

# PATENTICS

## 审查员必读

微信合集



## 目录

Patentics 语义检索精解 .....	3
Patentics 相关度解析 .....	5
Patentics 搜索透镜(1) .....	9
Patentics 搜索透镜(2) .....	12
Patentics 快捷功能按钮---审查员必读 .....	18
SIPO 与 EPO 英文审查报告结果自动比对 .....	21
用大数据分析认定、辩护公知技术推断 .....	24
Patentics 语义检索过程标准化 .....	26
审查员幸运模式与 Patentics 自动测试比对 .....	33
机器翻译+机器标引+机器语义排序 = EPO 特殊的检索工具或检索方式 .....	36

# Patentics 语义检索精解

许多用户在使用 Patentics 的语义检索中，经常误解语义检索“缺少可控性”。其实不然。

Patentics 语义检索的精髓是排序。

排序可以在全部数据库中排序。比如，`r/公开号`，`r/cn1128593`，`r/us4567890`，等等；

如果是新颖性检索，则加上时间限制。如，`r/公开号 and di/公开号`，即对该公开号申请日之前公开的所有文档进行排序。同时，还可以给出年月日时间数据，指定为该日之前公开的所有文献进行排序，如 `r/公开号 and di/yyyymdd`。如果希望做抵触申请，则时间限制用 `da/公开号`，或 `da/yyyymdd`，为对该公开号申请日（或 `yyyymdd`）之前申请的所有文档排序。这就是我们所谓的“大海捞针”般的全数据库排序，即完全依靠我们的计算智能，按文本内容进行语义排序。排序结果是全部数据库，显然那些排在后面的不相干文档，没有任何用处。因此，Patentics 给出最相关的 400 个输出。而这往往给人以“漏检”的印象。其实，如果需要 `n` 个，你可以后加 `and ctop/n` 来控制输出最相关的 `n` 个文档，例如，`r/cn 1128593 and ctop/10000`，就是用 `cn1128593` 对全库排序，输出最相关的 10000 个。

很显然，如果用户想对排序结果进行控制，除了上述输出个数以外，对于排序对象也可加以限制。例如，用户认为参与排序的文档必须包括某一关键词，如 `cdma`，则可以用布尔检索式 `b/cdma` 来加以限制定义，`r/CN1128593 and di/cn1128593 and b/cdma`。

必须强调，Patentics 支持所有传统布尔检索方法的命令。用户可以用任何传统检索方法来对排序的结果集数据进行控制（限制）。这样，采用传统检索来限定一个数据集，帮助我们的计算机算法从全库范围排序缩小到一个限定集排序，相当于人的智慧与机器的人工智能结合，产生更精准的排序结果。

什么是最理想的布尔检索式来限制排序数据集？答案是一切检索策略中有效的表达式都可以用。如关键词（`b/关键词`，`t1/关键词`，`abst/关键词`，`spec/关键词`，等等）、IPC/（CPC/）分类，标准化申请人（`ann/申请人`）、申请人（`an/`

申请人)。而且这些检索要素都可以像传统检索策略一样，通过 **and**、**or**、**andnot** 来组合。

与传统检索方法相比，什么是语义排序的优势？大家知道，传统检索输出的结果，没有相关度排序，每一篇文档都是一样重要，都必须阅读，不然就是漏检。因此，用户需要设计 4-6 个检索策略，将数据集限定在可浏览的范围（100-200 篇），然后浏览每一篇文献。这样，许多相关的文献就会被 4-6 个检索策略排除在外，发生漏检。

而采用语义排序，被排序的结果集可以很宽泛。采用一个简单的关键词限定排序集在几十万篇中排序，经常也可达到理想结果。用户可以根据相关度从最相关到次相关依次浏览。往往浏览前面几篇，就可获得想要结果。而且根据系统给出的相关度排序，如果用户认为还需要进一步限定结果集，用户可以进一步添加检索式。

用户进行查新检索，可以从简单到复杂，逐步渐进。

**r/CN1128593**

**r/CN1128593 and di/cn1128593**

**r/CN1128593 and di/cn1128593 and b/cdma**，或

**r/CN1128593 and di/cn1128593 and ann/高通**，或

**r/CN1128593 and di/cn1128593 and icl/h04L**，或，

**r/CN1128593 and di/cn1128593 and icl/h04L and b/cdma** 等。

排序结果的最前面几篇是必读的。Patentics 大数据分析告诉我们，第一篇就是审查员心目中的对比文件的概率为 8-9%。而如果公开号被绿色标记，则表示该篇文献已经被语义透镜聚焦，其命中概率是 20%以上。

当然，请记住！每引入一个检索要素进行限制，就会产生漏检。而这正是传统检索技术无法避免的。Patentics 语义检索，可以将漏检限制在没有（r/）或很小的范围（r/ and b/）。

# Patentics 相关度解析

在使用 Patentics 智能搜索系统进行搜索时，系统会根据搜索表达式和专利文本含义自动算出两者之间的相关度，并以百分比表示。

当我们分别用“手机”和“汉字输入法”进行概念搜索，我们发现，与“手机”最相关专利的相关度仅为 91%，而“汉字输入法”的最相关专利的相关度却高达 100%。

公开号	标题	申请人	发明人	CPC	国际分类	相关度
CN201985937	带有接听装置的超薄卡片式移动通信终端	上海华勤通讯技术有限公司	陈文峰		H04M	92%
CN203167013	一种具无线路由功能的腕带式手机	华锐	华锐		H04M	89%
CN201332417	带打火机功能的手机	康佳集团股份有限公司	林麒伟		H04M	87%
CN2011130963	一种分体式蓝牙手机	康佳集团股份有限公司	刘红	H04M	H04M	86%
CN202004827	一种带有指南针的手机	上海闻泰电子科技有限公司	夏映微		H04M	86%

公开号	标题	申请人	发明人	CPC	国际分类	相关度
CN1207519	双笔画码汉字输入方法	张文安	张文安	G06F	G06F	99%
CN1570818	计算机汉语五码(符型)输入方法	孙国	孙国	G06F	G06F	99%
CN1848053	一种拼音分隔码输入法	周常德	周常德	G06F	G06F	99%
CN1397864	计算机音形码汉字输入法	甘锋利	甘锋利	G06F	G06F	99%
CN1096112	一种汉字声母编码输入法及其所用键盘	莫重文	莫重文	G06F	G06F	99%
CN101266519	一种汉字拼音识别码输入法	甄鲲鹏	甄鲲鹏   崔永振   甄小飞	G06F	G06F	99%
CN1338674	形音数三码合一汉字输入法系统	李一新	李一新	G06F	G06F	99%
CN1581037	计算机汉语(左右码)输入法	刘汉秉   孙国	刘汉秉   孙国	G06F	G06F	99%
CN101833382	笔画加部首输入法	张新芳	张新芳		G06F	99%

我们分别查看含有关键字“手机”和“汉字输入法”专利的数量。

公开号	标题	申请人	发明人	CPC	国际分类	相关度
CN201985937	带有接听装置的超薄卡片式移动通信终端	上海华勤通讯技术有限公司	陈文峰		H04M	92%
CN203167013	一种具无线路由功能的腕带式手机	华锐	华锐		H04M	89%
CN201332417	带打火机功能的手机	康佳集团股份有限公司	林麒伟		H04M	87%
CN2011130963	一种分体式蓝牙手机	康佳集团股份有限公司	刘红	H04M	H04M	86%
CN202004827	一种带有指南针的手机	上海闻泰电子科技有限公司	夏映微		H04M	86%

b/汉字输入法

搜索帮助 | 论坛  
搜索过滤

搜索帮助 字段组合 智能搜索向导 用户手册 可视化检索

中国申请 快速浏览全文

2250项结果: 相关性(%) 0 过滤 专利号:

公开号	标题	申请人	发明人	CPC	国际分类	相关性
CN1207519	关键字含有“中文输入法”的全部专利	张文安	张文安	G06F	G06F	99%
CN1848053	一种拼音分隔码输入法	周常德	周常德	G06F	G06F	99%
CN1397864	计算机音形码汉字输入法	甘锋利	甘锋利	G06F	G06F	99%
CN101266519	一种汉字拼音识别码输入法	甄鲲鹏	甄鲲鹏   崔永振   甄小飞	G06F	G06F	99%
CN1338674	形音数三码合一汉字输入法系统	李一新	李一新	G06F	G06F	99%
CN101833382	笔画加部首输入法	张新芳	张新芳	G06F	G06F	99%
CN101063905	音数码汉字输入法	蓝章礼	蓝章礼   曹建秋	G06F	G06F	99%
CN1858680	五行码汉字输入法	付加高	付加高   罗艳琴	G06F	G06F	99%
CN1265484	合成码汉字输入法	程红伟	程红伟	G06F	G06F	99%
CN102566777	数码汉字输入法	谭钧元	谭钧元	G06F	G06F	99%
CN1567157	五码加加汉字输入法	廖远	廖远	G06F	G06F	99%
CN1460915	汉字输入法	梅雪倩	梅雪倩	G06F	G06F	99%

含有关键字“手机”的专利有 320499 篇，“汉字输入法”有 2250 篇。

手机是现在人的生活必需品，在食品、化学、医疗、家居和生物等等领域都有所涉及，衍生出的产品也数不胜数，已不仅仅是最早纯粹的通信工具。如果需要查找手机应用于某领域的专利，我们需要阅读包含“手机”的全部八万多篇专利，犹如大海捞针。为此，用 Patentsics 重排序功能，可对这些含有关键字“手机”的专利进行重排序。

例如，“b/手机 and r/饭菜”，“b/手机 and r/DNA”。

b/手机 and r/饭菜

搜索帮助 | 论坛  
搜索过滤

搜索帮助 字段组合 智能搜索向导 用户手册 可视化检索

中国申请 快速浏览全文

320499项结果: 相关性(%) 0 过滤 专利号:

公开号	标题	申请人	发明人	CPC	国际分类	相关性
CN104887021	内置式沸水保温托盘	李良友	李良友   李水   李洲	A47G	A47G	82%
CN103637669	烹饪器具自动控制多种食物同时烹饪的方法	林志鹏	林志鹏	A47J	A47J	81%
CN203969676	智能餐饮系统	祁国祥	祁国祥	A47B	A47B	80%
CN104082975	智能餐饮系统	祁国祥	祁国祥	A47B	A47B	80%
CN104188554	带USB接口的饭盒加热装置	四川长虹电器股份有限公司	罗彬	A47J	A47J	79%
CN203168676	一种改进型饭碗	王雅慧	王雅慧	A47G	A47G	79%

用“饭菜”对含有关键词“手机”的专利排序，我们居然发现有许多引用“手机”的专利与“饭菜”最高相关度为 82%。

公开号	标题	申请人	发明人	CPC	国际分类	相关度
CN102373265	一种检测遗传性耳聋的试剂盒	博奥生物有限公司   清华大学	张冠斌   项光新   邢婉丽   程京		C12Q	86%
CN102453761	一种磁珠与发光体共标记以检测遗传性耳聋的试剂盒	博奥生物有限公司   清华大学	高华方   项光新   蒋迪   邢婉丽   程京		C12Q	85%
CN102428191	热稳定性内切核酸酶在产生报道分子中的应用	塞昆纳姆股份有限公司	M·A·罗伊   P·A·厄斯		C12Q	84%
CN103882131	确定样品中是否存在目标物的方法	清华大学	杜亚楠   谢丽萍		C12Q	82%
CN103890191	单细胞全基因组扩增方法	哈佛大学校长及研究员协会	谢晓亮   宗诚航   陆思嘉		C12Q	81%

