申请人名称标准化研究报告

国际局编拟

## 导言和背景

1. 世界各地的工业产权局（IPO）所产生的工业产权（IP）信息量和文献量正在稳定增长。与之相伴的是工业产权信息数据库所采用的技术迅猛发展。如果不对信息进行充分的标准化处理，专利信息用户进行很多工作将愈加困难，如组合评估、数据统计分析。
2. 经过标准化处理，关于申请人和当前工业产权所有人的准确可靠的信息将为统计分析提供显著的便利，并帮助公司作出战略性决策，如识别竞争者和潜在的商业伙伴，或评估某个公司的工业产权组合。因此，该问题对于整个工业产权界，尤其是工业产权信息用户和申请人至关重要。
3. 在名称标准化的框架内，利益攸关方讨论了一系列范围广泛的问题：
* 技术问题，如在公布的申请中更正印刷错误，或与翻译或音译名称有关的问题；
* 程序方面，如希望每个专利族只有一个受让人，或需要对工业产权转让进行登记；
* 法律考量，如由于寻求保护所在地区国内立法的某些规定而使用公司名称的“不同表示方式”。
1. 各相关方，尤其是工业产权局和工业产权信息提供商，为解决上述若干问题而投入了大量资源。例如，韩国知识产权局（KIPO）根据社保号或法人的国家注册，分别对自然人和法人实行申请人代码制度；汤森路透为申请提交量超过500件的申请人分配唯一代码。另一种方法是欧洲专利局（EPO）和经济合作与发展组织（OECD）合作开发的能够发现可能的不一致之处并提出名称“标准化”版本建议的软件。
2. 在若干地区和国际论坛，各利益攸关方尝试对这些做法进行简化，使其更为高效。申请人名称的标准化是工业产权五局以及其他工业产权局议程中的一项；它在工业产权信息用户的各会议上得到讨论。
3. WIPO对工业产权文献中申请人名称及其表示方式的标准化工作在WIPO标准委员会（CWS）前任机构的框架内开展。通过开展这些工作，WIPO成员国通过了WIPO标准ST.20，这项标准推行在名称索引中采取统一的名称表示方式，以及工业产权局以统一的方法整理索引中的名称。尽管ST.20的当前版本生效于1993年12月，但标准中提出的一些建议仍具有相关性。但是，由于该标准主要侧重于当时普遍且广泛使用但现在已非如此的印刷体名称索引，因此某些条款和参考文献已过时，可对其作出修改。
4. WIPO国际局还在PCT体系和马德里体系相关统计报告中实施了申请人名称标准化这一程序。所采用的方法依据的是计算各申请人名称和地址的相似度，并同时对申请量最大的各申请人进行关键词检索。
5. 2016年9月，WIPO国际局组织了申请人名称标准化讲习班。讲习班的主要目标是明确申请人名称标准化工作的优先事项，并确定WIPO在这项工作中能够发挥怎样的作用以协助工业产权局和其他利益攸关方。

## 范围和目标

1. 本研究报告总结了工业产权文献中申请人名称标准化的状况，尤其是侧重于遇到的问题和到目前为止识别的解决方案；目标是调查WIPO在这一过程中能够发挥的作用。
2. 本研究报告的范围仅限于包含在专利申请中的名称，不包括其他类型的工业产权。但是，适用于专利申请的名称表示方式方面的结论可能同样适用于商标和工业品外观设计。

## 参考文献

1. WIPO标准ST.20“专利文献名称索引编制建议”
2. 申请人名称标准化WIPO标准讲习班材料，日内瓦，2016年9月5日
3. PDG IMPACT工作队关于申请人名称的材料
4. 统一专利统计数据的数据生成方法：专利权人名称的统一，欧盟统计局，欧盟委员会，2006年
5. 欧盟委员会专利统计数据：地区化的方法、部门分配和名称的统一，欧盟统计局，2011年
6. 第4轮讨论报告“有意义的专利可视化处理和名称的统一——现行做法和趋势”，欧洲专利局专利信息会议，2015年
7. 韩国知识产权局，“申请人名称的标准化”，全球专利卷宗第3工作队（GDTF）会议，2016年[演示文稿]
8. 韩国知识产权协会（KINPA），“标准化的申请人名称”，全球专利卷宗第3工作队（GDTF）会议，2016年[演示文稿]
9. 美国知识产权法律协会和知识产权所有人协会，“仅限产业界的全球专利卷宗第3工作队总结”，2016年[演示文稿]
10. Andrey Sekretov，欧亚专利局，“欧亚专利局名称处理的特定方面”，东方遇见西方，2016年[演示文稿]
11. Stephen Adams，马吉斯特有限公司，“名称中包含着什么？”国际专利信息会展——大师班，罗马，2015年[演示文稿]
12. Julie Callaert，研究与发展监测中心——鲁汶大学，“申请人名称的统一以实现更好的检索结果”，欧洲专利局专利信息会议，2015年[演示文稿]

