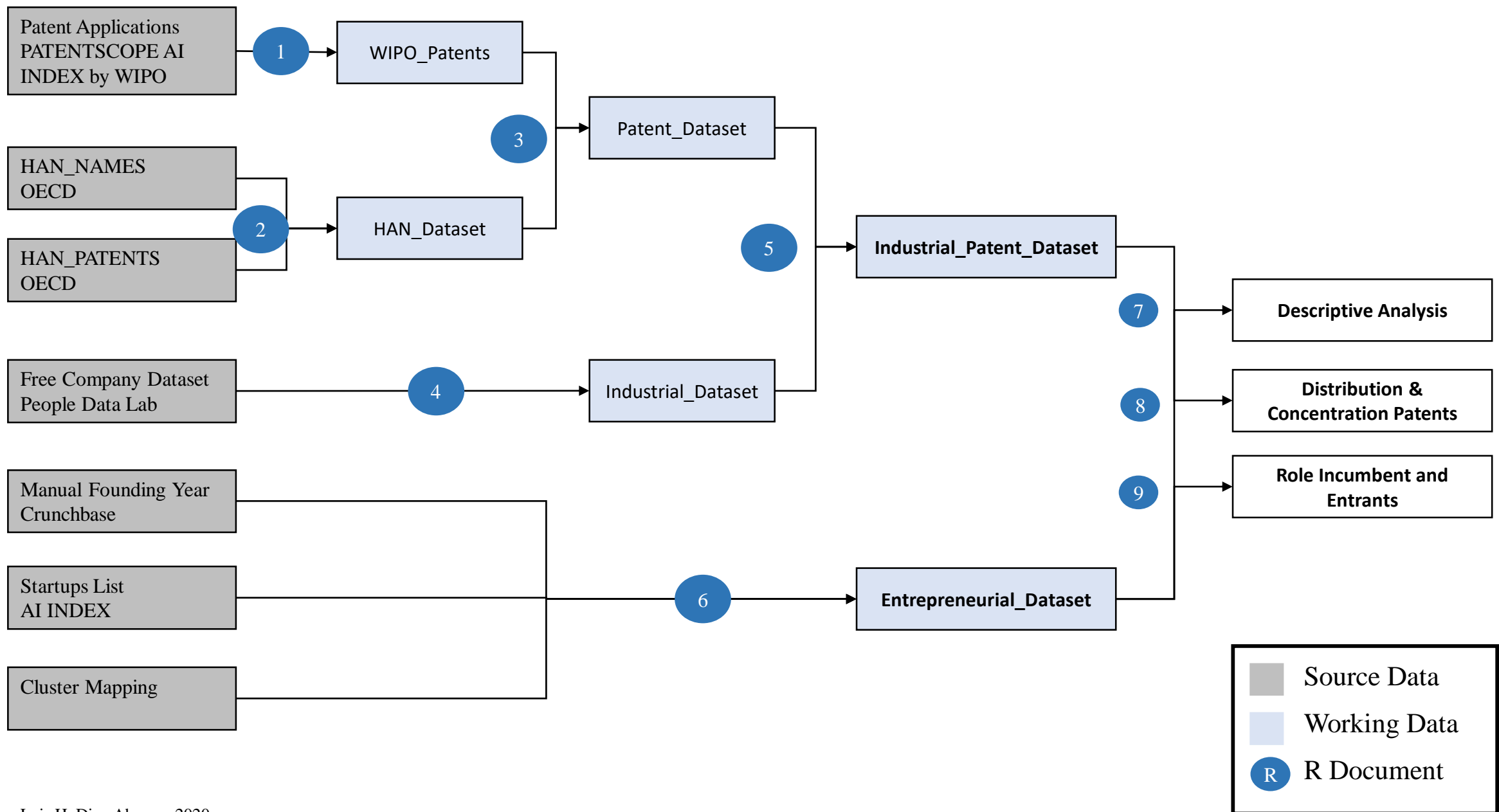


Incumbent and Entrant Dynamics in the Development and Diffusion of Artificial Intelligence

Data Preparation and Analysis Overview

Luis H. Diaz Abram – June 2020



R Documents Listing

1. WIPO-Patent_Data-Preparation
2. HAN_Data_Preparation
3. Patent_Harmonization
4. PDL_Data_Preparation
5. Match_Patent_Industrial
6. Entrepreneurial_Data_Preparation
7. Analysis-01-Descriptive
8. Analysis-02-Distribution_Concentration_Patents
9. Analysis-03-Role_Entrants_Incumbents

Overview of R Documents 1-2

1. WIPO-Patent_Data-Preparation

- Consolidation 66 queries into one dataset
- Classification patents based on Category and Field
- Identification unique patent applications

2. HAN_Data_Preparation

- Combination HAN_Name and HAN_Patents datasets provided by OECD

Overview of R Documents 3

3. Patent_Harmonization

- Merge WIPO_Patents and HAN_Dataset
- Removal remaining 2019 applications
- Selection patents from top 20 countries by number of applications
- Identification of type of applicant
- Removal of legal designation and non alpha-numeric characters by country
- Addition keyword for posterior fuzzy match
- Identifications patent applications with multiple applicants
- Label GPT market cluster (i.e. labeling patents as Applied or Core AI)

Overview of R Documents 4

4. PDL_Data_Preparation

- Removal organizations for which Name, Year of Foundation and Industry are blank
- Elimination Duplicates
- Country Selection
- Removal legal designation and non alpha-numeric characters by country
- Removal of organizations based on Industry
- Addition keyword for posterior fuzzy match

Overview of R Documents 5

5. Match_Patent_Industrial

- Direct Matching
- Convoluted Fuzzy Matching
- Calculation Applicant Age when patenting
- Elimination industrial information for applicants with negative age

Overview of R Documents 6

6. Entrepreneurial_Data_Preparation

- AI INDEX
 - Selection only Private Investments
 - Country selection
 - Removal legal designation and non-alphanumeric characters by country
- Merger AI Index Company list with data about Foundation year
- Removal companies for which foundation year is not available
- Mapping of activity clusters into ACS-CMM fields
- Calculation age of companies at time of investment