

## Patentics 与世界专利审查员检索报告自动比对测试

本测试报告对 Patentics 直接通过输入专利号所得搜索结果, 与各国专利局通过 PCT 提交的国际检索报告中引用的 X、Y、A 专利文献进行自动比对。测试结果报告如下

- 中国专利局专利检索报告 (中文) 自动比对测试 (简介);
- 欧洲专利局专利检索报告 (英文) 自动比对测试 (简介);
- 美国专利局专利检索报告 (英文) 自动比对测试 (简介);
- 日本专利局专利检索报告 (英文) 自动比对测试 (简介);
- 南朝鲜专利局专利检索报告 (英文) 自动比对测试 (简介);
- 中国专利局专利检索报告 (英文) 自动比对测试 (简介);
- 中国专利局专利检索报告 (中文 --> 英文) 自动比对测试 (简介);
- 美国专利发明人引用文献自动比对测试 (简介);
- 中国专利发明人引用文献自动比对测试 (简介);
- 世界原创性专利发明人引用文献自动比对测试 (简介);
- 美国专利诉讼案例专利发明人引用文献自动比对测试 (简介)。

Patentics 数学模型库为根据美国专利/申请全文至(1971-2009/10/8)共 6,176,828 篇, 中国专利申请全文至(1985-2009/10/7)共 2,869,697 篇构建。

在没有任何限制, 没有任何检索策略, 没有任何人工干预, 直接用专利公开号进行自动检索, 并与世界专利审查员检索报告进行自动比对, 本身在世界专利史是首次尝试, 也是在世界通用搜索技术领域首开先河, 难免有考虑不周之处, 敬请提出宝贵意见。

详细报告将在欧洲专利局举办的 2009 年专利信息年会上公布, 同时在 [patentics.com](http://patentics.com) 上公开, 敬请关注。

# 中国专利局专利检索报告（中文）自动比对测试（简介）

## I.

本测试部分对 Patentics 直接通过输入专利号所得搜索结果，与中国专利局于 2004-2009 年，通过 PCT 提交的国际检索报告中引用的中文（CN）X、Y、A 专利文献进行自动比对。

