Калибровка силы патента

Постановка задачи

# Дано:

## Исходные данные

Есть две таблицы: ‘PS\_Strength’ (красная) и ‘PO\_Strength’ (зеленая):

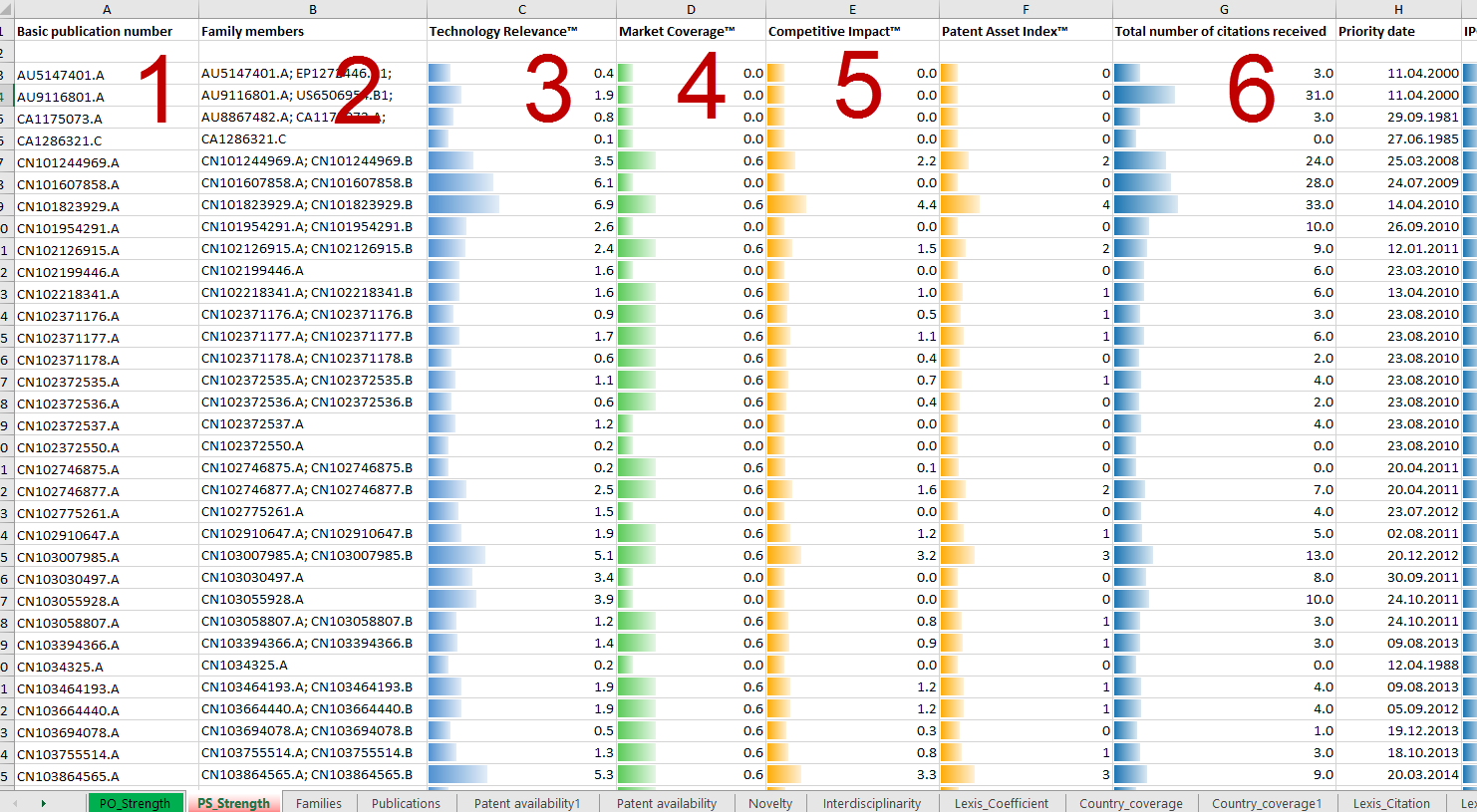


Таблица ‘PS Strength’ (красная)