## 利益攸关方及其面临的困难

### 专利信息用户

1. 产业界工业产权部门的很多工作需要对海量数据进行分析。申请人和所有权信息对于这些工作的重要性如下：
* 自由实施（FTO）检索——所有权信息对于确定与拥有“阻碍性”专利的公司的关系（开始许可谈判、评估风险或准备诉讼）十分重要。
* 公司分析——掌握完整的所有权信息对于分析一个（专利数量可能达上千件）公司的专利组合至关重要。
* 组合对标——将公司的专利组合（或部分专利组合）与该公司竞争对手可能涵盖上千个专利家族的专利组合进行比较。所有权信息对于确定组合的范围至关重要。
* 技术态势——确立技术态势的目的是为了了解专利活动“看起来怎么样”以及谁在某个技术领域较为活跃。从申请人信息中可以找到第二个问题的答案。
1. 专利信息用户的主要关注点是明确专利所有人。他们在尝试确定某个时间点某个专利的所有人的过程中面临很多困难，其中之一是申请人信息不明确。产业界希望工业产权文献中的申请人名称具备以下特点：可通过唯一标识进行识别、标准化和书写正确。
2. 以下因素会对专利分析的结果产生影响：
* 以**子公司**名义注册的专利通常不被计入母公司所拥有的专利数量
* 专利文献可能包含同一申请人的**不同名称**
* 在**一个专利家族中**可能**有不同的受让人名称**
* 一件国际申请中可能包含**多个申请人**（针对不同的指定国）
* 申请人名称可能有**拼写错误**
* **发明人名称**可能作为受让人名称出现

### 专利局

1. 工业产权局承认缺少申请人名称的标准化，并尝试找到解决方案，这将加强数据的可用性及其与其他数据源的连通性。
2. 很多工业产权局公布“所提交的”名称信息，因为国内立法的规定通常要求由申请人采取对著录数据进行更正这一行动，这可能涉及额外的费用。这种做法使工业产权局无法
* 高效地更正错误，尤其是申请人名称中的打字错误
* 使用同一申请人名称的一种拼写方式（“IBM”对“I.B.M.”）
* 一致地说明申请人名称的法律形式和其他“非重要部分”[[1]](#footnote-1)
* 一致地使用申请人名称的首字母缩略语
1. 另一个障碍是在以工业产权局的一种或几种工作语言表示外国名称方面，以及在为了国际数据交换的目的为载于工业产权局所公布申请的著录数据项目提供翻译方面没有建议。如果申请是通过工业产权保护地区或国际（如PCT）体系提交，情况甚至会更为复杂[[2]](#footnote-2)。
2. 工业产权局还面临一系列与申请人名称相关的技术困难，例如：
* 申请人名称过长，数据库字段或屏幕表格无法容纳
* 损坏的附加符号和其他特殊字符
* “看起来相同的”字符有不同的代码（如斯拉夫字母“P”的UTF‑8hex代码是0620，拉丁字母“P”的UTF‑8hex代码是0050）

### 专利信息提供商

1. 商业专利信息提供商为更新专利申请人和权利人信息投入了大量资源。它们的主要挑战是“原始”申请人名称信息质量低下。在处理的专利中有20%需要进行数据更正或审查。这相当于每周11,000件专利，需要人力知识附加价值。（更多信息见“现有解决方案/汤森路透”。）

### 总结和结论

1. 缺少申请人名称的标准化对于专利信息用户是存在已久的挑战。
2. 商业工业产权信息提供商为将申请人名称的标准化信息纳入它们的产品投入了大量资源；这项工作极具挑战，因为来自一次文献（由工业产权局公布）的信息并未经过充分的标准化处理。
3. 工业产权局承认专利信息用户面临的困难，但它们缺少资源、指导和法定权力来以统一的形式接受和公布申请人名称并“从源头上确保质量”。