用“DNA”对含有关键词“手机”的专利排序，引用“手机”的专利与“DNA”最高相关度为 86%。

查看排在最前面结果，发现不少内容主题的确与“饭菜”和“DNA”相关的专利，同时也引用了关键词“手机”。说明了“手机”主题的发散、不明确性。

相关度的计算，可以理解为 Patentics 自动将含有关键字“手机”的专利按主题排序，比如说 Patentics 发现在全部中国专利申请（目前为止 960 万）中引用关键词“手机”的多数申请都是与通信相关的。因此当用户仅输入“手机”搜索时，系统自动将最可能（92%）与通信相关的申请排在最前面。打个比方说，快递以前只能靠陆运，随着发展，现在已能够海运、陆运、空运和河运，但绝大多数还是靠陆运。所以提起运输，我们首先想到的还是陆运。但陆运对运输来说，已不是百分之百的选择了。系统将其它主题的申请排在后面。例如，那篇与“饭菜”有关并引用关键词“手机”的申请 CN104887021 “内置式沸水保温托盘”，被系统以相关度 8%排在第 258160 位。所以当用户输入关键字检索“b/手机”，Patentics 已经将检索式自动扩展为“b/手机 and r/手机”。

同样用“饭菜”对含有关键词“汉字输入法”的专利排序，其相关度最高仅为 27%。说明该专利与主题“饭菜”没有关系。而不像“手机”，与“饭菜”的相关度为 82%。这反应了引用关键字“汉字输入法”的专利主题明确、一致性。其中，排在前几位的申请就讲一个主题“汉字输入法”，所以这些申请与“汉字输入法”的相关度很高也是理所当然的。

Patentics 代表新一代智能搜索技术，通过超大规模计算智能，能自动理解上千万篇专利文献中蕴含的各种主题。有些搜索主题很明确、单一，如“汉字输入法”，有些主题模糊、多意，如“手机”。对于后者，如果用户想要搜索与通信不相关的主题，则用户只需添加感兴趣的主体。

上述分析可以应用到利用公开号进行搜索。有些专利申请全文主题明确清晰，单一，有些申请全文的主题模糊、东拉西扯。对于前者，往往只需输入一个公开号即可找到对应相关对比文献。而后者，则需要一些帮助，限定一个大致范围，这就是我们经常介绍的 B/ and R/。

如果大家希望更详细了解 Patentics 语义排序，我们有中国发明专利“CN101359331B 对搜索结果重新排序的方法和系统”和美国发明专利“8,126,883 Method and system for re-ranking search results”可以参考。



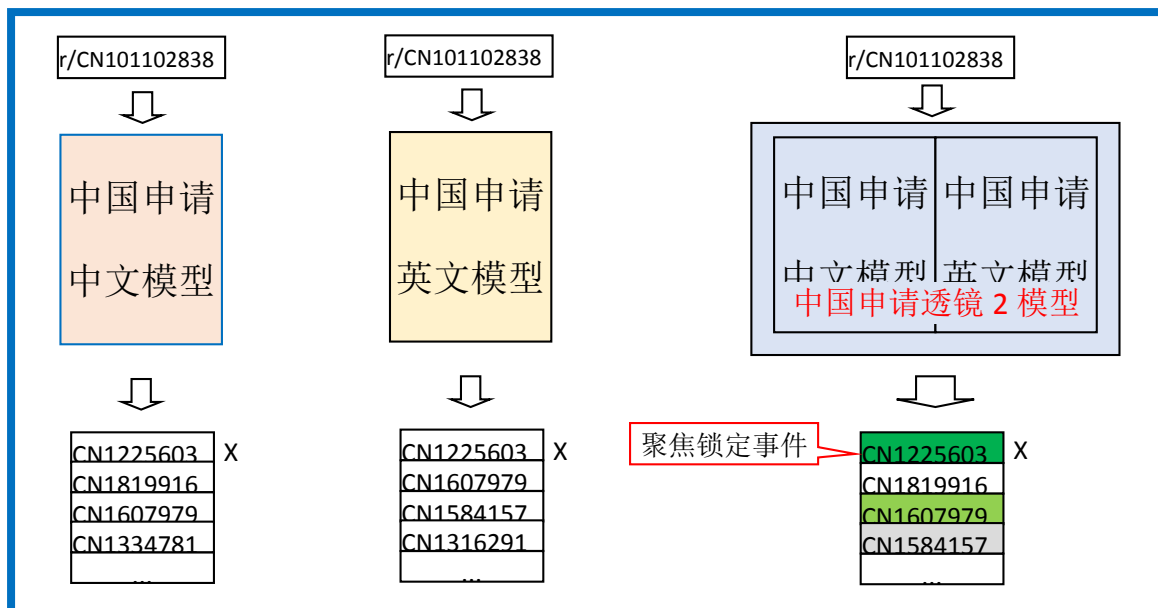
# Patentics 搜索透镜(1)

在检索过程中，经常会看到一些中国申请的公开号表有不同色度的绿色背景。这是 Patentics 搜索透镜发生智能“聚焦”。

The screenshot shows the Patentics search interface. At the top, there is a search bar with the query 'R/CN101035019'. Below the search bar, there are several navigation links: '搜索帮助', '字段组合', '智能搜索向导', '用户手册', and '可视化检索'. A red box highlights the '启用 Patentics 搜索透镜' (Enable Patentics Search Lens) button, with a note below it stating '必须只选用中国申请' (Must only select Chinese applications). Another red box highlights the '中国申请' (Chinese Applications) dropdown menu. Below the search bar, there are filters for '最相关400项结果' (Most relevant 400 results), '深绿表示第一位锁定' (Dark green indicates first digit lock), '中绿表示前三位锁定' (Medium green indicates first three digits lock), '浅绿表示前二十位锁定' (Light green indicates first 20 digits lock), '相关性(%)' (Relevance (%)), '过滤' (Filter), and '专利号' (Patent No.). The main part of the screenshot is a table of search results. The table has columns for '公开号' (Publication No.), '标题' (Title), '申请人' (Applicant), '发明人' (Inventor), '欧洲分类' (EPC Class), and '国际' (International). The table contains 15 rows of results. The first three rows have dark green backgrounds for the '公开号' column, the next three have medium green, and the last nine have light green. Red boxes with arrows point to these color-coded patent numbers and their corresponding annotations.

公开号	标题	申请人	发明人	欧洲分类	国际
CN101035019	快速收敛端到端业务的方法和装置	华为技术有限公司	李 莹	H04L12	H0
CN1009301	一种端到端业务快速收敛的方法和路由设备	华为技术有限公司	戴怡   吴小前	H04L12	H0
CN1933448	业务快速收敛	华为技术有限公司	戴怡   李莹	H04L12	H0
CN101355486	一种路由切换的方法、设备及系统	中国移动通信集团公司	葛 澍   戴 忠   孙金霞   尤 梦   孔令山   董智纯   段立鸿   叶 斌   秦 越   韦 芳   黄昭锐		H0
CN101163103	一种实现快速重路由的方法	孙先花	孙先花	H04L12	H0
CN101217457	实现快速重路由的方法及路由器	华为技术有限公司	李 莹   吕 鸿   宋宝华   怀 南   姜玉萍	H04L12	H0
CN101076029	一种边界网关协议递归路由更新的方法	中兴通讯股份有限公司	金 昊   朱 彤   王 征   钱 霄鸿	H04L12	H0
CN101697527	网络处理器实现快速收敛路由的方法及装置	中兴通讯股份有限公司	闫学涛   敖斌		H0
CN103595641	一种标签	杭州华三通信技术有限公司	姚跃腾   郭威   赵昌峰		H0
CN101848159	三层VPN路由重置方法和装置	中兴通讯股份有限公司	李宁   汪芳		H0
CN103297340	MPLS和BGP组网中的路由收敛方法和设备	杭州华三通信技术有限公司	王伟   樊超   王海		H0
CN101237409	MPLS VPN中实现快速重路由的方法及设备	华为技术有限公司	李 莹   吕 鸿   姜玉萍	H04L12	H0

Patentics 独创搜索透镜的基本思想是，在原有中国专利中文申请库上，通过中、英文机器翻译，构造出一个中国专利英文申请库。如果考虑语言是对事物的表示，则中国专利中文库与中国专利英文库是对相同发明内容的不同语言表示。通过我们的大量实验证明，中国专利英文库，具有与中国专利中文库相似的相关性语言模型。这两个模型相关但不相同，信息互补而不冗余。我们发现，如果对两个模型的相关排序结果进行特殊处理，就像光学装置中透镜一样，系统会对相关度排序输出结果，产生智能“聚焦”作用！



对二个语言模型进行自动测试。测试均无需任何检索策略，仅输入被审公开号，Patentics 排序计算结果与中国审查员给出国际检索报告（ISR）中的 X 文献进行比对。其中，TOP-i 表示前 i 篇自动计算排序结果中包含审查员给出 X 文献的命中率。如，TOP-1 表示，仅输入公开号，Patentics 计算排序第一位就有 9%的可能是审查员给出 X 文献。必须注意，在测试中我们严格假设，没有被审查员引用的其它相关文献都不算命中。

	TOP-1	TOP-20	TOP-100	TOP-400
中国申请中文模型	9.0%	27.39%	43.73%	59.44%
中国申请英文模型	9.1%	27.76%	44.78%	61.16%
中国申请透镜模型 (双模型-聚焦前)	9.1%	29.79%	47.15%	63.43%

从测试中，我们可以发现，就单语言模型比较（中国申请中文、中国申请英文）中国申请英文模型略好于中国申请中文模型。将两个语言模型通过搜索透镜算法聚焦融合，同样是输入被审公开号、同样是没有任何检索策略，其 TOP-20、TOP-100 的 X 命中精度，均提高 10%。

搜索透镜不仅提供了一个新的合成检索源，更为重要的是，把两个相关的检索源通过搜索透镜算法融合，构建成一个全新的智能检索决策源。通过对两个语言模型相容的排序结果的智能融合处理，我们可以计算和捕获单模型不可能出现的事件。比如说，当两个模型都将某一篇文章排序在第 1 位，更一般地，当两个模型都将某一篇文章排序在前 i 位（简称为“第 i 锁定”），这些锁定事件对 X 命中率的影响将会如何？

实际测试表明，如果某一篇文章被锁定在第 1 位，则该篇被锁定的文档是 X 文献的概率从 9% 跳升到 20%！在单语言模型中要通过人工浏览近 20 篇才能达到的精度，现在只需浏览被锁定的第一篇即可达到。如果两个模型都锁定某一篇在前 3 位，则被锁定前 3 篇中为 X 文献的概率为 34.3%，在单模型中要通过人工浏览 40 篇才能达到的精度，现在只需浏览最多 3 篇即可达到。如果两个模型都锁定某一篇在前 20 位，则被锁定前 20 篇中为 X 文献的概率为 67.9%，在单模型中要通过人工浏览多于 400 篇才能达到的精度，现在只需浏览最多 20 篇即可达到。

	TOP-1	TOP-3	TOP-20
X 命中率	20%	34.3%	67.9%

换句话说，Patentics 搜索透镜技术，创建一个新的智能语义模型，从另一个语言视角自动理解文本意思，**聚焦**最相关文献，帮助用户节省阅读时间，提高检索、浏览精度！