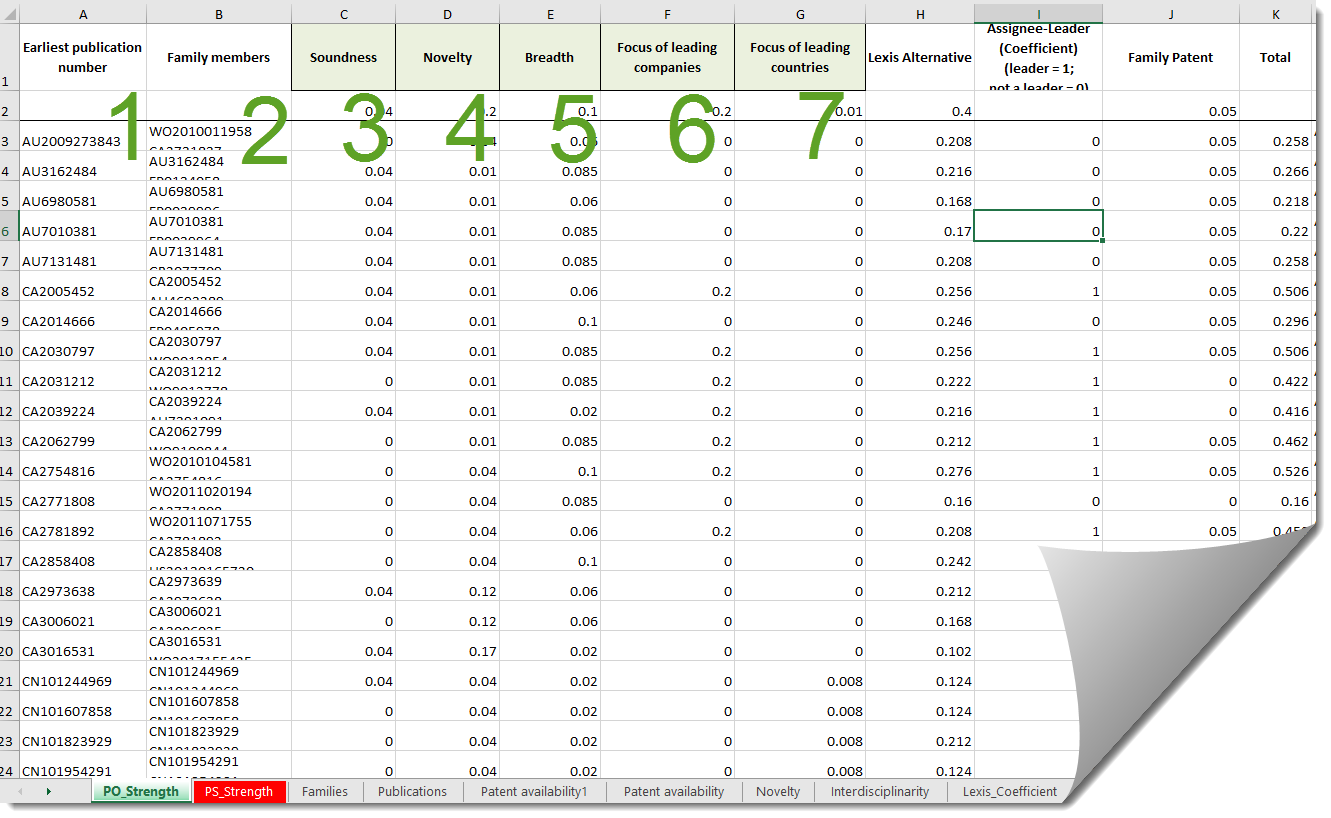


Таблица ‘PO Strength (зелёная)’

В каждой таблице – набор патентных семейств. Семейство представлено базовой публикацией (позиция 1 на рисунке) и одним или несколькими членами семейства (позиция 2).

В вырожденном случае в ‘Family members’ всего один документ, и он совпадает с базовой публикацией.

## Красная таблица

Красная таблица – эталонная. В ней для каждого семейства определено два индикатора: ‘technology relevance’ (позиция 3) и ‘market coverage’ (позиция 4).

Для каждого семейства есть интегральный показатель ‘competitive impact’, который как-то (они не раскрывают) сворачивает эти два показателя (3 и 4) и использует что-то еще для расчёта. Позиция 6 – общее число цитирований, которое почти всегда используется при расчёте ценности. Вполне возможно этот показатель учтен при расчете ‘technology relevance’.

Для справки. ‘market coverage’ больше апеллирует к ВВП стран. Если ВВП США больше чем ВВП России, то патент, полученный в США, имеет большую ценность, чем полученный в России.

## Зелёная таблица

Зелёная таблица – таблица с нашими показателями (которые мы можем подготовить). Сейчас таких показателей всего 5, позиции 3-7 на рисунке.

Мы в дальнейшем будем включать гораздо больше показателей (см. таблицу ниже, это кандидаты на features при расчете ценности патента). Таких показателей-кандидатов уже 50, мы прогоним их через серию мозговых штурмов, часть отвалится, какие-то новые могут добавиться.

