FILE REG

=> L1 Uploading...¥STN¥1.str(将 str1 上传到 REGISTRY 中)

=> L2 SCREEN 2043

=>L3 S L1 AND L2 SSS FUL (用 AND 运算符联合 screen2043 (聚合物普通

分类) 和结构)

=> FILE CAPLUS

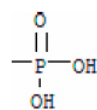
=>L4 S L3(将在 REGISTRY 中检索得到的结果转入 CAPLUS 数据库中作进一步的文献检索)

以上检索得到 5 篇文件，对检索结果进行分析可知，检索得到能够评价权利要求 1 (当 R_1 为 $\text{H}_2\text{N}-$ 时) 创造性的 X 文件 JP 昭 62-7749A，还检到了本申请人在同一申请日提交的另一篇系列申请。

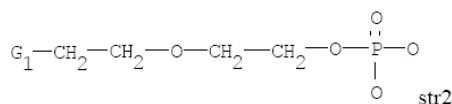
由于考虑到 STN 可能并未将对比文件中与本申请的聚乙二醇磷酸二氢酯具有相同结构的化合物标记为聚合物，因此，调整检索思路，将结构式 str1 上传到 REGISTRY，然后直接采用 SSS 对其结果进行检索，检索得到 21 篇文献，未再检得到更多的能够评价权利要求 1 新颖性和/或创造性的文件。

4. 调整 STN 中的检索结构式

进一步分析可以看出，构造的检索结构式 str1 中具有磷酸根结构



和羧基 $-\text{COOH}$ ，重新构造如下式所示的检索结构式 str2：



其中， G_1 的定义为 H_0 、 H_2N 或 H_0OC 。再将结构式 str2 上传到 REGISTRY，然后采用 SSS 对其结果进行检索，最后将检索得到的结果转入 CAPLUS 数据库中作进一步的文献检索得到 39 篇文件。其中又有

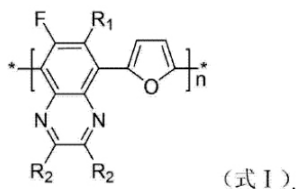
4 篇文件可以评述权利要求 1 结构 (I), 当 R_1 为 $HO-$ 、 H_2N- 或 $HOO C-$ 时的新颖性和 / 或创造性: 其中 3 篇为专利文献: JP 昭 48-34832B、JP 特 开 2001-3083A、US2002/0142329A1, 1 篇为非专利文献: “Phosphopantetheinyl Transferase Catalyzed Site-Specific Protein Labeling with ADP Conjugated Chemical Probes”, Yekui Zou, et al, 《J. AM. CHEM. SOC.》, 第 131 期, 7548 - 7549 页, 公开日 2009 年 5 月 14 日。

用以上两个检索结构式 str1 和 str2 检索的结果存在差异是因为聚乙二醇磷酸二氢酯的末端基团磷酸基和羧基结构中的氢很容易被取代, 它们既能以酸的形式存在, 也能以盐或酯的形式存在, 而这几种形式的共性是均可以标记为磷酸根或羧酸根。

(二)【案例 2】

1. 案情介绍

案例 2 涉及一种氟代喹啉与呋喃的共轭聚合物。权利要求 1 为: 一种氟代喹啉与呋喃的共轭聚合物, 其特征在于, 结构通式如式 I 所示:



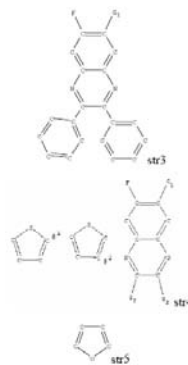
式中, R_1 为 H 或 F 原子; R_2 是间位或对位碳原子数为 4 到 20 的直链或支链的烷氧基苯基, 或者是碳原子数为 4 到 20 的直链或支链的 2- 烷基噻吩基或 2, 3- 二烷基噻吩基。