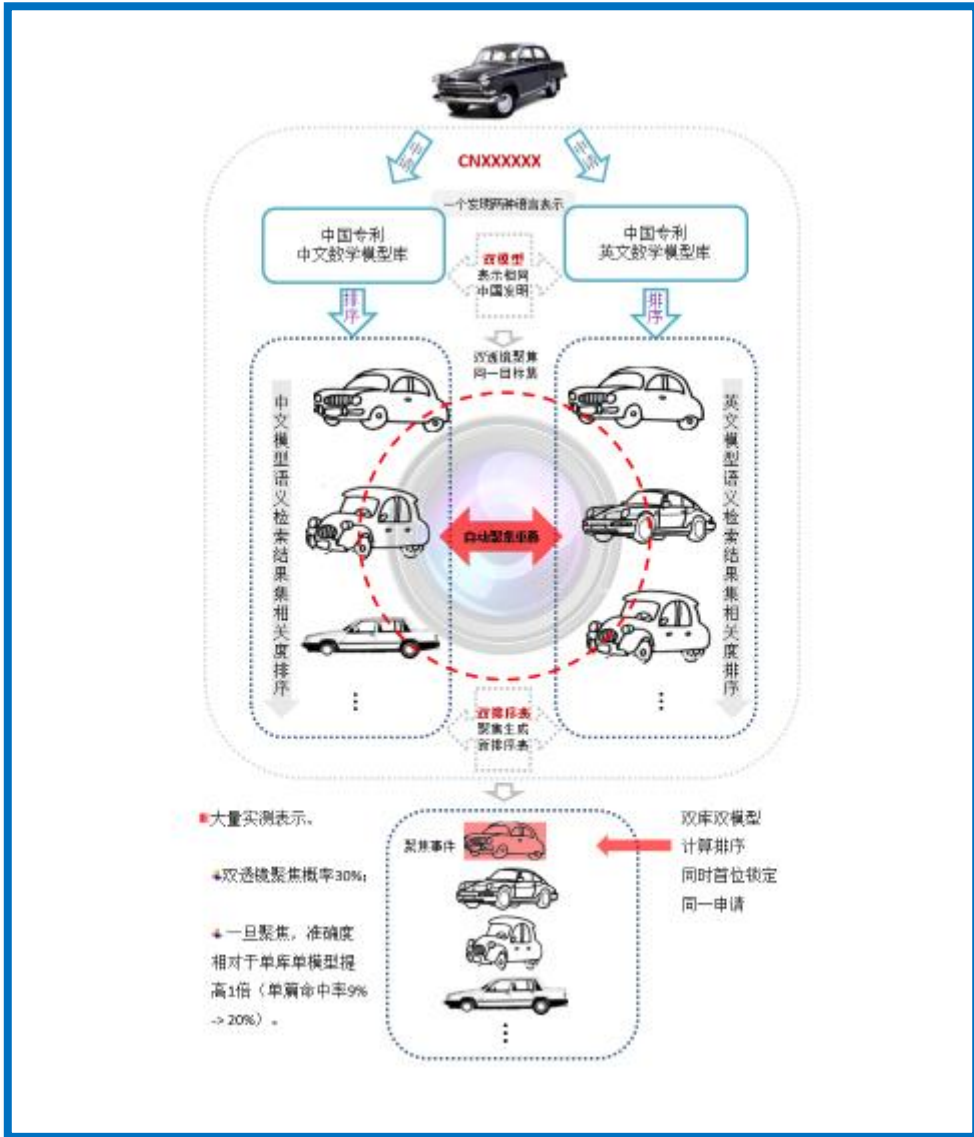
必须注意，为了使 Patentics 搜索透镜发生作用，目前支持数据库必须选“中国申请”库。这对检索中国对比文献，不是一个限制。因为，中国申请包含中国授权。

## Patentics 搜索透镜(2)

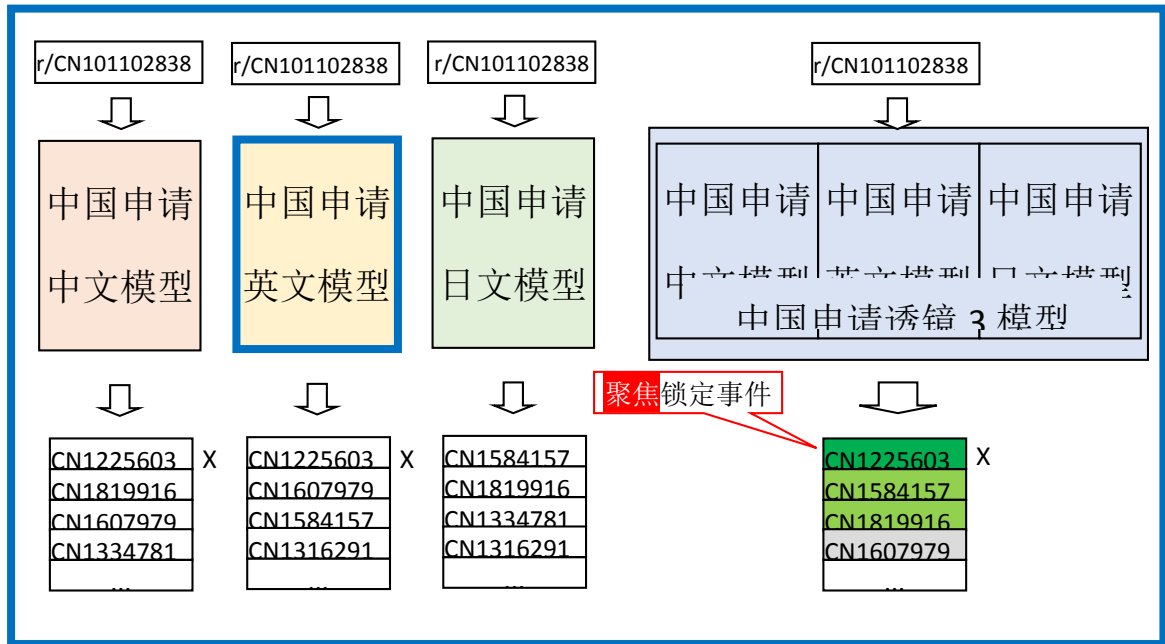
Patentics 搜索透镜的成功，是建立在计算机硬件越来越便宜，算法越来越复杂，用成倍、数十倍计算量来换取排序精度的提高的判断上。当 2009 年 1 月，Patentics 语义检索首次登上世界专利检索舞台，我们的语义引擎是简单的单模型-单知识库架构。当时，我们做了一个判断，随着 CPU、内存的性价比每年成倍提升，使我们可以通过大幅度提升算法的复杂度来提高排序精度的成为可能。

第一次尝试是采用单模型-多知识库架构，将一个全局知识库，切分成 10-20 个领域知识库。就像原来是一个全科医师诊治全部病人，现在变成多个专科医师诊治对应专科病人一样，多个领域知识库被用于对相关领域的专利全文进行排序。这样，Patentics 排序计算量翻了许多倍，但是排序的精度提升了 10%。虽然我们 must 淘汰我们原来的硬件，但是排序准确率提升 30%，对应的就是用户的检索、浏览效率提高了 30%。

第二次尝试是采用双模型-多知识库架构。我们知道，每个发明都有其固有的特征量。但是，通过不同语言表示，往往会对这些固有的特征量带来不同的信息“失真”。我们发明的多语言模型-多知识库架构，就是希望通过不同语言的互补表示模型，构造不同视角的“透视引擎”来对被描述的发明进行“全息透视”。又一次以算法复杂度成倍提升，计算量成倍增加为代价，换来了 Patentics 的排序准确率提升 10%，更关键的是，通过智能融合不同语言引擎的排序结果，我们又获取了一个新的排序决策量 -- 搜索“聚焦”，使前 20 篇的浏览效率提高 50%-100%。



上述 Patentsics 搜索透镜是通过中、英文模型排序输出互为参考、对比实现。作为我们新一轮的尝试，Patentsics 新一代搜索透镜将采用多于两种语言（中、英、日）表示中国发明。这样，**聚焦**融合与决



策算法可以采用多数表决来提高聚合排序命中率和**聚焦**命中率。

Patentics 搜索透镜对于中、英、日文用户来说，与正常检索一样，输入都是公开号。只是对输出排序结果中被**聚焦**的那些文档号进行特殊标记处理。如 CN101207555 被标绿色，表示该文献被中、英文模型都计算排序在第 1 位。以此颜色提示用户对该文献加以特殊关注，因为根据大量统计试验，这些被第一位**聚焦**的文献为 X 文献的概率是 20%。CN1819916 分别被中、日文模型排序在第 2 位，标浅绿色提示用户加以关注。

从此，全文搜索技术中出现一个新的词汇，您的搜索结果有没有被“**聚焦**”！

希望更多了解 Patentics 搜索透镜技术，可以参考中国专利申请 CN201310274941.X。

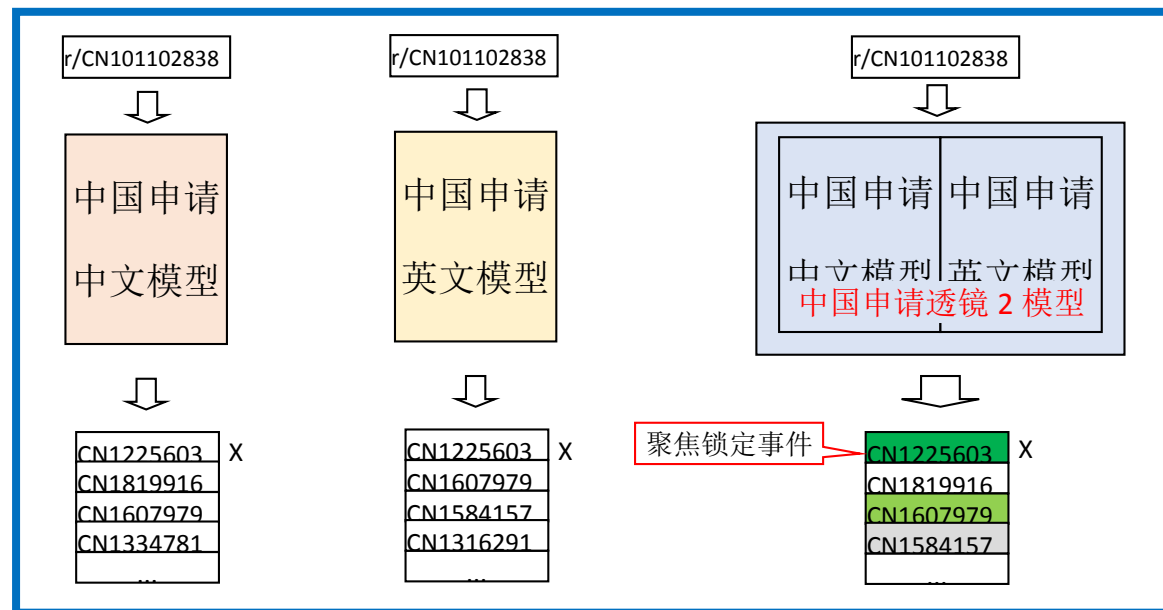
Patentics Search Interface Screenshot

Search Query: R/CN101202734 (Still input public number)

Search Results (Most Relevant 400 results):