Таблица – Показатели для расчёта ценности патентов

|  |  |  |  |
| --- | --- | --- | --- |
| № | Показатель | Комментарий к расчету | % |
|  | «Наличие хотя бы одного действующего патента в семействе» (soundness) |  |  |
|  | «Новизна» (novelty) |  |  |
|  | «Междисциплинарность» (breadth) |  |  |
|  | «Внимание стран-лидеров» |  |  |
|  | «Внимание компаний-лидеров» |  |  |
| Suggestions | | | |
|  | litigations |  |  |
|  | security interest (залог в банках) |  |  |
|  | forward citations |  | 45 |
|  | backward citations |  |  |
|  | crowdedness of space |  |  |
|  | independent claim count |  | 14 |
|  | claim 1 word count |  | 12 |
|  | prosecution time |  |  |
|  | patent age |  | 19 |
|  | family size and international filing |  | 10 |
|  | assignee quality (maturity) |  |  |
|  | geographic coverage |  |  |
|  | related applications |  |  |
|  | patent literature cites |  |  |
|  | well-defined patent |  |  |
|  | broadly scoped |  |  |
|  | potentially fundamental |  |  |
|  | significant offensive value |  |  |
|  | significant defensive value |  |  |
|  | well aligned with key products |  |  |
|  | well aligned with strategic goals and business interests |  |  |
|  | How many Continuations, Divisionals, CIPs? |  |  |
|  | How many times was the patent rejected before issue? Office actions? |  |  |
|  | Was the patent litigated, re-issued, re-examined? |  |  |
|  | What is the word length of claims, Claim count? (Breadth scope indicator) |  |  |
|  | Does the patent read on any standard (SEP) |  |  |
|  | ‘National priority’ takes the value 1 (0 otherwise) if the patent has first  been filed in a national patent office, and then transferred to EPO after the priority year |  |  |
|  | ‘EPO priority’ is equal to 1 if the applicant goes directly at EPO, without filing a national priority  application before |  |  |
|  | the PCT application which have been transferred to the EPO – should be of high value |  |  |
|  | Cross-border ownership |  |  |
|  | IP5 protection |  |  |
|  | Count of cited refs nonpatent literature |  |  |
|  | Наличие РСТ заявки |  |  |
|  | триадное семейство |  |  |
|  | совладения (межстрановое, межинституциональное) |  |  |
|  | наличие лицензий, восстановлений, свидетельств дополнительной охраны |  |  |
|  | срок рассмотрения |  |  |
|  | Citation Velocity |  |  |
|  | наличие представителя |  |  |
|  | число чертежей/таблиц |  |  |
|  | наличие немашинного перевода на язык неприоритетной стран |  |  |
|  | факт преобразования вида объекта на этапе делопроизводства |  |  |
|  | наличие корреспондирующего промобразца |  |  |
|  | число авторов |  |  |
|  | межстрановой авторский коллектив |  |  |

В общем случае в зеленой таблице будет какое-то число показателей ***n***, по которым мы будем собирать сведения для семейств.

Каждый показатель имеет или булевую оценку (есть / нет) или количественную оценку (например, «доля действующих патентов в определенном семействе»). Те значения, которые в таблице сейчас, будут пересмотрены; на них не смотри пока.

Все показатели будут в диапазоне от 0 до 1.

## Рассогласование семейств в таблицах

Данные в таблицах рассогласованы из-за разных принципов группирования патентных семейств.

Нужно согласовать семейства: ассоциированные красные и зеленые семейства должны иметь одну и ту же базовую публикацию.

Может быть несколько видов несоответствий:

1. базовые публикации одинаковые (позиции 1 в таблицах совпадают), но состав семейства разный – разное число членов и сами члены разные. В этом случае принимаем их сопоставленными, это одно семейство, несмотря на разный состав членов.

2. базовая публикация в красной таблице входит в состав членов семейства с отличающейся базовой публикацией в зеленой таблице. В этом случае нужно найти красную базовую публикацию в числе членов семейств зеленой, сделать ее базовой, а старую зеленую базовую публикацию включить в состав членов этого семейства.

3. красная базовая публикация не обнаруживается в зеленой таблице (ни среди базовых, ни среди членов). В этом случае исключаем это семейство из расчетов.

4. в зеленой таблице есть семейства, ни один из членов которой не найден в красных базовых публикациях. В этом случае ищем каждый член зеленого семейства в членах красных семейств. Если находится, ставим для него в качестве базовой зеленой красную базовую публикацию.

5. сложный случай: зеленая публикация найдена в нескольких красных семействах. В этом случае волюнтаристски ассоциируем зеленое семейство с любым из красных (например, самым большим, самым ранним и пр.).

6. совсем какие-то сложные случаи, с ними будем разбираться по ходу.

# Нужно сделать:

1. Выполнить сопоставление красных и зеленых семейств. Итогом работы должны быть две таблицы, в которых одинаковое число строк и одинаковый перечень базовых публикаций. Семейства, как красные, так и зеленые, для которых не удается выполнить сопоставление, удаляются из таблиц.  
   Результат можно свести в одну таблицу: {базовая публикация}; {красные показатели}; {зеленые показатели}. Семейства лучше отсортировать по убыванию ‘competitive index’ или по ‘basic publication number'.
2. Рассчитать начальное значение интегрального показателя ценности зеленых семейств. Сначала будут использованы только 5 имеющихся показателей (позиции 3-7 в зеленой таблице), затем по мере добавления новых показателей, они тоже д.б. включены в расчет интегрального показателя. Начальное значение интегрального показателя – все показатели имеют один и тот же вес, функция аддитивная, равнозначны для оценки ценности и пр. допущения.
3. Дальше хитрыми методами с ансамблями и деревьями нужно сделать так, чтобы интегральные показатели зеленых семейств были очень близки интегральному показателю ‘competitive index’ соответствующих красных семейств, т.е. решить задачу:

Φ → min δ(|νred – νgreen|)

Минимизировать модули разности интегральных показателей ценности красных и зеленых семейств, подобрав соответствующие веса зеленых показателей (сейчас 5, в перспективе 50).