2. 在 STN 以外的专利数据库和非专利数据库中的检索

首先以氟代喹啉与呋喃为基本检索要素, 以关键词和 / 或分类号为入口在专利数据库中进行常规的检索, 并未检索到合适的对比文件; 然后再在非专利数据库中进行检索, CA 中只给出了本申请中产物的 CAS 登记号, 根据这些登记号只检索到了本申请本身, 而用从氟代喹啉和呋喃中提取的关键词在 CA 中检索的噪音较大, 难以提取合适的对比文件; 鉴于本申请的权利要求 1 请求保护以结构为特征的聚合物, 在 STN 中进行结构式检索。

3. 在 STN 中检索 X 文件

第一步进行评价新颖性或单篇评价创造性的 X 文件的检索, 按照权利要求 1 式 I 所示的聚合物绘制如下所示的检索结构式 str3、str4 和 str5:



其中, G_1 的定义为 H 或 F, G_2 定义为 str4 中左边的两个噻吩片断, “1”和“2”分别代表这两个噻吩基团与喹啉基团的连接点。由于如式 I 所示的结构通式中 R_2 是间位或对位碳原子数为 4 到 20 的直链或支链的烷氧基苯基, 或者是碳原子数为 4 到 20 的直链或支链的 2- 烷基噻吩基。

吩基或 2,3-二烷基噻吩基,考虑到在所属技术领域,烷氧基苯基和烷基噻吩基中的烷氧基或烷基存在与否、或是以其他基团如烷基的形式存在、以及这些基团的取代位置对所述共轭聚合物的性能的影响是可以预期的,从而绘制如上所示的检索结构式 str3 和 str4。检索过程如下:

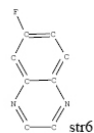
=> FILE REG
=> L1 Uploading...STN¥3. str(将 str3 上传到 REGISTRY 中)
=> L2 Uploading...STN¥4. str(将 str4 上传到 REGISTRY 中)
=> L3 Uploading...STN¥5. str(将 str5 上传到 REGISTRY 中)
=> L4 S (L1 OR L2) AND L3 SSS FUL (用 OR 运算符联合结构 str3 和 str4 后再用 AND 运算符联合结构 str5)
=> FILE CAPLUS
=> L5 S L4(将在 REGISTRY 中检索得到的结果转入 CAPLUS 数据库中作进一步的文献检索)

以上检索得到 1 篇文件,为本申请。

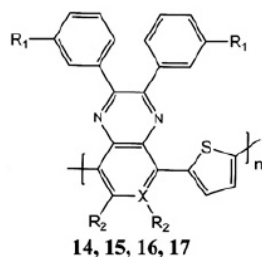
4. 在 STN 中检索 Y 文件

由于没有检到合适的对比文件,因此改变检索策略。经过以上对现有技术的检索以及对本申请的进一步的解读发现,本申请的发明点就在于在喹喔啉这种缺电子单元上引入氟原子,然后与富电单元聚合制备成含氟 D-A-D 型给体聚合物,从而降低材料的 HOMO 能级,提升开路电压,提高器件光伏性能,因此,继而检索能够评价权利要求 1 创造性的 Y 文件,即检索“包含氟代喹喔啉单元的聚合物”。考虑到对

比文件中可能并没有在结构中标引氟代喹喔啉基团中的其他取代基,绘制如下所示的检索结构式 str6:



再将结构式 str6 上传到 REGISTRY,用 AND 运算符联合 screen2043 和结构,然后采用 SSS 检索,最后将检索得到的结果转入 CAPLUS 数据库中作进一步的文献检索得到 59 篇文件,浏览检索得到的文件,发现其中 1 篇为 Y 文件“New quinoxaline and pyridopyrazine-based polymers for solution-processable photovoltaics”,ReneeKroon, et al,《Solar Energy Materials & Solar Cells》,第 105 卷,第 280-286 页,公开日 2012 年 7 月 13 日。该文件中公开了用于溶液法制备太阳能电池的新型含吡啶并吡嗪聚合物:



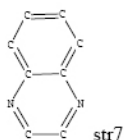
其中,

TQ1 (14): R₁:OC₈H₁₇, R₂: H, X: C TQ-8A (16): R₁:C₈H₁₇, R₂: H, X: C
TQ-N (15): R₁:OC₈H₁₇, R₂: H, X: N TQ-F (17): R₁:OC₈H₁₇, R₂: F, X: C,