PN	Title	Assignee	Inventors	ECLA	ICL	Rank
CN101202734	一种完全基于...的方法	中兴通讯股份有限公司	金昊   林光辉	H04L12	H04L	100%
CN101807524	一种路由黑洞避免过程中过位自动清除的方法	中兴通讯股份有限公司	金昊   林光辉	H04L12	H04L	99%
CN101789900	组播转发路径的查询方法、中间节点和管理节点	福建星网锐捷网络有限公司	王秀文   杨敬民		H04L	97%
CN102739543	一种...的最大序列号翻转的处理方法和装置	杭州华三通信技术有限公司	林长望		H04L	97%
CN102916894	一种...的方法	杭州迪普科技有限公司	石东豪		H04L	97%
CN102217374	一种...的方法	华为技术有限公司	王江胜   王自强		H04W	97%
CN101394354	一种避免中间系统邻居关系震荡的方法和一种中间系统	杭州华三通信技术有限公司	赵昌峰		H04L	97%
CN1543129	BGP路由快速收敛方法	港湾网络有限公司	陈国文	H04L12	H04L	97%
CN1689279	一种...的方法	艾利森公司	H·德西内尼   K·阿瓦萨拉拉   D·奥伦   K·P·施瓦茨	H04L12	H04L	97%
CN101350789	一种...的方法	杭州华三通信技术有限公司	宋新超   钱雪彪   魏志广   张蓓		H04L	97%
CN101909065	网络设备中的地址解析协议表的老化更新方法和网络设备	北京中创信测科技股份有限公司	贾林   朱建仪		H04L	96%
CN102035738	一种获取路由信息的方法及装置	中兴通讯股份有限公司	张明之		H04L	96%
CN101771616	一种报文转发和终结的方法和系统	中兴通讯股份有限公司	张胜友		H04L	96%
CN101969395	节点配置和路径探测方法、系统和网络设备	华为技术有限公司	宋建民   马鹏   陈大鹏		H04L	96%
CN101257448	一种提高RSVP-TE隧道可靠性的方法	中兴通讯股份有限公司	张新村   周燕青	H04L12	H04L	96%
CN101834772	一种基于以太网的介质访问控制地址分类方法和装置	中兴通讯股份有限公司	宋晓丽		H04L	96%
CN102625402	一种关于移动Ad-Hoc网络路径发现和选择的方法	昆明理工大学	邵玉斌   龙华   许伶俐   董金升		H04W	96%
CN101035080	一种...的方法	杭州华为三康技术有限公司	姜冬华	H04L12	H04L	96%
CN102025642	一种路由器标识冲突的检测方法、装置及路由设备	北京星网锐捷网络技术有限公司	余峰楼		H04L	96%
CN102664816	一种快速查找MPLS转发表的装置及方法	迈普通信技术股份有限公司	孟庆超   孙健		H04L	96%

Page 1 / 20 [1] [2] [3] [4] [5]



对二个语言模型进行自动测试。测试均无需任何检索策略，仅输入被审公开号，Patentics 排序计算结果与中国审查员给出国际检索报告（ISR）中的 X 文献进行比对。其中，TOP-i 表示前 i 篇自动计算排序结果中包含审查员给出 X 文献的命中率。如，TOP-1 表示，仅输入公开号，Patentics 计算排序第一位就有 9%的可能是审查员给出 X 文献。必须注意，在测试中我们严格假设，没有被审查

员引用的其它相关文献都不算命中。

	TOP-1	TOP-20	TOP-100	TOP-400
中国申请中文 模型	9.0%	27.39%	43.73%	59.44%
中国申请英文 模型	9.1%	27.76%	44.78%	61.16%
中国申请透镜模型 (双模型-聚焦前)	9.1%	29.79%	47.15%	63.43%

从测试中，我们可以发现，就单语言模型比较（中国申请中文、中国申请英文）中国申请英文模型略好于中国申请中文模型。将两个语言模型通过搜索透镜算法**聚焦**融合，同样是输入被审公开号、同样是没有任何检索策略，其 TOP-20、TOP-100 的 X 命中精度，均提高 10%。

搜索透镜不仅提供了一个新的合成检索源，更为重要的是，把两个相关的检索源通过搜索透镜算法融合，构建成一个全新的智能检索决策源。通过对两个语言模型相容的排序结果的智能融合处理，我们可以计算和捕获单模型不可能出现的事件。比如说，当两个模型都将某一篇文章排序在第 1 位，更一般地，当两个模型都将某一篇文章排序在前 i 位（简称为“第 i 锁定”），这些锁定事件对 X 命中率的影响将会如何？

实际测试表明，如果某一篇文章被锁定在第 1 位，则该篇被锁定的文档是 X 文献的概率从 9%跳升到 20%！在单语言模型中要通过人工浏览近 20 篇才能达到的精度，现在只需浏览被锁定的第一篇即可达到。如果两个模型都锁定某一篇文章在前 3 位，则被锁定前 3 篇中为 X 文献的概率为 34.3%，在单模型中要通过人工浏览 40 篇才能达到的精度，现在只需浏览最多 3 篇即可达到。如果两个模型都锁定某一篇文章在前 20 位，则被锁定前 20 篇中为 X 文献的概率为 67.9%，在单模型中要通过人工浏览多于 400 篇才能达到的精度，现在只需浏览最多 20 篇即可达到。



	TOP-1	TOP-3	TOP-20
X 命中率	20%	34.3%	67.9%

换句话说，在搜索透镜技术中，通过创建一个新的智能语义模型，从另一个语言视角自动理解文本意思，**聚焦**最相关文献，帮助用户节省阅读时间，提高检索、浏览精度！

还是输入公开号

还是选择中国申请中文库

聚焦锁定第 1 位，20%可能是 X

聚焦锁定前 3 位，3 篇中 34%可能是 X

聚焦锁定前 20 位

聚焦锁定前 20 位

PN	Title	Assignee	Inventors	ECLA	ICL	Rank
CN101202734	一种完全... 堆栈删除的方法	中兴通讯股份有限公司	金昊   林光辉	H04L12	H04L	100%
CN101202734	一种由黑洞避免过程中过载位自动清除的方法	中兴通讯股份有限公司	金昊   林光辉	H04L12	H04L	99%
CN101789900	组播转发路径的查询方法、中间节点和管理节点	福建星网锐捷网络有限公司	王秀文   杨敬民		H04L	97%
CN102739543	一种... 的最大序列号翻转的处理方法和装置	杭州华三通信技术有限公司	林长望		H04L	97%
CN102916894	一种... 的方法和装置	杭州迪普科技有限公司	石东豪		H04L	97%
CN102217378	一种... 的方法和装置	华为技术有限公司	王江胜   王自强		H04W	97%
CN101394354	一种避免中间系统邻居关系震荡的方法和一种中间系统	杭州华三通信技术有限公司	赵昌峰		H04L	97%
CN1543129	BGP路由快速收敛方法	港湾网络有限公司	陈国义	H04L12	H04L	97%
CN1689279	一种... 的方法和装置	艾利森公司	H·德西内尼   K·阿瓦萨拉拉   D·奥伦   K·P·施瓦茨	H04L12	H04L	97%
CN101350789	一种... 的方法和装置	杭州华三通信技术有限公司	宋新超   钱晋彪   魏志广   张蓓		H04L	97%
CN101909065	网络设备中的地址解析协议表的老化更新方法和网络设备	北京中创信测科技股份有限公司	贾林   朱建仪		H04L	96%
CN102035738	一种获取路由信息的方法及装置	中兴通讯股份有限公司	张明之		H04L	96%
CN101771616	一种报文转发和终结的方法和系统	中兴通讯股份有限公司	张胜友		H04L	96%
CN101969395	节点配置和路径探测方法、系统和网络设备	华为技术有限公司	宋建民   马鹏   陈大鹏		H04L	96%
CN101257448	一种提高RSVP-TE隧道可靠性的方法	中兴通讯股份有限公司	张新林   周慈善	H04L12	H04L	96%
CN101834772	一种基于以太网网的介质访问控制地址分类方法和装置	中兴通讯股份有限公司	宋晓雨		H04L	96%
CN102625402	一种用于移动Ad-Hoc网络路由发现选择的方法	昆明理工大学	邵玉斌   龙华   许伶俐   董会升		H04W	96%
CN101035080	邻居路由器及广播网络	杭州华为三康技术有限公司	姜冬华	H04L12	H04L	96%
CN102025642	一种路由器标识冲突的检测方法、装置及路由设备	北京星网锐捷网络技术股份有限公司	余峰楼		H04L	96%
CN102664816	一种快速查找MPLS转发表的装置及方法	迈普通信技术股份有限公司	孟庆超   孙健		H04L	96%

Page 1 / 20 [1] [2] [3] [4] [5]

从此，搜索技术中出现一个新的词汇，您的搜索结果有没有被“**聚焦**”！

## Patentics 快捷功能按钮---审查员必读

Patentics 有许多功能，帮助用户，特别是每天需要检索、浏览大量专利文献的审查员高效完成任务。鉴于绝大多数审查员都没有受过任何 Patentics 培训，除了相互介绍 Patentics 语义搜索外，对于我们专门为专业专利检索设计的其它许多功能，都没有很好地用起来。为此，我们在这里将系列介绍 Patentics 功能。如果有审查员在我们的订阅号上，希望大家能把相关信息转发给你们的同事们。帮助大家用好 Patentics，快速找到 X 和 Y 文件，回家！

Patentics 有许多快捷功能，帮助用户高效浏览文献。最常用的快速比对功能。同时打开 20 篇（搜索结果）快速浏览。包括快速比对摘要、比对权利要求、比对 20 篇全部专利附图（下图所示）。

140项结果: 相关性(%):0 过滤 专利号: [ ]

公开号	标题	申请人	发明人	分类	国际分类	相似度
8,301,130	检测和接通无线电话服务和local phone networks	斯普林特通信	Gunasekara; Don   Chapman; Shane David	455	H04M	79%

摘要|主权利要求|登录|参考引用|分类|**图片**|索引|相关概念及专利|新颖分析|侵权分析|同族|信息

8,301,130 Detect and connect wireless phone services and local phone networks

缩略图 首页 | 上一页 | 下一页 | 尾页 | 全屏 | 新增缩图

2005/0282492 无线手机概念

摘要|主权利要求|登录|参考引用|分类|**图片**|索引|相关概念及专利|新颖分析|侵权分析|信息

2005/0282492 Wireless phone concept

缩略图 首页 | 上一页 | 下一页 | 尾页 | 全屏 | 新增缩图

2004/0266478 wireless phone adapter

摘要|主权利要求|登录|参考引用|分类|**图片**|索引|相关概念及专利|新颖分析|侵权分析|信息

King, Anthony G. | King, Francis | 455 | H04M | 76%

20篇专利附图同时打开

Patentics 除了把全球专利的数十亿个专利附图全部嵌入专利全文，提供原位显示（EPO 想做也没有做到，这也是我们的一个亮点），我们还提供任意 4 个专利的附图比对（上图所示）。这些附图快捷功能，对于机械、装置类专利审查员，有非常大的帮助！

公开号	标题	申请人	发明人	分类	国际分类	相关性
8,301,130	检测和接通无线电话服务和local phone networks	斯普林特通信	Gunasekara; Don   Chapman; Shane David	455	H04M	79%

摘要|主权权利要求|超录|参考引用|分类|图片|索引|相关概念及专利|新颖分析|侵权分析|同族|信息

8,301,130 Detect and connect wireless phone services and local phone networks

被体现的有计算机可执行指令的一个或多个非短暂的计算机可读存储,在第一电话上由计算机通信来电executed到第二电话的当,电脑可读媒体包括:确定,是否连接到标志网关的多个电话,其中配置通信的标志网关来电到一个多个电话,并且,其中标志网关包括至少天线,发射机,制式转换器,电力管理器,landline-interface 模块,瘦客户机模块,发音器,接收器,和双音多频(双音多频)模块;配置连接到标志网关的多个电话,其中具有唯一环音的每个电话rings;发送来电到第一电话的预定目的地;发送第一电话的环音,以致连接到具有第一电话的环音的标志网关ring的多个电话;提供环音到第二电话,其中第一电话的环音是表示的,以致应答第二电话理解来电的用户最初地驶往第一电话。

4567890 1 3

2 输入对比专利号

还有许多有这个按钮的地方,都实现原位对比

本专利附图

4567890 专利附图

里面还有许多小按钮,提供许多 Magic,我们都会向大家介绍。

## SIPO 与 EPO 英文审查报告结果自动比对

至 2012 年底，采用 SIPO 递交 PCT 的国际检索报告申请中，共有 11135 篇进入 EP。为此，EPO 根据欧洲专利法，对这些进入 EP 的专利申请重新做了 Supplementary European Search Report（欧洲补充检索报告）。

The screenshot shows the European Patent Register interface for application EP2499770. The main title is "About this file: EP2499770". The search results show "EP 2499770 - DOWNSTREAM FRAME SYNCHRONIZATION PATTERN PROTECTED BY HEADER ER GIGABIT PASSIVE OPTICAL NETWORK".