### 1. 测试数据采样

获取全部中国专利局于 2004-09 年通过 PCT 提交的专利文献及其国际专利检索报告，总计 15,522 篇<sup>1</sup>，以作对比。

### 2. 自动测试过程设计

### 3. 实测数据

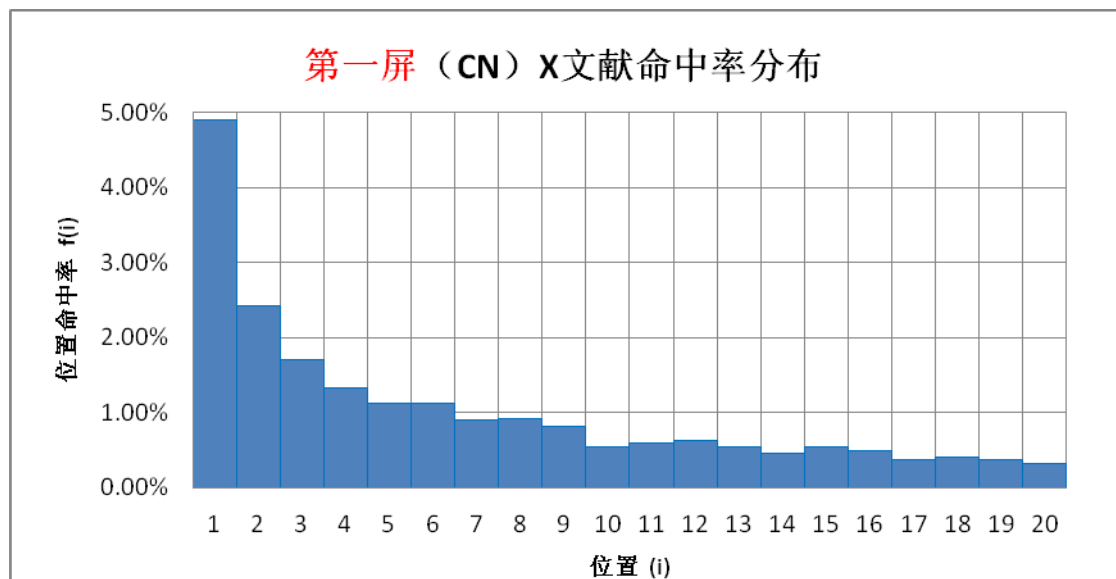
例：

778	该篇专利在样本中的序号
PN - WO2004096205 A1 20041111	该篇专利 WO 专利号及申请日期
FM - CN1541646 A	该篇专利同族专利号
IC - A23K1/16; A23K1/175; A23L1/304; A23L1/305; A61K31/195;	该篇专利 IPC 分类号
CT - CN1338254 [X] - 0 (99)	审查员引用的 CN X 文献 CN1338254
Patentics 所得结果与审查员检索报告中引用的专利文献对比	CN1290488 [X] - 2 (98) 在 Patentics 第一位命中，且与该篇专利的相关度为 99%。
	CN1333655 [A] - 334 (77)

测试报告中没有命中的都以 400 标注，更多说明可参见[实测数据说明](#)。

### 4. 测试数据分析

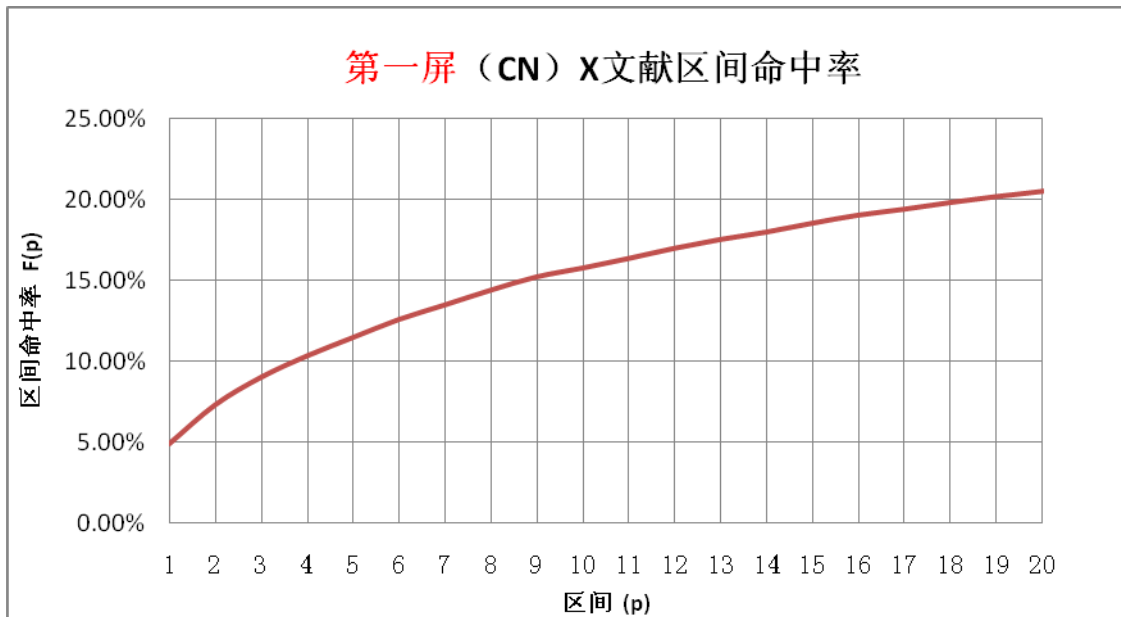
该测试数据包括审查员在 15,522 篇测试样本中，共 4,796 篇国际专利检索报告中引用 6,791 个（CN）X 文献：