该文件的表 3 和表 4 还给出 TQ-F (含氟代喹喔啉单元)依次比 TQ1 和 TQ-8A (均不含氟代喹喔啉单元)具有更低的 HOMO 能和更高的开路电压,即由 TQ-F 依次比由 TQ1 和 TQ-8A 为给体材料的光伏器件的光

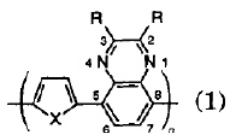
伏性能更好。

接下来需要检索另一篇 Y 文件,即包含喹喔啉单元和呋喃单元的聚合物。绘制如下所示的检索结构式 str7:



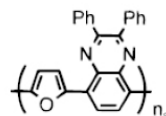
再将结构式 str7 和上述结构式 str5 上传到 REGISTRY, 用 AND 运算符联合 screen2043 和结构, 然后采用 SSS 检索, 最后将检索得到的结果转入 CAPLUS 数据库中作进一步的文献检索得到 16 篇文件, 浏览检索得到的文件, 发现 2 篇 Y 文件:

“New π -Conjugated Heteroaromatic Alternative Copolymers with Electron-Donating Thiophene or Furan Units and Electron-Withdrawing Quinoxaline Units: Preparation by Palladium-Catalyzed Polycondensation and Characterization of the Copolymers”, TAKAKI KANBARA, et al, 《Journal of Polymer Science: Part A: Polymer Chemistry》, 第 33 卷, 第 6 期, 第 99-1003 页, 公开日 1995 年 4 月 30 日, 该文件中公开了如下式 (1) 所示的聚合物:



其中 $X=O$, $R=Ph$; “ π -Conjugated Donor-Acceptor Copolymers Constituted of π -Excessive and π -Deficient Arylene

Units. Optical and Electrochemical Properties in Relation to CT Structure of the Polymer”, Yamamoto et al., 《J. Am. Chem. Soc. Vol.》, 第 118 卷, 第 43 期, 1996, 这 2 篇文件中均公开了包含喹喔啉单元和呋喃单元的聚合物, 该文件中公开了具有如下结构式所示的聚合物:



四、检索策略总结

对于难以用关键词和分类号表达或准确检索的高分子通式化合物的权利要求, 可以选择在 STN 中进行结构式检索。因为当所述高分子通式化合物在所属领域不具有通用的命名时, 对该高分子通式化合物的命名一般包含以下两种情况: 1. 只给出了申请人自定义的命名, 如案例 1 中的含磷酸二氢根的聚乙二醇磷酸二氢酯; 2. 权利要求中没有给出高分子通式化合物命名, 只是给出了其部分结构片段的名称, 在说明书部分也只能提取该通式化合物的具体制备原料的名称, 如案例 2 中的氟代喹喔啉与呋喃的共轭聚合物。这时如果再用常规的关键词或关键词加分类号的检索手段进行检索会出现问题: 使用申请人自己定义的高分子通式化合物的名称作为关键词或 CA 给出的该高分子通式化合物的 CAS 登记号检索, 可能只是检索到本申请, 如案例 1 和案例 2 中均使用了聚乙二醇磷酸二氢酯产物和氟代喹喔啉与呋喃的共轭

聚合物产物的 CAS 号进行检索，分别只检到了本申请自己；而用这些高分子通式化合物的结构片段或原料名称作为关键词进行检索，其结果也未必理想，一个极端是可能会遇到噪音过大、检索结果的数量过多从而难以筛选的问题，如用案例 1 中在所属领域常见的聚乙二醇、磷酸酯为基础扩展的关键词在专利数据库和非专利数据库中进行常规的检索，由于这些关键词出现的频率较高，检索的噪音较大，难以提取合适的对比文件；另一个极端就是检索结果的范围过窄，漏掉了有效的对比文件，如案例 1 和案例 2 中均使用了实施例中的具体的原料的 CAS 登记号进行检索，也只能检索到本申请。