Key details from the record include:

- Status:** Request for examination was made. Database last updated on 29.10.2013.
- Most recent event:** 03.04.2013, Amendment by applicant.
- Applicant(s):** Huawei Technologies Co., Ltd. (Shenzhen, Guangdong 518129 / CN).
- Inventor(s):** LUO, Yuanqiu (Cranbury, New Jersey 08520 / US) and EFFENBERGER, J. Frank (Freehold, New Jersey 07728 / US).
- Representative(s):** Kreuz, Georg Maria, et al. (München / DE).
- Application number, filing date:** 10637015.6 (06.12.2010), WO2010CN79453.
- Priority number, date:** US20090287024P (16.12.2009), US20100884566 (17.09.2010).
- Publication:** Type: A1 (Application with search report), No.: WO2011072578, Date: 23.06.2011, Language: EN.

Red boxes and arrows highlight the following search reports and documents:

- International and Supplementary search report(s):** International search report - published on: CN; Supplementary European search report - dispatched on: EP. A red box labeled "EP审查员 引用2个X" points to the EP report.
- Documents cited:**
  - Search: EPO ESP (D0 \* US2009262130, D0 \* EP1265548)
  - International search: SIPO ISP (WI \* CN1912410, WI \* C\*\*01047470, WI \* KR20010048322)

A red box labeled "SIPO审查员 引用4个A, 没有发现X" points to the SIPO ISP entries.

在 11135 篇对应的 EPO 申请中，我们发现 1598 篇中国专利局递交的 PCT 国际检索报告，仅有 A 文献，而 EPO 做的欧洲补充检索报告恰引用了 X 文献。

以该 1598 篇申请案，我们采用 EPO 补充检索报告中 EPO 审查员给出的英文 X 文献为测试标准，对 Patentics 的智能排序技术，进行了自动测试。

测试分两部分进行，

1. 采用英文申请全文文本，没有任何检索策略，仅输入英文申请公开号，在 900 万英文全文库中进行智能排序，并与 EPO 审查员给出 X 文献比对；



2. 假设，没有英文申请全文文本，仅有中文申请全文文本，没有任何检索策略，仅输入中文申请公开号，由 Patentics 智能跨语言检索系统的机器翻译部件，自动将中文翻译成英文，在 900 万英文全文库中进行智能排序，并与 EPO 审查员给出 X 文献比对。

将 Patentics 智能排序计算的结果进行比对，记录匹配位置，统计得如下结果，

	第 1 位命中	第 20 位命中	第 100 位命中	第 400 位命中
英文申请全文	5.4%	18.63%	33.82%	50.05%
中文申请全文	4.8%	16.32%	31.15%	48.68%

显然，直接用中国公开号(利用 Patentics 机器翻译的中国申请全文)，要比直接用英文文本(人工翻译)检索，损失 2 个百分点。

更进一步，对于 11135 篇进入 EP 的 SIPO 递交的 PCT 国际检索报告，我们同样采用国际检索报告中 SIPO 审查员给出的英文 X 文献为测试标准，对 Patentics 的智能排序技术，进行了自动测试。

	第 1 位命中	第 20 位命中	第 100 位命中	第 400 位命中
英文申请 全文	4.9%	14.82%	26.98%	41.80%
中文申请 全文	3.9%	13.85%	26.76%	40.50%

显然，与采用 EPO 审查员的英文 X 文献的命中率相比较，SIPO 审查员给出的英文 X 文献与 Patentics 的智能排序结果匹配命中率要低 8-10%。这是否表示 EPO 审查员给出的英文 X 文献要比 SIPO 审查员给出的英文 X 文献更加相关。

# 用大数据分析认定、辩护公知技术推断

发明人、申请人在回答审查意见，或审查员撰写审查意见，断定技术特征为公知技术之前，应该参考 Patentics 对该申请的技术特征的大数据分析。

我们曾有一实际案例，很能说明问题，特将当时步骤重复如下。

CN200810105725.1, 是我们公司的一篇申请。该申请通过复审，最终获授权。其中一次审查意见中，审查员提出“在进行搜索时，采用重排序表达式表达用户的重排序需求，是本领域技术人员的常用技术手段”。当时，我们正在研发 Patentics 技术特征大数据分析，因此属于好奇，我们就把该申请号直接输入 Patentics 专利输入框。点击标题，点击“主权权利要求”（这是不用检索，直接快速浏览一篇专利信息的方法）。

15项结果: [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标] [图标]

输入公开号 或申请号 专利号: CN200810105725.1 [图标]

公开号	标题	申请人	发明人	欧洲分类	国际分类
CN101359331B	对搜索结果重新排序的方法和系统	索意互动(北京)信息技术有限公司	裘钢	点击	G06F

摘要 | 主权权利要求 | 题录 | 参考引用 | 分类 | 图片 | 索引 | 相关概念及专利 | 新颖分析 | 侵权分析 | 同族 | 法律状态 | 信

89% [35|14, 37|11] 一种对搜索结果重新排序的方法，包括：  
步骤一：根据接收到的搜索表达式对数据源进行搜索并获得搜索结果；  
步骤二：获得用户兴趣，该用户兴趣是由用户输入的重排序表达式，其中重排序表达式包括重排序内容，其中重排序表达式与搜索表达式同时输入，根据重排序表达式，对步骤一的搜索结果中的所有文档进行重排序；  
步骤三：以所述排序来显示所述搜索结果中的所有文档。

- 重排序.2275
- 同时输入.8033
- 排序.67844
- 收到的.174739
- 重排序内容.2
- 搜索表达式.60
- 搜索.109268
- 重新排序.7304
- 显示.1345828
- 搜索结果.12231
- 用户输入.78199
- 数据源.22589
- 重排序表达式.1
- 文档.43674
- 用户兴趣.1562

该申请主权项有 14 个技术特征。其中四种不同颜色是四类主题意思。每一技术特征旁的数字为引用该技术特征的文献篇数。

在这里，审查员认定技术特征“重排序表达式”为公知技术，显然得不到数据的支持。因为根据 Patentics 的大数据分析，“重排序表达式”仅被 1 篇文献引用过，就是本申请。

这是连我们自己也意想不到的结果。我们将该特征放到其它网站上搜索，都



只能得到本申请。为此，我们在回复中进行了根据数据实例的说明，最后审查员放弃公知技术的认定并给以授权。

因此，如果发明人拿到审查员的公知技术认定，不妨将该申请的技术特征要素来个大数据分析。如果被引用次数很小，自然可做基于数据事实的辩护，如果已经引用次数很多，放弃可能也是一个上策。审查员在认定该技术特征为公知技术前，最好也查一下这些被推断为公知技术的技术特征，确保从实际引用数据上站的住脚。



我们还对 1000 万中国专利申请、授权都进行了技术特征的大数据分析。上述图中 89%[35|14, 37|11] 分别表示该段文字(主权项)与本文档的相关度为 89%。35|14 表示本文档的权利项数(专利度)为 35, 主权项的技术特征数(特征度)为 14。37|11 表示对偶专利(授权专利)的专利度|特征度。

同时, 当该申请有授权版本时, Patentics 会显示一小图标, 点击后会对比显示申请与授权版本的变化。为了获得授权, 满足审查员的审查意见, 粉红为插入, 黄色为删除。如果大家需要评估该专利的稳定度, 一眼了解这些变化记录, 非常有帮助!

# Patentics 语义检索过程标准化

我们以科学为唯一索求。

Patentics 于 2009 年上线以来,大量的中国专利局审查员在每天的审查工作中使用我们的服务。我们的微信无非就是希望帮助审查员更好掌握新一代语义检索技术,为提高中国专利审查质量尽一份力。如果说我们的分析万一哪里冒犯了哪位领导,请看在科学的名义上,宽恕我们!

【大家一定读完,虽说下面的开头看起来像是昨天的删除版,但是我们加了许多新的内容,附图也重做了,还包括从来没有写过的 **Patentics 语义检索标准化过程**】

许多审查员对 Patentics 用户界面提出了改进建议。其中,最突出的是希望对搜索结果的排序与搜索结果的附图的高速批量显示集成。更进一步,提出希望能在浏览每一篇文献的全部附图(不是首图)时,能够同时快速阅读每一篇文献的摘要。就是说希望图文结合,最相关的最先看!

为满足审查员快速检索、浏览、理解直至找到高相关度的对比文件,Patentics 已与近日推出智能语义图文检索系统界面 [www.patentics.com/scn.htm](http://www.patentics.com/scn.htm) (中文界面)和 [www.patentics.com/s.htm](http://www.patentics.com/s.htm) (英文界面)。当检索结果出来时,第一时间显示图文信息,而且每个页面的显示项数 20/50/100 可调。

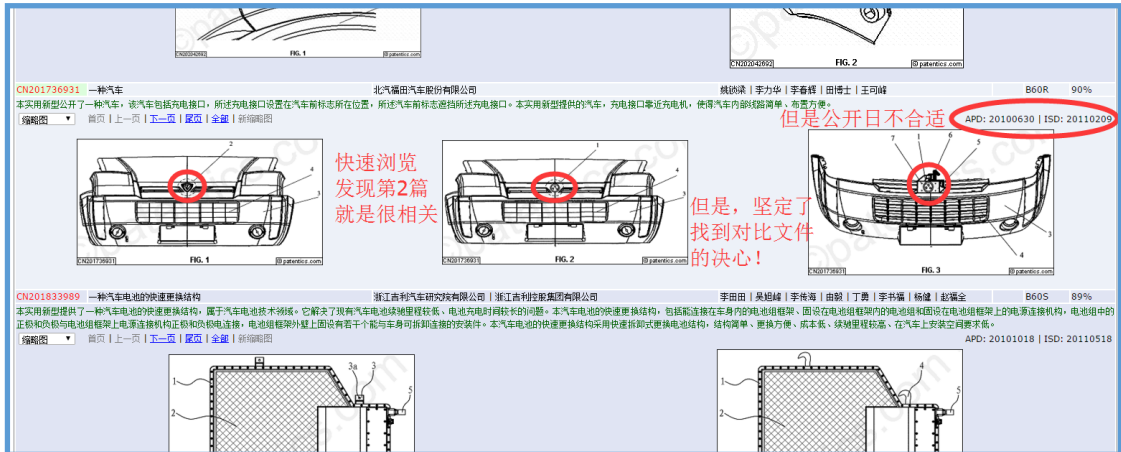
下面，我们还是以“一种可电驱动的车辆”的检索过程分析为实例。通过本实例，我们希望发展一种标准检索过程，帮助审查员能快速找到中文、英文对比文件。

输入(申请号或公开号)

