Data Sharing Workshop: Insights from IPRoduct

Gaétan de Rassenfosse

March 17, 2021 via Zoom

www.iproduct.io
Welcome to IPRoduct

Welcome to the IPRoduct data platform. IPRoduct is an on-going academic research project that seeks to link products to patents. We source the data from virtual patent marking websites and focus primarily on U.S. patents.

After a proof-of-concept phase and a scaling-up phase, the project has now entered its third phase: data collection and enrichment through user contributions.

Everyone can register and contribute to improving the data using tools that we have developed. Each of your contributions gives you credits that you can use to download the final data.

The data platform is currently a public beta version. This means that we are eager to get your feedback and improve the platform to meet your needs.

Please register, explore the platform and help us improve it!

... and register on iproduct.io 😊

Follow us on Twitter...
Agenda

1. The IPRoduct database in brief
2. The biggest challenge
3. Ensuring a smooth user experience
4. Questions for discussion

With funding from:
Agenda

1. The IPRoduct database in brief
2. The biggest challenge
3. Ensuring a smooth user experience
4. Questions for discussion
A database that links products to patents...

Logitech Marathon Mouse M705
... using patent marks
Data sourced from virtual patent marking webpages

LOGITECH VIRTUAL PATENT MARKING

The Logitech products listed in the PDFs below are protected by patents in the United States. This webpage is provided in compliance with the virtual patent marking provisions of the America Invents Act.

These downloadable lists may not be all inclusive. The Logitech products listed may be protected by additional patents, and other patents may be pending. Any Logitech products not listed may be protected by one or more patents.

LOGITECH PATENT LIST - APRIL 15, 2020
LOGITECH PATENT LIST - DECEMBER 31, 2018
LOGITECH PATENT LIST - SEPTEMBER 19, 2017
LOGITECH PATENT LIST - NOVEMBER 11, 2015
Understanding VPM

https://www.nber.org/papers/w24288

https://doi.org/10.1093/jiplp/jpaa063
<table>
<thead>
<tr>
<th>Company Name</th>
<th>Domain</th>
<th>Sector</th>
<th>HC Country</th>
<th>Foundation Year</th>
<th>Number of employees</th>
<th>Name</th>
<th>Product Type</th>
<th>Authority</th>
<th>Number</th>
<th>Design</th>
<th>Kind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX200 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX150 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 2000</td>
<td>Logitech MX120 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX100 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX320 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX180 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX250 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX900 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logitech</td>
<td>logitech.com</td>
<td>Consumer Electronics</td>
<td>Switzerland</td>
<td>1981</td>
<td>5001 to 5000</td>
<td>Logitech MX900 Mouse</td>
<td>PRODUCT</td>
<td>KS</td>
<td>7008357</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Agenda

1. The IPRoduct database in brief
2. The biggest challenge
3. Ensuring a smooth user experience
4. Questions for discussion
Economists know the recipe for producing data

- Characterized by **fixed production cost and low marginal cost of distribution**.

- **Public funders** (or philanthropists) cover the cost of data production.
  - Incentives to share the data are even stronger if funders strongly encourage/require it. more and more

- The researcher invests her time in producing the data with the hope of publishing papers exploiting these data.
  - Incentives are even stronger if the data as such can attract citations.

- **Low marginal cost of reproduction.** more and more
  - Datavarses have been tremendously useful to drive down distribution costs to zero.
  - Proper documentation limits "customer support" requests.
IPRoduct: Cost of data production is **not fixed**

- New documents pop up at (known and unknown) places. → Need to continuously crawl the web
- No standard data structure across documents.
- Changing data structure for the same document. → Hard to automatize
- A lot of potential for data augmentation (e.g., product information). → Project grows bigger with time.

IPRoduct is open-ended project.
How to maximize production and diffusion

- **IPRoduct** cannot rely indefinitely on external funding. → Need to find ways of financing a fairly complex data production process.

- Solutions: crowd sourcing (with incentives)

- Three phases
  1. Club good approach (June–December 2020). Users receive credits in exchange of contributions, allowing them to download the data.
  2. Differentiated pricing (January 2021). Option to buy the data at reduced price for university researchers and full price for private firms. What price?
  3. Full subsidization of university researchers when payments from private firms covers running and data acquisition costs.
Five ways to contribute (so far)

Classify
Sort the wheat from the chaff by telling us which webpage contains relevant information. We have plenty of pages to show you.

Send Pictures
You found a physically marked product? Great! Send us a picture and we will add it to the database.

Share Web Documents
You think we should look more closely at some web documents you have found? Share them with us and we will ask the crowd to validate them.

Train classifier and validate results

Extension to physical marking

Enrich Products
Enrich the products by providing information you have gathered online about companies in the database.

Enrich Companies
Enrich the data by providing information you have gathered online about companies in the database.
<table>
<thead>
<tr>
<th>PRODUCT DESCRIPTION</th>
<th>MARKET INFORMATION</th>
<th>PROTECTION DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product name</td>
<td>Release Date</td>
<td>Protected by U.S. trademarks</td>
</tr>
<tr>
<td>ADNS-7550</td>
<td>Indicative Price</td>
<td>Add a value (i.e. 75234541), then press ENTER...</td>
</tr>
<tr>
<td>Product clean name</td>
<td>Currency</td>
<td>You can search on the WIPO Website for US Trademark numbers and then report them one by one in the text area above.</td>
</tr>
<tr>
<td>ADNS-7550</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item Description</td>
<td>no longer on the market since:</td>
<td>OPEN WIPO &gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Harmonized Commodity Description and Coding Systems (HS)**

- **Type**: Physical Product
- **Category**:

**Trademark data**

**Link to Amazon**

**Product enrichment data sheet**
About user contributions

- Managing the user community (i.e., validating the entries) is time consuming. Three levels:
  - Newcomer: contributions needs to be validated by the admin
  - Trusted: when enough contributions have been validated
  - Read-only: when too many contributions have been rejected

- So far, we have 100+ registered users but only a handful have contributed. Why?
  - Because the data are not yet available for download?
  - The interface is not user friendly enough?
  - The reward structure is unclear/unattractive?
Agenda

1. The IPRoduct database in brief
2. The biggest challenge
3. Ensuring a smooth user experience
4. Questions for discussion
Intuitive export module

Data Export
Get the data and explore your previous purchases. Exports in the public beta version are limited to a test sample.

Our complete database is composed of 760 companies, 1268 web pages, 22297 products, 28999 U.S. patents and 120567 relationships.

Create your query
Select the data you want and export them to CSV or Excel files.

What does it cost?
The data are currently not available for sales. They are available in exchange of credits, which you earn by helping us.
Making life easier for the user

- Creation of a FAQ or forum.
- Creation of a user guide—but who reads them?
- Connectivity:
  - Patents. DOCDB compatible ‘US-7650331-B1’; no link to PATSTAT appln_id (because I would need to buy the new release every time).
  - Products. ASIN, UPC and EAN
  - Trademarks. U.S. trademark numbers
- Suggestions welcome:


Agenda

1. The IPRoduct database in brief
2. The biggest challenge
3. Ensuring a smooth user experience
4. Questions for discussion
Questions for discussion

1. What should the terms of use be?
2. What can such projects learn from Wikipedia and other participative efforts?
3. What price points and reward structure?
4. Can I share the raw data (the actual VPM pages)?
5. I have been tempted to share the simple product-patent table but I am worried that it will kill the long-term survival prospects of the project… Maybe with a one-year lag?