通过对以上案例的实际检索过程可以看出，在使用 STN 进行结构式检索时，检索的关键在于检索结构式的选择与构建。而在选择与构建检索结构式时需要注意以下几点：

首先需要注意 STN 对高分子通式化合物的标引形式。一般来说，STN 对文献中出现的高分子通式化合物的结构有两种标引形式：对高分子通式化合物对应的所有具体的高分子化合物的结构式进行标引，如案例 1 中的含磷酸二氢根的聚乙二醇磷酸二氢酯通式中对应的各个具体的聚乙二醇磷酸二氢酯结构均被 STN 所标引；不对高分子通式化合物对应的所有具体高分子化合物进行标引，而只是标引了部分实施例中出现的具体的高分子化合物的结构，如案例 2 中，STN 就只标引了其实施例中出现的氟代喹啉与呋喃的共轭聚合物。

鉴于以上 STN 对高分子化合物的标引形式，在选择与构建检索结构式时，如果高分子通式化合物的结构比较单一，如案例 1 那样只对应四种不同的具体化合物，可以首先对高分子通式化合物的每种形式进行检索，即将通式中给出的不同的取代基用变量 G-group 定义，从而构建成一个检索式；如果高分子通式化合物代表多种不同结构的化合物，并难以用少数几个检索结构式表达出来的时候，可以先选择实施例中的具体高分子化合物的结构构建检索结构式。

若以上两种检索方式的检索结果不理想，就需要进一步选择与构建检索结构式，也就是将最能够代表权利要求中的高分子通式化合物的那部分结构构建检索结构式，如考虑到端基官能团可能的变化，在构建检索结构式时，只选择端基官能团的一部分或不选择端基基团，如案例 1 中考虑到对比文件中的羧酸和磷酸基团中的氢均能被其他基团所取代，因此在进一步构建检索式时只选取了磷酸根离子和羧酸根离子；对于以重复单元的结构为特征的高分子通式化合物的权利要求，如果只检索高分子通式化合物中完整的重复单元结构得不到合适的对比文件（即 X 文件）时，需要进一步将重复单元中的片断拆分开来，检索对发明点有贡献的那部分片段，或根据实际的需要决定是否将拆分后的片段分别进行检索，并将得到的检索式用“and”联合，如在案例 2 中，用同时含有氟代喹啉片断和呋喃片断的重复单元构建的检索式检索只能检索到本申请本

身,进一步理解发明可以得出“含有氟代喹喔啉基团的聚合物”就是本申请的发明点,从而检索具有同样的发明点的对比文件,也就是只检索由氟代喹喔啉片断构建的检索式,最终得到了一篇合适的Y文件;

还有一点值得注意的是,是否选择使用运算符联合对聚合物检索有用的screens与上传到REGISTRY中的结构的结果视具体的情况而定,这与REGISTRY中是否将高分子通式化合物标记为“polymer”有关。

五、结语

以上是对用STN进行高分子通式化合物的结构式检索的一些浅谈。当用常规的检索手段在专利库和非专利库中难以对包含高分子通式化合物的权利要求进行检索时,利用STN结构式检索可能能够对其精确地检索,起到事半功倍的效果。但STN也有一些局限性,比如STN不可能做到把文献里出现过的所有结构式都进行标引,过分依赖STN的检索结果可能会导致漏检,且STN的结构式检索费用较高。因此,应当合理、有效地利用STN进行高分子通式化合物的结构式检索。

参考文献

- [1] 国家知识产权局专利局专利审查协作北京中心. 化学领域文献实用检索策略[M]. 北京: 知识产权出版社, 2011: 107-135
- [2] 刘延淮. 利用STN国际联机检索解决化学技术文献检索难题——记STN国际联机检索培训班[J]. 审查业务通讯, 1997, 3(1): 1-6
- [3] 王星亮. STN专利数据库中的专利文献种类代码[J]. 专利文献研究, 2003(4): 1-15

(专利审查协作北京中心 刘文霞 审校)