**cn200710146226.2**

快速浏览到第二篇，从图中发现很相关文件。查看公开日是在本申请后，虽不能用作对比文件，但是表示排序的方向是正确的，坚定了找到对比文件的

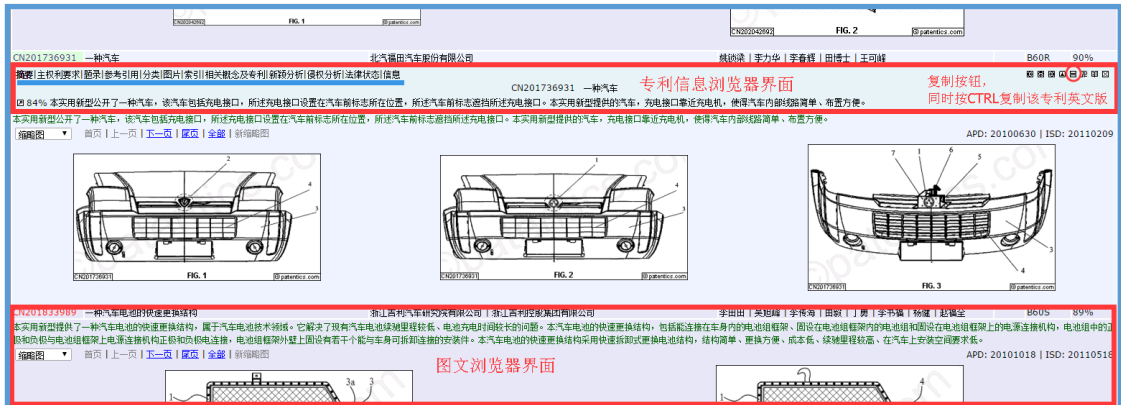
决心！



虽然该申请不能做对比文件，但是这并不妨碍我们通过该申请获取相关信息，比如说英文技术特征。

做为标准化操作流程，我们介绍**图文信息浏览器**界面与**专利信息浏览器**间快速切换，通过**专利信息浏览器**界面快速获取该专利的英文翻译版本。

大家一定熟悉Patentics**专利信息浏览器**界面，里面包含摘要，机器标引主权项，机器标引关键词项，同族、法律状态等信息。许多信息对于查新审查检索，可能暂不需要。当输入公开号或申请号进行相关度排序时，审查员一定是着急找到最相关的对比文件。为了帮助大家第一眼就能看出个大概，最大的帮助莫过于把专利附图与摘要同时快速原位显示，快速浏览。

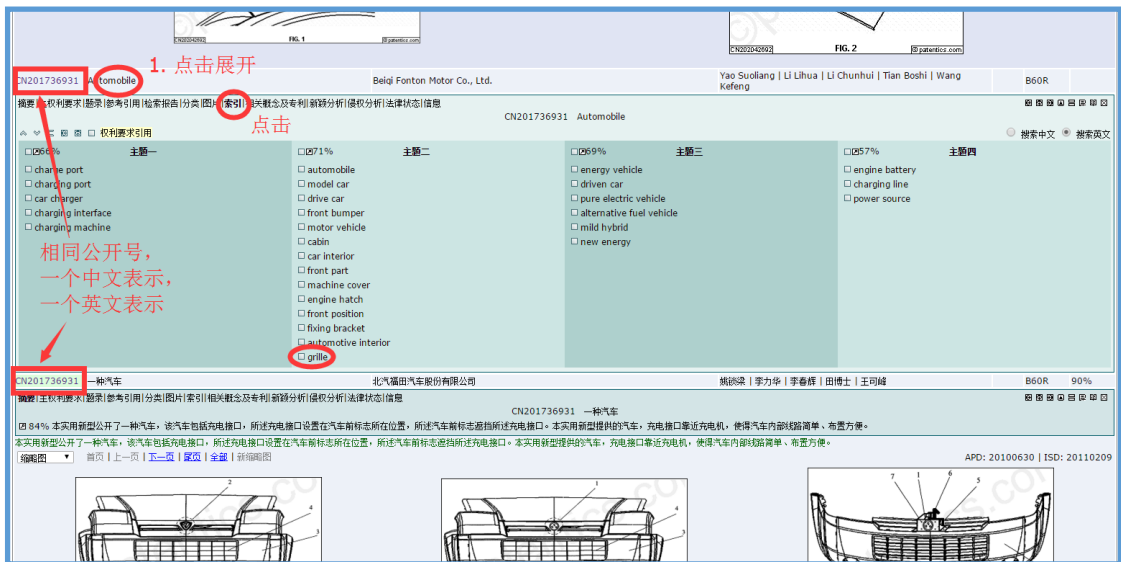


Patentics带有高准确度的中国专利申请的全文英文库（近1000万篇）以及对应的智能机器标引的英文关键概念（1000万中文+1000万英文，人工标引？）。通过2个点击，我们就可以快速、原位获取这些关键信息。

点击标题，获得**专利信息浏览器**界面。按下CTRL键并同时点击右上角“**复制**”小按钮，复制该中国申请号的英文翻译版本。



点击**专利信息浏览器**中的标题，点击“索引”，即可获得机器标引的该申请CN201736931的关键技术特征。我们发现“grille”（格栅）是一个非常好的关键技术特征的英文描述。



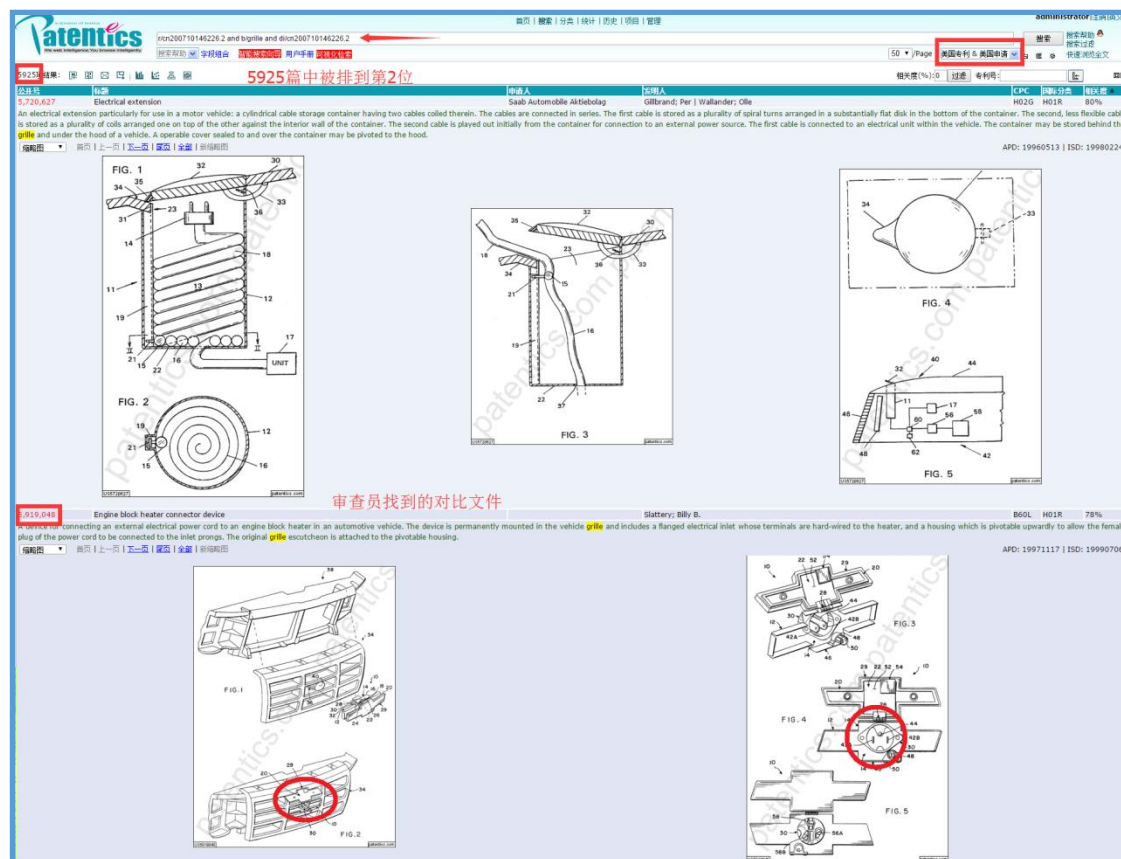
选择美国全文授权、申请库，还是用该中国申请做排序基准。这时候系统会自动采用中国专利申请的英文翻译全文，对英文文献进行排序。新的检索式为，

**r/cn200710146226.2 and di/cn200710146226.2 and b/grille**

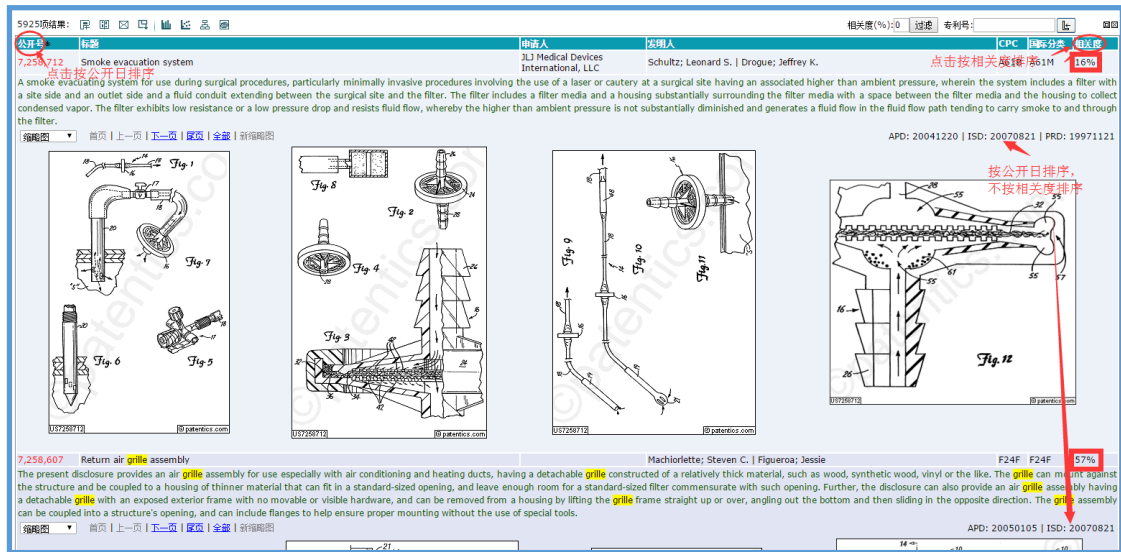
对所有公开日在cn200710146226.2申请日前（20070829）并包含关键词

grille的美国专利、申请，按中国申请号cn200710146226.2（的英文版）的语义进行自动排序。

快速浏览美国专利图文信息，被Patentics从5,925篇中排在第2位的**US5919048A**，只需扫一眼该美国专利的附图，就知道这是一篇大家想要找的对比文件！



有人可能认为，如果知道关键词“grille”，不用语义排序也可以找到**US5919048A**。这个我们可以试试看。大家可以点击“公开号”，这样检索结果就不按语义相关度排序，5,925篇美国专利将按公开日排序，**US5919048A**会出现在什么位置？出现在第1,559位。显然，没有语义相关度排序，5,925篇都一样重要，每一篇都必须阅读，最后必须借助于检索策略将数据集缩小到可阅读范围。这时，希望对比文件非常有可能已经被排除在外。



相比之下，Patentics语义排序，检索结果一篇都不少（5925篇），每一篇通过计算与被审发明内容的语义相关度，即两两间发明内容的相关关系来进行排序。相关度高的排在前面，相关度低的排在后面，决不会像传统检索靠去除检索策略来缩小阅读量那样的漏检。

Patentics语义检索，把专利查新变成一个标准化的过程。我们总结一下，

1、先用公开号或申请号对中国申请数据库排序（对1000万申请全文进行排序），

**r/cnxxxxxx and di/cnxxxxxx**

一般快速浏览前100篇，根据我们与国际检索报告的自动测试表明，前100位包含X对比文件的概率为48%左右；

2、如果通过快速浏览没有发现合适的对比文件，可以考虑提供一个简单检索策略，将排序范围从全部数据库（千万级）压缩到特定范围（百万、数十万... 都可）。可以是大家常用的布尔检索式定义，如关键词**b/**，IPC分类**icl/**等等，这个特定范围可以很大，因为最后可以通过排序将最相关的排到前面；

**r/cnxxxxxx and di/cnxxxxxx and b/xyz**

**r/cnxxxxxx and di/cnxxxxxx and icl/xxxx**

...更多请见“字段组合” [搜索帮助](#) [字段组合](#)。

3、如果在中文库里不能找到合适对比文件，可以进一步考虑到英文专利申请库继续找。为了找到合适的关键技术点来限定范围，可以考虑先在中国全文

库中找到最相关的申请，然后通过CTRL+“复制”按钮获得该申请的英文全文，就如本文前面所介绍的过程。借助于Patentics中国专利申请英文库，还是用原来的中国公开号、申请号为排序基准，只需把数据库指到美国、EP等即可。

