Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

# Cross-nation cooperation mining from USPTO

M.H. Wang *, C.M. Huang, C.Y. Fan, P.H Lin, S.H. Chang

*Science and Technology Policy Research and Information Center, NARLabs, Taipei, Taiwan*

ARTICLE INFO

*Article history:*
Received 28 March 2016
Received in revised form
24 May 2016
Accepted 24 May 2016

*Keywords:*
IPR
USPTO
Simpson's Index

ABSTRACT

Patents are highly valued intellectual property and an observable indicator of new technology. In this study, we analyze the 2014 and 2015 USPTO patents and disassemble cross-nation inventor strength. Recently, problems that are unmanageable have emerged in different fields primarily because of the rapid and sudden occurrence of these problems. In this way, original knowledge and its protective framework can unfortunately and suddenly become obsolete. Therefore, frequent and repetitive references to the diverse latest safety devices are the best means of preventing a sudden outbreak of a crisis. Discovering each country's dominant inventors and potential industries is a fundamental way of resolving such crisis.



## 1. Introduction

Information from the United States Patent and Trademark Office (USPTO) shows that for new industries to have a decisive opportunity in the market, protection of the intellectual property (IP) in their product range is crucial. Through open market information, it is possible to not only predict the launch of products from each company but also discover the competitiveness of each of their scientific advances and technologies. The United States (US) is one of the principal consumer markets in the world and is the most competitive, and its companies have the highest invention rate. To understand each country's competitiveness, we can quickly obtain information concerning strong industries and already existing partnerships in the market from the USPTO using assignee and inventor relationships.

## 2. Related works

Previous IP discovery work can be categorized as summarizing the tendencies of IP applications in measuring patent quality (Park and Yoon, 2014) and government–university–industry cooperation (Srihari and Mathew, 2014; Lee et al., 2014). In the field of text mining, some studies have analyzed patent classification and content (Qu et al., 2014) and some have constructed topic models and existing technology aggregations (Hu et al., 2014). Using statistics, such work can review past events and discover hidden connections (Oh et al., 2014). Using the available information, this study utilizes invention as a connection to cross-nation mining.

## 3. Data source and methods

The data source was the USPTO patent xml (USPTO) posted openly on the Internet. After identifying each xml DTD version (United States Patent and Trademark Office DTD), we extracted the xml into a single database. There are six types of DTDs (Table 1). Although the tag name or element value might be different in each version, the major elements are the same. The patent grant comprises six DOM elements: us-bibliographic-data-grant, abstract, description, drawings, claims, and claim-statement. The most complicated element is us-bibliographic-data-grant, which includes basic patent data, e.g., patent number, date, title, inventor, assignee, classification, citation, relation information, figure, examiner, agent, and applicant (Fig. 1).

After extracting the xml, we calculated the relationship between every assignee and their foreigner inventors. To measure the diversity, we used Simpson's Index, which was introduced in 1949 by Edward H. Simpson to measure degrees of concentration when individuals are classified into groups (Simpson, 1949).

## 4. Equations

n = the total number of patents from country B in country A (foreigner relation)

N = the total number of patents in country A (foreigner relation)

$$D = \sum(\frac{n}{N})^2 \quad (1)$$

* Corresponding Author.
Email Address: mhwang@stpi.narl.org.tw (M.H. Wang)

## 5. Figures

The USPTO patent xml has four major DOM layers. The first layer is the root layer, and the detail tags mostly belong to the bibliographic-data-grant layer (Fig. 1).

The Diversity of Patent Assignee and Foreigner Inventors from USPTO are shown in Fig. 2.

## 6. Tables

There are six types of patent xml DTDs. The major elements are the same; however, the tag names and values are different (Table 1).

We only compare those assignee countries, which have more than 10 foreigner inventor countries (Tables 2 and 3).