## 现行做法和解决方案

1. 申请人名称的标准化过程有以下4个层面[[3]](#footnote-3)：
* 正规化——更正“细微”错误（同一申请人仍可能有多个名称）
* 统一——一个申请人使用一个名称（可能不是最终所有人，因为受益方为母公司时工业产权可以以子公司的名义注册）
* 通过公开（提交申请时的）公司结构，对申请人名称信息作出综合考量
* 对所有权进行准确、动态的记录
1. 标准化应达到哪个层面的问题比较复杂，取决于标准化工作的目标。
2. 对各利益攸关方所采取的做法解释如下。

### 工业产权五局的合作

1. 申请人名称的标准化是工业产权五局“全球专利卷宗”项目的优先事项之一；它旨在通过统一工业产权五局专利文献集中的申请人名称，更好地检索现有技术，改进专利文档管理。
2. 这项工作的目标是将同一申请人的多个名称统一为一个标准化条目，并提供一个工业产权五局文献中所列原始名称和标准化名称之间的对照表。
3. 最初的想法是为一组属于同一申请人的不同名称分配一个单一数字代码。计划利用句法距离、地址和电话号码信息来对名称进行分组。由于以下因素，工业产权五局并未建立这一代码系统：
* 一些申请人有句法上并不接近的不同名称
* 对地址和电话号码进行比较被证明是一项人工密集型工作
* 由于法律限制和来自产业界的反对，并非所有工业产权五局都接受实行申请人代码系统

#### AIPLA的观点

1. 美国知识产权法律协会（AIPLA）的代表要求工业产权局在将申请人名称改为标准化版本或分配标识符前与申请人进行磋商。

### 统一的名称

#### 为欧洲专利授权程序统一客户数据（NACM项目）

1. 欧洲专利局旨在为每个法人实体或个人创建一个单一的唯一标识，为拥有统一名称和地址的欧洲申请人建立一个主数据中心存储库（NACM项目）。为了实现这一目标，欧洲专利局正在为名称的标准化和重复名称的删除制定“清洗战略”；它还制定了质量标准（名称的句法分析）和数据质量政策。计划是申请人将对标准化的结果进行确认。

#### PATSTAT中的申请人名称信息

1. PATSTAT是欧洲专利局开发的用于对专利数据进行复杂统计分析的工具。PATSTAT纳入了若干标准化解决方案，并为用户提供以下选项：
* 原始名称
* DOCDB标准化名称（每年更新两次）
* HAN申请人名称（OECD）（每年更新两次）
* PATSTAT申请人和发明人统一名称（每年更新两次）

#### DOCDB中申请人名称的标准化（STAN）

1. DOCDB包含一个分配给公司、高校和其他法人实体的标准化申请人名称的字段。在DOCDB中分配标准化名称的过程分为两个步骤：根据专门开发的算法进行自动化分析，以及在结论不明确的情况下进行人工检查。
2. DOCDB目前包括约250,000个标准化名称，涵盖约945,000个不同的“所提交”的申请人名称。原始提交的申请人名称信息也被纳入DOCDB条目中。

#### OECD编制的统一申请人名称（HAN）

1. 经济合作与发展组织（OECD）利用专利指标来：
* 评估公司在某一经济部门的技术效绩
* 识别新兴技术
* 研究技术变革的知识传播和动态以及发明过程的地理特性
* 调查创新和公司动态
* 评估发明的经济价值
* 评估高校在技术发展中的作用
* 研究研发活动全球化进程
1. 为了开展这些分析，OECD需要将专利数据与拥有专利的公司的最新信息相关联。OECD为此建立了各国专门的申请人名称库（OECD HAN数据库），通过定制的算法Imalinker（Idener多算法链接器）将专利申请人名称与列于ORBIS©数据库的公司名称进行匹配。
* 按**国别**进行匹配
* 通过**国别名称库**对名称进行统一
* 使用**字符串匹配**算法（基于记号/字符串度量）
* **匹配分数的阈值非常高**，以便尽量减少假匹配或假不匹配
* 采取**手动控制**调整准确率[[4]](#footnote-4)