关于数据库的选择，我们建议根据全文数据的构成考虑。对于中国专利数据库，建议使用中国申请库，不加授权库。因为中国申请库包含每一篇授权的公开披露。加入授权库，里面许多信息都是冗余的。美国全文库则不一样，因美国的授权库与申请库不是全部重合的，因此一般需选择授权、申请。

不靠运气，你能捡到我检不到还有他也检不到。靠的是r/cnxxxxxxx排序。这个cnxxxxxxx号就是大家都知道的标准号-公开号或申请号！所以我们说现代专利检索技术与传统布尔检索技术的最大差别，就是检索过程标准化了！

下一节我们还要介绍，如何使用Patentics排序加简单检索策略，对EPO对比文件进行回溯分析，找出“推测欧专局可能具有特殊的检索工具或检索方式，所以获得了这两篇文件”。



# 审查员幸运模式与 Patentics 自动测试比对

许多审查员表示，在他们的审查工作中，有 20%到 30%的案子，不需要任何检索策略，只需一个申请、公开号，就能在第一屏（前 20 篇）快速找到对比文件。审查员们称之为 lucky 模式（幸运模式）。这个 lucky 模式，其实和我们的数十万检索报告的自动测试结果相符。

	TOP-1	TOP-20	TOP-100	TOP-400
中国申请中文 模型	9.0%	27.39%	43.73%	59.44%
中国申请英文 模型	9.1%	27.76%	44.78%	61.16%
中国申请透镜模型 (双模型-聚焦前)	9.1%	29.79%	47.15%	63.43%

以 SIPO 递交的国际检索报告为自动测试比对的标准，我们有个程序能输入公开号和时间限制，

`r/cnxxxxxx and di/cnxxxxxx`

通过算法自动排序，获得最相关的 400 篇文献。把自动排序结果与审查员在该申请的检索报告中给出的 X 文献号比对，如果排序在第一位与 X 文献相同，就在第一位统计加 1，如果在第二位相同，就在第二位统计加 1，一直到 400 项。如果没有在 400 项之内出现，就算没有命中。最后对这些命中数据进行统计处理。

其中，最重要的是被 Patentics 排在第一位的(TOP-1)，如果不看其它文献，就看第一篇，9%的可能该文献就是审查员心目中的 X 文献。大家可以做个简单的统计试验，每次输入公开号排序，浏览第一篇，同时把是否 X 文献记下。做 100 个案子，一定有 9 篇左右是你认定的 X 文献。

看第一篇，9%成功率显然不够。许多审查员往往浏览第一屏，即排序前 20 篇（TOP-20）。这就是大家说的 luck 模式。审查员反映的 luck 模式命中率（20%-30%，取 25%）与我们自动测试的 TOP-20 完全相符。在对 TOP-20 进行统计时，

我们将第 1 到第 20 篇的命中次数进行累加，最后除以全部测试篇数，得出 27% 左右的命中率。就是说，如果你浏览前 20 篇，就会有 27% 的概率找到 X 对比文件。

如果再花点时间多看几篇，比如 100 篇。X 文献的命中率 TOP-100 是 44% 左右。显然增加的命中率与投入的时间不成正比。多看 80 篇，命中率仅增加 17%。至于再多看 300 篇，到 400 篇的命中率 TOP-400 增加更少，仅 16%。

这本身充分验证了 Patentics 语义排序的科学性。最相关的排在最前，排在最前的与审查员人工找出的 X 文献相符率最高。排序越到后，相关度越低，与审查员人工找出的 X 文献相符率越低。

有了这个统计结果，大家就可以好好规划自己的检索过程。第一屏一定要看。做 4 个案子一定有一个已经排在里面了！

而且，根据我们的自动测试，命中率几乎与 IPC 学科无关。最多是小于 1% 的上下浮动。

上面的介绍，都是基于输入一个公开号，没有任何检索策略，当然没有任何漏检的可能，一步找到对比文件。为理想检索模式。许多审查员反应，在理想模式中找出的对比文件由于相关度高，不需要公知技术搭接，更容易结案。

对于剩下的 75% 案子，许多审查员问，是不是可以做下去，一直到完成检索任务。我们的答案是肯定的。

25% 命中率的方案是不给 Patentics 任何帮助，在全部数据库（上千万篇）里“大海捞针”。如果给系统一点帮助，用一个简单的检索式圈定一个（大）范围。不是在上千万里面排序，而是在几百万、几十万、直至你认为合适的范围内排序（像上一节介绍的在 5,925 篇美国专利中排序），Patentics 的智能排序精度就会提高数倍。

r/cnxxxxxx and di/cnxxxxxx and b/abc

r/cnxxxxxx and di/cnxxxxxx and icl/h04L

r/cnxxxxxx and di/yyyymdd and ...

...

Patentics 有全套（扩充）的布尔检索系统支持各类检索需求。当你的检索获得结果，你可以在任意检索式后面加上排序，可以是一个词，一句话，一段文字，一个公开号（对应的全文），把自己感兴趣的主体文献排到最前面。

为了帮助审查员尽快找到合适的对比文件。我们下一个主攻方向有 2 个。一是要帮助大家快速浏览前 n 篇文献，尽快确定是否有对比文件在里面。刚刚上线的图文浏览界面 [www.patentics.com/scn.htm](http://www.patentics.com/scn.htm)，就是我们的一个努力成果。二是通过新的算法自动比对多元技术要素，快速分析、发现可能的对比文件。

对于审查员来说，应该尽量充分利用 Patentics 已经开发的资源。如对全球 5000 万中、英文全文的机器标引，在“索引”中按技术主题列出。对于中国申请，我们有近 1000 万中国申请的全文英文翻译，提供英文机器标引的“索引”词。对中国申请的中文版的主权项，我们进一步根据技术特征进行按主题的机器标引和聚类。当然，我们还发明了中英文检索透镜 (CN201310274941.X)，来进一步提高排序精度。大家可以阅读微信“Patentics 搜索透镜(1)”，理解不同的颜色对应不同的检索透镜作用最终对应不同的命中率。

这样，不需任何检索策略仅输入一个公开号或申请号，Patentics 的检索精度，可以通过自动分析来统计、预测。大家可以多花点心思，记下 luck 模式下，Patentics 找出的 X 文献，看看与我们上面的统计测试结果是否相符？

# 机器翻译+机器标引+机器语义排序 = EPO

## 特殊的检索工具或检索方式

Patentics 具有高精度基于对文本意思自动理解的机器标引系统。该机器标引系统不仅能对中国专利全文（中文）进行自动标引，而且还能对美国、EP、WO、JP、KR 等进行英文自动机器标引。此外，我们还将 1000 万篇中国申请全部采用基于语义的机器翻译成英文，对于这些从中文机器翻译成英文，再对这些中国专利英文全文进行机器标引的准确度如何，我们通过实例来验证。

EPO 做出的检索报告是国际上公认的检索结果相关度高的检索报告。为了缩小差距，研究 EPO 审查员使用的检索策略，审查员们会对 EPO 案例进行检索过程的分解，根据 EPO 的检索结果研究其使用的分类号、关键词、检索式、数据库等内容。

选取聚氨酯领域的一个 EPO 案例，通过对技术方案的分析，选择合适的分类号和关键词，对欧洲审查员的检索策略进行了分析和推测，以期对本领域的检索提供指导。

[案例要点] 关键词的积累扩展并注意不同语言中表达方式的异同

[案情介绍]

申请号：200980149754.0

发明名称：一种异氰酸酯三聚催化剂体系，前体制剂，使异氰酸酯三聚的方法，由其制备的硬质多异氰脲酸酯/聚氨酯泡沫体，和制备这种泡沫体的方法

权利要求：一种三聚催化剂体系，其包括：

磷阳离子；和

诱导异氰酸酯三聚物的阴离子；

其中所述三聚催化剂体系的三聚活化温度为小于或等于 73℃。

[IPC 分类] IC C08G18/02 IC C08G18/09 IC C08G18/16 IC C08G18/18

[检索要素]

确定检索要素时主要应注意对磷阳离子在关键词上的扩展。活化温度虽然是本发明的改进之处，但除非现有技术中使用与本申请同样的描述，否则对于活化温度在检索中并不容易提炼出合适的关键词。本案为 PCT 申请，在对中文库进行检索时没有得到合适的对比文件，也没有得到欧专局给出的国际检索报告中的 X 类文件 US4602080A 和 US3876618A。通过对外文库的检索发现，异氰酸酯三聚之后形成六元环，所以也是一个环化的过程，而 US4602080A 的摘要中正是记载了“ring”，却没有记载三聚，所以以“ring”为关键词时才能在外文摘要库中检索到。对于常用的关键词，US4602080A 在全文中使用但不完全且精准，所以即使在外文全文库也不一定能保证检索到上述两篇 X 类文件。推测欧专局可能具有特殊的检索工具或检索方式，所以获得了这两篇文件。

检索要素表如下：

检索要素 表达形式		异氰酸酯 化合物	三聚	磷阳离子	活化 温度
分 类号	I PC	C08G18/02 、 C08G18/09 、 C08G18/16 、 C08G18/18			
关 键词	中 文	异氰酸、聚 氨酯、异氰脲酸	三聚	磷、磷三 轮、阳离子、正 离子	活化 温度、低 温
	英 文	polyureth ane, isocyanate, urethane, isocyanurate	cycl ic, cyclo, trimer, ring	phosphon ium, cation	acti ve temperat ure

[常规检索过程]

1 CNABS 334 (陶氏 or DOW)/pa and (菲利普 or 阿西 or 纳森 or 内森 or 威尔莫特 or 维尔莫特 or 理查德 or 基顿 or 戴维 or 巴布 or 塞西尔 or 波耶 or 博耶 or 蒂莫西 or 莫利)/in

2 CNABS 3827 ((异氰酸 or 聚氨酯 or (C08G18/02 or C08G18/09 or C08G18/16 or C08G18/18)/ic) and 三聚) or 异氰脲酸

3 CNABS 1709 磷 or (磷 s 三轮)

4 CNABS 868 诱导 and (阴离子 or 负离子 or 羧酸根 or 甲酸根 or 乙酸根 or 辛酸根 or (乙 s 己酸根) or 苯甲酸根 or 碳酸根 or ((酚 or 氨化 or 脘化 or 亚胺 or 磷 or 氟) s (阴离子 or 负离子)) or 氰酸)

5 CNABS 20 3 and (阳离子 or 正离子) and ((活化 s 温) or 低温)

6 CNABS 61 (1 and 2) or (2 and 3) or (3 and 4)

1 CNTXT 25 (((异氰酸 or 聚氨酯) and 三聚) or 异氰脲酸) and ((磷 or (磷 s 三轮)) p (阳离子 or 正离子)) and 诱导 and (阴离子 or 负离子 or 羧酸根 or 甲酸根 or 乙酸根 or 辛酸根 or (乙 s 己酸根) or 苯甲酸根 or 碳酸根 or ((酚 or 氨化 or 脘化 or 亚胺 or 磷 or 氟) s (阴离子 or 负离子)) or 氰酸) and ((活化 s 温) or 低温)

1 DWPI 292181 +isocyanate+ or +urethane or +isocyanurate or (C08G18/02 or C08G18/09 or C08G18/16 or C08G18/18)/ic

2 DWPI 578 (+isocyanurate s ring?) or +tirmet+

3 DWPI 3544151 +cation or +phosponium+

4 DWPI 35497 activ+ s temperature

5 DWPI 59 1 and 2 and (3 or 4)

6 DWPI 1 5 and us4602080/pn

1 USTXT 0 us3876618/pn

[案例启示]