$$f(i) = h_i / 6791$$

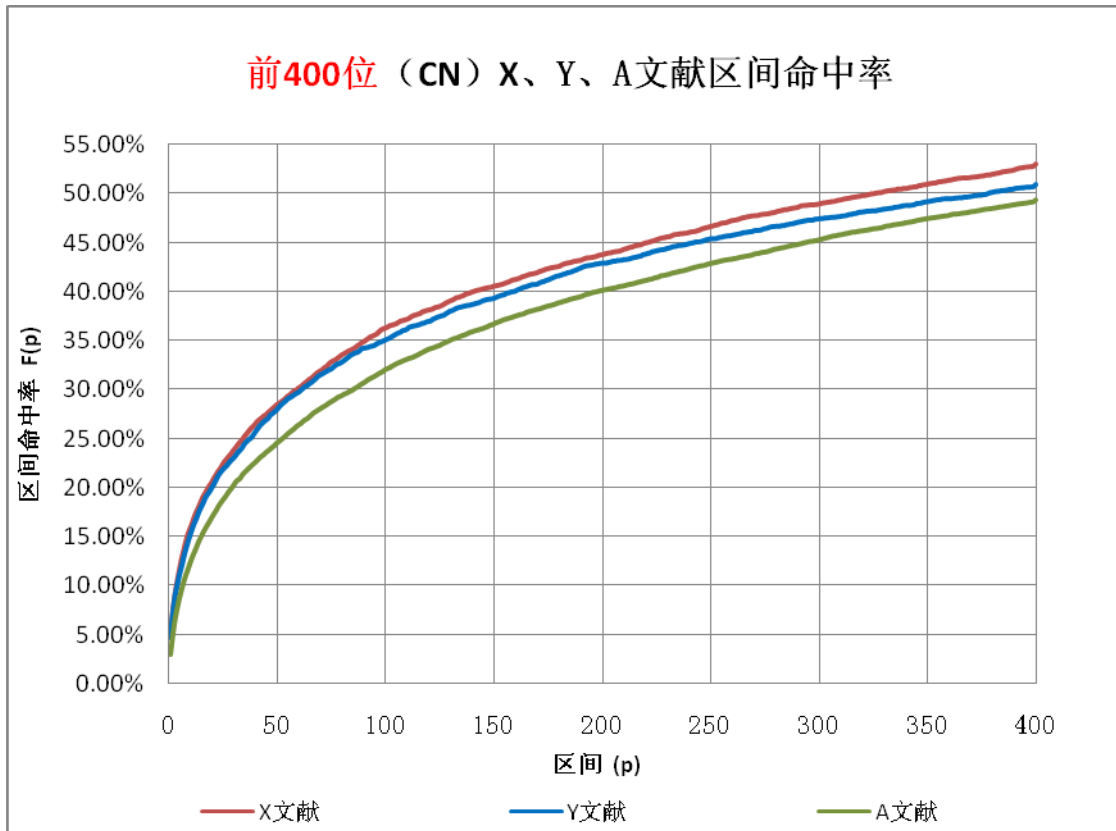
X 文献最可能的命中位置为 Patentics 结果的第一位，命中率已接近 5%，即，仅输入一专利公开号，在 Patentics 结果第一位就有近 5% 的可能性命中审查员引用的 X 文献。

<sup>1</sup> 样本文件 wo2004-09



$$F(p) = \sum_{i=1}^p h_i / 6791$$

第一屏（即，前 20 位），Patentics 计算命中审查员引用 X 文献的概率已达到 F(20)=20%以上。即，仅输入一专利公开号，在 Patentics 第一屏（前 20 位）结果中，有 20%以上的可能性命中审查员引用的 X 文献。



$$F(p) = \sum_{i=1}^p h_i / 6791 \quad F(p) = \sum_{i=1}^p h_i / 5563 \quad F(p) = \sum_{i=1}^p h_i / 18523$$

**Patentics 系统呈现出非常强的趋势：X 文献命中率高于 Y、A 文献。**

当审查员引用对比文献时，通过人工阅读、理解，已经对该文献的相关度作了评判，X 为最相关、Y、A 为依次次之。从实验统计中可发现，Patentics 通过算法，已能模拟人工阅读、理解能力，准确地将人工判定的 X(最高)、Y(次高)、A(最低)的相关度关系反映出来。

附表 按 IPC 分类统计 Patentics 前 400 位 (CN) X 文献命中情况

IPC 分类	命中篇数	总篇数	命中率
A	353	597	59.1290
B	233	360	64.7222
C	204	342	59.6491
D	36	46	78.2609
E	68	99	68.6869
F	183	278	65.8273
G	287	481	59.6674
H	1464	2593	56.4597

从命中率按 IPC 分类可见, Patentics 对中国专利文献中各学科知识的机器自动理解已趋于各向同性。