#### PATSTAT中的统一名称

1. PATSTAT中申请人（和发明人）名称的统一过程分为两个步骤；该程序由鲁汶大学开发，并作为欧盟委员会所开展的欧盟统计局项目的一部分提供。[5]
2. 第一步（在欧盟统计局的专有名词中称之为方法第一层级）的目标是对看似相似但由于拼写或语言的不同而有所区别的名称进行匹配；通过将每个名称与所有其他名称进行比较完成匹配。这一步骤完全为自动化操作，重点是准确性，必要时可忽略完整性。该阶段的流程如下：
* 数据预处理
* 字符清洗（转换为ASCII字符）
* 标点符合清洗（预解析）
* 名称清洗
	+ 法律形式名称处理（法律形式信息被统一并移至特别字段）
	+ 删除常见的公司词汇（如“company”或“corporation”）
	+ 统一不同的拼写方式（不同的拼写方式和此前发现的如“系统”等常见词汇的拼写错误被正确的单一拼写所代替）
	+ 压缩（删除所有非字母数字字符）
	+ 统一元音变音
1. 第二步（方法第二层级）通过更为细致地检查不同的名称表示方式进行进一步统一。检查基于的是近似字符串匹配、进一步手动验证和质量控制。该阶段最重要的因素是：
* 完整性
* 高准确性（“保守”规则）
* 专利数量方面的高覆盖率
1. 所显示的情况是检索出的专利数量超过10件的名称占专利文献中所包含名称总数的99.6%；因此第二步的检查范围仅限于这些名称（“前500个申请人”）。

### 申请人标识符

1. 韩国知识产权局开发了一个自1987年以来一直使用的申请人代码系统。1999年KIPOnet（韩国知识产权局信息技术自动化系统）启动时向已有的申请人分配了代码，此后韩国知识产权局一直在根据所注册的社保号或公司注册号为所有新申请人分配代码。
2. 韩国知识产权局的申请人代码由12位数字组成：DYYYYNNNNNNC，D是区分符（用于识别本国和外国公司、机构、本国和外国自然人、律师等），YYYY——年份，NNNNNN——序列号，C——校验位。
3. 2009年，韩国知识产权局对规则进行了修改，纳入了申请人应使用单一申请人名称和申请人代码的要求。实行申请人代码有以下好处：
* 申请人信息得到更为有效的管理
* 申请人不必重复输入信息
* 申请人名称和其他相关信息在同一申请人提交的所有申请中可同时变更
* 准确的工业产权统计数据[[5]](#footnote-5)

#### 汤森路透

1. 汤森路透运行一个包含超过180万个条目的受让人名称内部数据库。德温特世界专利索引（DWPI）处理的每个受让人名称都会对照该数据库进行检查，所识别的错误被修正，名称经过缩写和标准化处理。对于识别出的公司，专利受让人代码（PACO）会被添加至条目中。对于新公司，会为其分配一个新代码并纳入内部数据库。这些代码有助于区分名称相似的公司以及识别名称中不包含母公司名称的子公司[[6]](#footnote-6)。

### 总结和结论

1. 所描述的大部分解决方案处理的是专利申请人名称标准化的前两个层面：正规化和统一，而未触及所有权问题。
2. 申请人名称的标准化主要有3个方法：
* 分配申请人代码（韩国知识产权局）
* 在专利信息数据库中为正规化或统一的名称新增字段（欧洲专利局PATSTAT）
* 运行申请人数据库，其中每个申请人都有一条包含统一名称和地址的单独记录（欧洲专利局NACM）
1. 工业产权局留存正确的申请人记录并从源头确保质量的最为直接的方式看起来是分配申请人代码，但由于法律问题（应公布“所提交”的申请人数据）以及来自产业界的阻力（见AIPLA的观点），这种做法难以在一些国家辖区实施。注册法人实体和自然人的不同国家做法可能使在地区和国际层面实施该方法更为复杂。但看起来有用的做法是与各工业产权局分享该做法，以便它们在兼顾国内立法的同时能够找到适当、合理的方法来说服申请人尽可能以统一的方式说明名称。
2. 在专利信息数据库中为正规化或统一的名称新增一个字段需要来自工业产权局的大量资源，并且只有当不同的工业产权局适用相同的申请人名称正规化和统一规则时才对国际数据交换具有价值。如欧洲专利局等一些工业产权局已在这一领域获得了经验和专门知识，如果它们能与其他工业产权局分享这些经验和知识，这将使工业产权界受益匪浅。还可以使正规化规则标准化，并为申请人名称的统一提出建议。
3. 建立并运行申请人数据库与第一种方法（分配申请人代码）相类似，因为实际上等同于代码的数据库标识码将被用于把申请与申请人信息相关联。但是，由申请人本人对数据库中包含的信息进行确认可能有助于避免上述若干问题。这一方法也需要工业产权局的大量投资。替代方案可能是建立国际申请人名称数据库。
4. 汤森路透和其他商业工业产权信息提供商采取措施确立了专利权的最终受益方（确定公司结构），并向它们的客户提供这些信息（标准化第三层面和第四层面）。这是一项极为人工密集型的工作，通常超出工业产权局的权限。