- (1) 关键词与分类号仍是主要的检索入口，对于来自欧洲的申请应注意利用 EC 分类体系进行检索；
- (2) 平时应注意积累关键词在不同语言环境下的扩展方式，而且在检索过程中发现检索要素新的表达方式后，应随时调整检索策略。

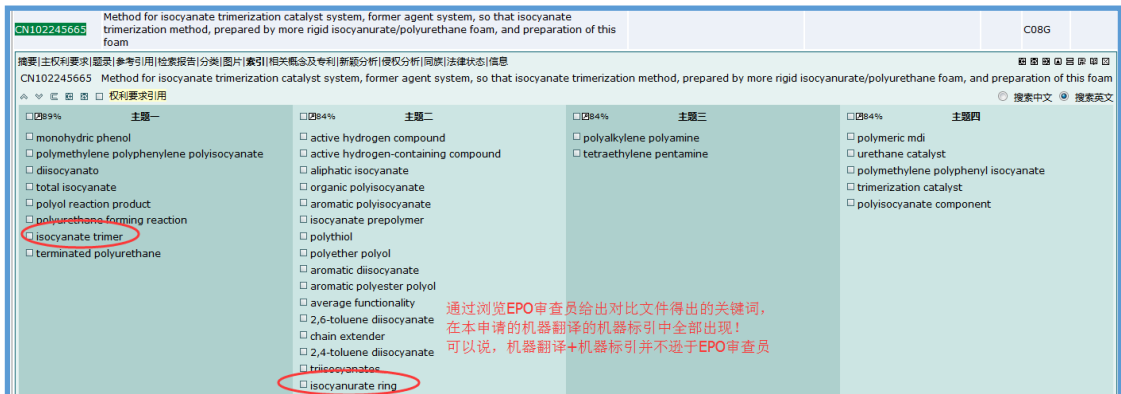
注意，关键技术概念“环化”是“通过对外文库的检索发现，异氰酸酯三聚之后形成六元环，所以也是一个环化的过程，而 US4602080A 的摘要中正是记载了“ring”，却没有记载三聚，所以以“ring”为关键词时才能在外文摘要库中检到”。所以审查员在进行中文检索策略构造时，环化概念都没有考虑。一直要浏览 EPO 给出的对比文件 US4602080A，才能以“+isocyanurate s ring?”为检索要素。

令人吃惊的是，这些关键技术概念在中文申请 200980149754.0 的英文全文的机器标引中全部出现。就是说，如果中国专利局审查员点击“英文复制按钮”，浏览该中国申请的英文机器标引“索引”，借助 Patentics 语义排序，也可以像 EPO 审查员一样轻而易举找到那 2 个被“推测欧专局可能具有特殊的检索工具或检索方式，所以获得了这两篇文件”。

2 个点击，快速获取机器翻译+机器标引的英文关键词。



而这些通过浏览 EPO 审查员给出的对比文件才得出的关键词，在本申请的机器翻译的机器标引中全部出现。可以说，机器翻译+机器标引并不逊于 EPO 审查员。



下面，就用该 2 个关键词限定检索范围，再通过语义排序将与 CN200980149754.0 最相关的对比文献排到最前面。

r/cn200980149754.0 and di/cn200980149754.0 and b/"isocyanate trimer"

公开号	标题	申请人	发明人	CPC	国际分类	相关度
6,515,125	Liquid partially trimerized polyisocyanates based on toluene diisocyanate and diphenylmethane diisocyanate	拜尔	Slack; William E.   Kemp, II; Hersel T.   Yonek; Kenneth P.	C08G	C08G	93%
5,104,930	Polyurea gel compositions and articles therefrom	瑞倪	Rinda; James A.   Marcer; Frank W.   Dawes; Keith   Bukownik; Rudolf R.	C08G	C08K	93%
7,595,396	Process for preparing polyisocyanates containing iminoxadiazinedione groups	拜尔	Richter; Frank	C08G	C07D	93%
2006/0079694	Process for preparing polyisocyanates containing iminoxadiazinedione groups	拜尔	Richter; Frank	C08G	C07D	93%
4,780,485	Isocyanurate foam and method for producing the same	Fischer; Harry A.	Ashida; Kaneoyoshi	C08G	C08G	93%
4,607,103	Methods for making blocked isocyanurates	通用电气	Zamek; Otto S.   Stein; Richard J.	C07D	C07D	93%
3,789,053	OXAZOLIDINONE MODIFIED TRIGLYCIDYL ETHER OF TRIHYDROXY TRIPHENYL METHANE AND DERIVATIVES THEREOF	陶氏化学	Clarke; James A.	C08G	C07D	93%
4,732,945	Synthesis by reacting oxirane resin with carboxyl compounds	阿什兰石油	Dammann; Laurence G.	C08F	C08F	92%
4,826,617	Selective resin synthesis and dual functional resins made thereby	阿什兰石油	Dammann; Laurence G.	C08F	C08F	92%
5,266,630	Aqueous coating composition containing a functional organic compound, a curing agent, and a dispersing agent	阿弗苏诺贝尔	Noomen; Arle   Peters; Petrus J.	C09D	C08G	92%
2004/0106762	Method for the preparation of low-viscosity (poly) isocyanates	罗狄亚公司	Chamiere, Eugenie   Bernard, Jean-Marie   Revelant, Denis   Randu, Stephane   Carlo, Michel		C08G	92%
3,676,397	OXAZOLIDINONE-CONTAINING EPOXY RESINS AND PROCESS FOR THEIR PREPARATION	陶氏化学	Clarke; James A.	C08G	C08G	92%
3,876,618	Oxazolidinone modified triglycidyl ether of trihydroxy triphenyl methane and derivatives thereof	陶氏化学	Clarke; James A.	C08G	C08G	91%
4,033,908	Polyisocyanurate compositions and foams of improved friability and process of preparing same	General Latex and Chemical Corporation	Hopkins, Jr.; Henry S.   Vargo; Donald L.	C08G	C08G	91%
3,926,909	Dibenzyl hydroxyl amine stabilizer for spandex	Globe Manufacturing Co.	Wei; Peter	C08K	C08K	91%
2009/0023823	POLYOLS AND POLYURETHANES, POLYURETHANEFORMS USING THE SAME	现有方法不能找到的文件就在第32位	Kim; Do Gyun   Kim; Hyosung	C08G	C08I	91%
5,859,135	Aqueous coating compositions comprising functional group-containing crosslinkable resins	阿弗苏诺贝尔	Doomen; Willy Frans Anna   van Wingerde; Mario Martinus Maria	C08G	C08G	91%
6,111,048	Blocked polyisocyanate and uses thereof	旭化成	Asahina; Yoshiyuki   Usui; Taketoshi	C08G	C08G	90%
2009/0053421	Non-Aqueous, Liquid Coating Compositions	纳昂尔杜邦	Flosbach; Carmen   Doessel; Karl-Friedrich   Becker; Wiebke	C08G	B05D	90%
4,433,680	Polyurethane casting material	美国强生	Yoon; Hee K.	C08G	A61L	90%

r/cn200980149754.0 and di/cn200980149754.0 and ac1m/"isocyanurate ring"

公开号	标题	申请人	发明人	CPC	国际分类	相关度
4,647,623	Process for producing polyisocyanate and resin composition comprising said polyisocyanate for use in paints	大日本油墨化学	Kase; Mitsuo   Okoshi; Noboru   Tsuyuzaki; Kazuo	C08G	C08G	88%
5,352,757	Sulfur compounds and new polymers obtained from these sulfur compounds	罗纳普朗克	Lavaurl; Sylvie   Valleret; Gerard	G02B	C07C	87%
4,602,080	Organic salts of polyisocyanate derivatives and the preparation thereof	Societe Carbochimique en abrege "CARBOCHIM"	Lambert; Pierre M.   De Aguirre-Otegui; Ionaco	C08G	C08G	87%
4,431,758	Heat resistant resin composition comprising reaction product of polyamideimide resin, alcohol and acid component.	日立	Osada; Yuichi   Kasai; Shozo   Okada; Yasunori   Uchigasaki; Isao   Oshima; Toyoji   Mukoyama; Yoshiyuki   Nishizawa; Hiroshi	C08G	C08G	86%
5,292,829	Isocyanurate ring-containing crosslinked polyurethane polyurea particulate polymer and process for producing the same	大日本油墨化学	Konno; Hidetoshi   Tamamoto; Yoichi   Muramatsu; Ichiro   Kase; Mitsuo   Okoshi; Noboru	C08G	C08G	86%
4,448,844	Heat resistant resin and process for producing the same	日立	Osada; Yuichi   Kasai; Shozo   Okada; Yasunori   Uchigasaki; Isao   Oshima; Toyoji   Mukoyama; Yoshiyuki   Nishizawa; Hiroshi	C08G	C08G	86%
5,252,623	Expandable vinyl chloride resin composition	三菱   三菱	Sugino; Satoru   Hotta; Hisaharu	C08I	C08G	86%
6,316,542	Crosslinkable dielectric composition	阿托菲纳公司	Berger; Noelle   Jay; Pierre	H01B	C08G	86%
4,138,372	Resin compositions containing polyepoxides and uretdiones	日立	Nishikawa; Aiko   Yokono; Hitoshi   Simizu; Ryuichi   Mukai; Junji	C08G	C08G	86%
4,212,917	Polyisocyanurate foam laminate	The Celotex Corporation	Skowronski; Michael J.   DeLeon; Alberto	B32B	B32B	86%
5,055,514	Paint resin composition	大日本油墨化学	Kumada; Hajime   Shoji; Akio	C08G	C08G	85%
5,021,534	Thermosetting urethane elastomer	日本聚氨酯工业株式会社	Kawasaki; Masahiro   Ohbuchi; Yukio   Maeda; Yoshiaki   Sato; Susumu	C08G	C08G	84%
5,719,227	Coating composition	阿姆斯特丹世界工业公司	Rosenberry; Angela S.   Rupp; Claude R.   Sethachayanon; Songvit	C08G	C08G	84%
4,614,782	Heat resistant resin composition	日立	Nishizawa; Hiroshi   Kasai; Shozo   Okada; Yasunori   Osada; Yuichi   Mukoyama; Yoshiyuki	H01B	C08G	84%
4,427,822	Process for producing particulate polymer having imide groups	日立	Nishizawa; Hiroshi   Mukoyama; Yoshiyuki   Hirai; Osamu	C08G	C08G	83%
2003/0149222	Thermosetting polyimide resin composition and process for producing polyimide resin	大日本油墨化学	Ichinose; Eiju   Yamashina; Yochoh   Ishikawa; Hidenobu	C08G	C08G	83%
6,887,967	Thermosetting polyimide resin composition process for producing polyimide resin, and polyimide resin	大日本油墨化学	Ichinose; Eiju   Yamashina; Yochoh   Ishikawa; Hidenobu	C08G	C08G	83%
5,594,049	Cationic electrodeposition coating composition	日本油漆株式会社	Yamada; Mitsuo   Kobata; Yusaku	C08G	C08G	83%
4,632,878	Magnetic recording medium	富士胶片	Okita; Tsutomu   Hashimoto; Hiroshi   Hashimoto; Akiyoshi   Nakagawa; Tetsuo   Kato; Hiroyuki	C08G	C08G	83%
4,304,818	Insulation system for winding of electric rotating machines and process of production thereof	日立	Hirata; Hideyo   Hakamada; Takeshi   Souma; Misao   Matsui; Masao   Suda; Yoshiyuki   Karasawa; Yoshiharu   Narahara; Toshikazu	C08G		83%

当然，我们可以和大家透个底，其实 EPO 也在数个场合被我们的语义检索折服，也在和我们接洽应用我们的语义检索技术！只是我们现在忙着“race to the top of the world”，还顾不上茬。