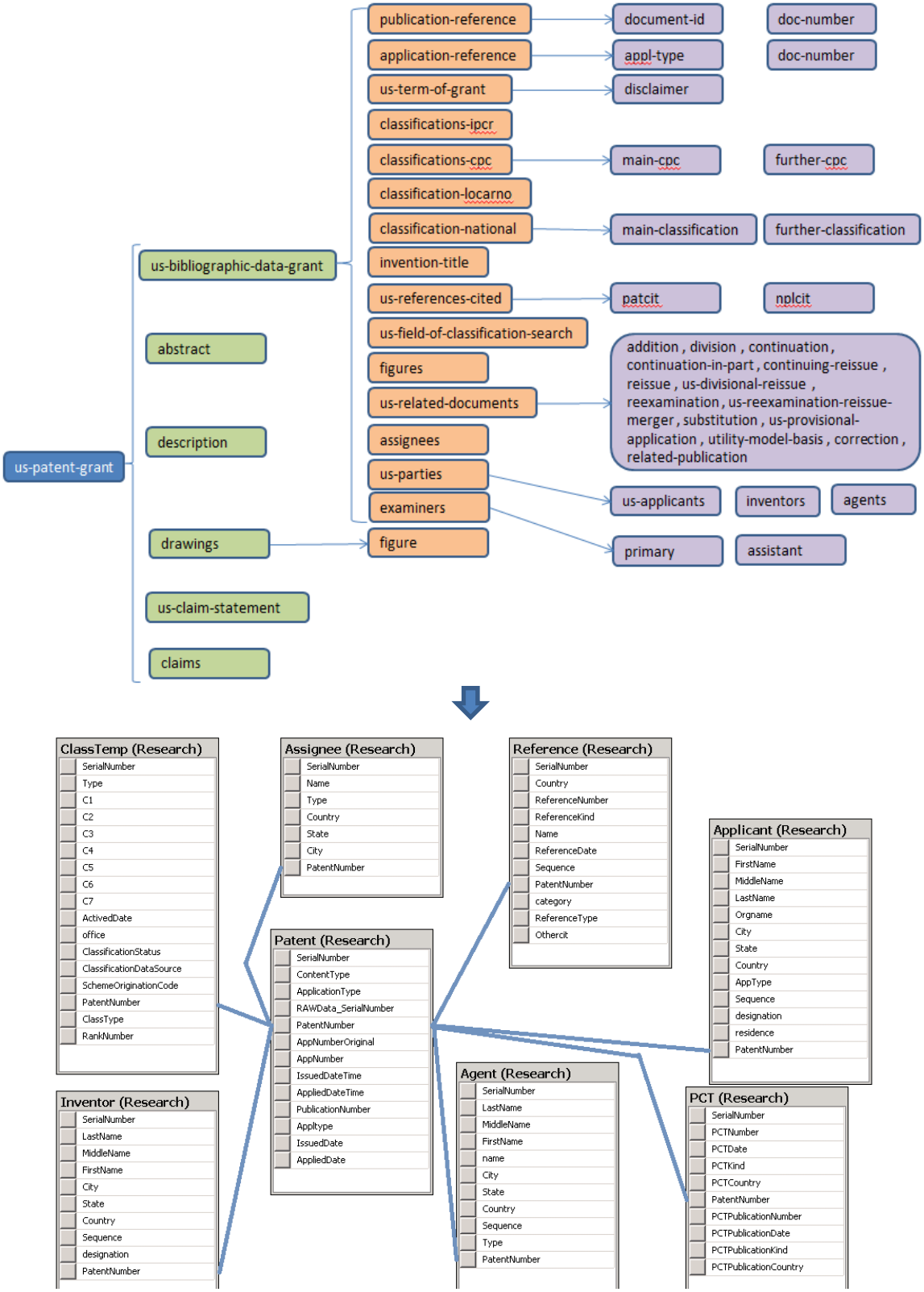


**Fig. 1:** Major elements of a USPTO patent grant

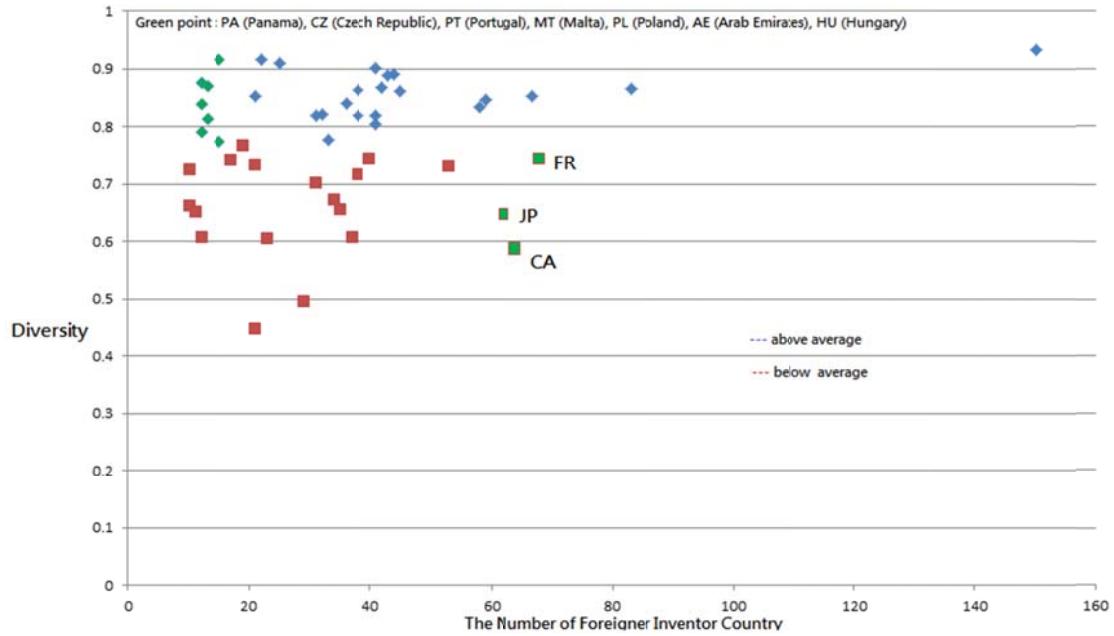


**Fig. 2:** Cross-nation diversity relationship for patents

**Table 1:** USPTO DTD versions and details

| Patent Year | DTD Version | File Numbers | File Intervals |
|---|---|---|---|
| 2015 | us-patent-grant-v45-2014-04-03.dtd | 52 | |
| 2014 | us-patent-grant-v44-2013-05-16.dtd | 52 | |
| 2013 | | 29 | 130618 ~ 131231 |
| | us-patent-grant-v43-2012-12-04.dtd | 22 | 130115 ~ 130611 |
| | us-patent-grant-v42-2006-08-23.dtd | 2 | 130101 ~ 130108 |
| 2012 ~ 2007 | | 52 * 5 + 53 (2008) | |
| 2006 | us-patent-grant-v41-2005-08-25.dtd | 52 | |
| 2005 | us-patent-grant-v40-2004-12-02.dtd | 52 | |

## 7. Results

Between year of 2014 to 2015, 147 countries had contributed to 653983 patents being filed to USPTO. After Simpson's Index calculating, we can gauge the diversity of the cross-nation strength. From Table 2 and Table 3, without question, the US is the country with the highest participation; however, a significant reason for this is that the data is from the US patent market. The interesting countries are France, Canada and Japan; they have more than 60 foreigner inventor countries but diversity index lower (Fig. 2). However, since their total number of patents is high, their contribution is easily diluted. Conversely, PA (Panama), CZ (Czech Republic), PT (Portugal), MT (Malta), PL (Poland), AE (Arab Emirates), HU (Hungary) are involved with less than 20 foreign countries; their diversity is above average (average: 0.773337).

## 8. Conclusion

From the USPTO patents, we were able to determine a trend in the raw data. There are several companies that do not include their confidential information in their patents to prevent hostile imitation by rivals. However, from the viewpoint of the application atmosphere of the largest business markets, intelligence assets can reveal the competitiveness and creativity of different countries as well as their preferences and expertise.