## WIPO在申请人名称标准化中的作用

### WIPO标准委员会

1. 申请人名称标准化讲习班的结论之一是与会人员要求就申请人标识符在多大程度上有用、推行这种做法可能带来的问题以及工业产权局是否正在使用申请人名称库对工业产权局进行调查。在标准委员会的框架内建立新工作队后，可通过编制调查问卷并由秘书处向各工业产权局分发来开展这项工作。调查结果将公布于WIPO《工业产权信息与文献手册》（《WIPO手册》）。
2. 与此同时，工作队可以研究是否能够以WIPO标准的形式制定建议，以帮助工业产权局所发布的专利文献中的申请人名称实现更好的统一。这些建议可涵盖以下领域：
* 申请人名称的正规化
* 翻译/音译的问题
* 不同国家中的不同名称结构
* 为数据交换新增字段（XML标准、INID代码等）
* 及其他。

### 在工业产权局之间交流经验

1. 很多造成专利申请中申请人名称不统一的问题超出了WIPO标准化活动的主要范围，因为它们的根源是工业产权局实行的内部程序、法人实体注册等国家做法、不同地区以往使用的不同名称结构等。但为了使工业产权局加深其对于所面临问题的认识，简化现有做法并为希望学习其他局经验的工业产权局提供帮助，有用的做法是在各工业产权局和其他相关方之间交流观点和经验，分享与申请人名称统一相关的做法和专门知识。
2. 考虑到以下事件或活动可能有助于为工业产权局之间的经验交流提供便利，国际局愿意应请求协助举办这些活动：
* 关于名称标准化**特定方面的讲习班**能够帮助工业产权局识别需要对国内立法作出的必要调整，以便对原始数据进行“清理”，或为实施报告所有权变更这一要求提供便利。
* 有关名称正规化算法或其他申请人名称信息技术解决方案的**培训课程**将帮助工业产权局建立必要的基础设施，以改进其专利信息产品。
* 关于业务流程的**做法交流（圆桌讨论）**将有利于确保“从源头保证质量”。它们可涵盖一系列范围广泛的问题，例如申请人在公布前对（数据已更正的）首页进行确认，对申请表格进行修改以及在缴费时要求确认当前的所有人。

## 结　论

1. 申请人名称的标准化是一项关系到所有利益攸关方的重要、复杂的工作：（发展中国家和发达国家的）工业产权局、专利信息用户和申请人。
2. 应在国际层面以及地区和国家层面寻求申请人名称标准化的解决方案。可以制定一项**WIPO标准**以处理申请人名称标准化的某些方面，但对于很多所识别的问题，制定WIPO标准看似时机不成熟或是不适当。为了协助成员的工作，标准委员会可以就工业产权局**使用标识符和名称“数据库”的情况开展一项调查**，并在《WIPO手册》中公布结果。WIPO还可以通过组织关于申请人名称标准化具体问题的**讲习班、培训课程和圆桌讨论**协助成员国交流经验和做法。

[附件和文件完]

1. 见WIPO标准ST.20建议13（1993年12月） [↑](#footnote-ref-1)
2. Sekretov，Andrey“欧亚专利局处理名称的特定方面”，申请人名称标准化WIPO标准讲习班的演示文稿[2] [↑](#footnote-ref-2)
3. Adams，Steven“问题的根源——这为什么发生？”，申请人名称标准化WIPO标准讲习班的演示文稿[2] [↑](#footnote-ref-3)
4. Dernis，Helene“OECD HAN数据库：专利统计中申请人名称统一的解决方案”，申请人名称标准化WIPO标准讲习班的演示文稿[2] [↑](#footnote-ref-4)
5. Yoon，Jaewook“与申请人名称相关的困难以及韩国知识产权局的现行做法”，申请人名称标准化WIPO标准讲习班的演示文稿[2] [↑](#footnote-ref-5)
6. Hajkowski，Steven“德温特世界专利索引中的受让人信息”，申请人名称标准化WIPO标准讲习班的演示文稿[2] [↑](#footnote-ref-6)