**Table 2:** Number and diversity of foreigner inventors by country (Below the average)

| Assignee Country | Number of patent | Foreigner Inventor Country | Diversity |
|---|---|---|---|
| Liechtenstein | 401 | 19 | 0.76667 |
| France | 4671 | 68 | 0.74462 |
| Singapore | 2127 | 40 | 0.74380 |
| New Zealand | 89 | 17 | 0.74258 |
| Brazil | 76 | 21 | 0.73407 |
| South Korea | 2531 | 53 | 0.73125 |
| Bulgaria | 48 | 10 | 0.72483 |
| China | 2562 | 38 | 0.71710 |
| Cayman Islands | 2005 | 31 | 0.70172 |
| Austria | 702 | 34 | 0.67170 |
| Mexico | 71 | 10 | 0.66217 |
| Israel | 533 | 35 | 0.65591 |
| Seychelles | 51 | 11 | 0.65052 |
| Japan | 6866 | 62 | 0.64830 |
| Taiwan | 3980 | 37 | 0.60756 |
| Iceland | 62 | 12 | 0.60614 |
| India | 149 | 23 | 0.60484 |
| Canada | 2957 | 64 | 0.58780 |
| Bermuda | 1894 | 29 | 0.49521 |
| Barbados | 862 | 21 | 0.44581 |

**Table 3:** Number and diversity of foreigner inventors by country (Above the average)

| Assignee Country | Number of patent | Foreigner Inventor Country | Diversity |
|---|---|---|---|
| USA | 56663 | 150 | 0.93312 |
| Panama | 27 | 15 | 0.91632 |
| South Africa | 49 | 22 | 0.91628 |
| Malaysia | 92 | 25 | 0.91092 |
| Spain | 200 | 41 | 0.90260 |
| Italy | 497 | 44 | 0.89145 |
| Sweden | 3259 | 43 | 0.88949 |
| Czech Republic | 31 | 12 | 0.87617 |
| Portugal | 33 | 13 | 0.87052 |
| Belgium | 1128 | 42 | 0.86808 |
| Germany | 8285 | 83 | 0.86504 |
| Luxemburg | 1311 | 38 | 0.86322 |
| Finland | 1411 | 45 | 0.86152 |
| UK | 2356 | 67 | 0.85414 |
| Cyprus | 120 | 21 | 0.85375 |
| Switzerland | 6689 | 59 | 0.84693 |
| Denmark | 898 | 36 | 0.84142 |
| Malta | 43 | 12 | 0.83937 |
| Netherlands | 4914 | 58 | 0.83466 |
| Saudi Arabia | 222 | 32 | 0.82136 |
| Island | 748 | 38 | 0.82038 |
| Norway | 299 | 31 | 0.82026 |
| Australia | 648 | 41 | 0.82020 |
| Poland | 31 | 13 | 0.81374 |
| Ireland | 1606 | 41 | 0.80467 |
| Arab Emirates | 50 | 12 | 0.79120 |
| Hong Kong | 960 | 33 | 0.77647 |
| Hungary | 51 | 15 | 0.77432 |

## Acknowledgment

We would like to express our gratitude to our colleagues (ML Chen, YT Lin) for providing their expertise and knowledge, which greatly assisted this research.

## References

Hu Z, Fang S and Liang T (2014). Empirical study of constructing a knowledge organization system of patent documents using topic modeling. Scientometrics, 100(3): 787-799.

Lee Y, Kim SY, Song I, Park Y and Shin J (2014). Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis. Scientometrics, 100(1): 227-244.

Oh S, Lei Z, Lee WC and Yen J (2014). Recommending missing citations for newly granted patents. In Data Science and Advanced Analytics (DSAA), 2014 International Conference on. IEEE: 442-448

Park H and Yoon J (2014). Assessing coreness and intermediarity of technology sectors using patent co-classification analysis: the case of Korean national RandD. Scientometrics, 98(2): 853-890.

Qu P, Zhang J, He Y, Zeng W and Xu H (2014). Term Extraction Using Co-occurrence in Abstract and First Claim for Patent Analysis. In Identification, Information and Knowledge in the Internet of Things (IIKI), 2014 International Conference on. IEEE: 60-63

Simpson EH (1949). Measurement of diversity. Nature, 163: 688

Srihari M and Mathew M (2014). Standardizing patent data cleaning in a university technology transfer office. In Management of Engineering and Technology (PICMET), 2014 Portland International Conference on. IEEE: 1368-1372.