

Berkeley Fung

erudite-marker-539:JEMS16.assignee__disambiguation

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. "Machine learning and natural language processing applied to the patent corpus." Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

"UCB Fung Institute Patent Data" by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15
Rows	5,272,283
Size	239.7 MB

Schema

View in BigQuery

- **PatentNo** STRING NULLABLE
Patent number
 - **pdpass** STRING NULLABLE
Pdpass (unique identifier of assignees)
 - **assignee_disambiguated** STRING NULLABLE
Standardized assignee name
-

erudite-marker-539:JEMS16.assignee__raw

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. "Machine learning and natural language processing applied to the patent corpus." Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

"UCB Fung Institute Patent Data" by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15

Stat	Value
Rows	8,579,322
Size	767.6 MB

Schema

View in BigQuery

- **id** INTEGER NULLABLE
System generated
- **PatentNo** STRING NULLABLE
Patent number
- **Company** STRING NULLABLE
Assignee name (can be companies, universities, government agencies, or simply person name)
- **Geography** STRING NULLABLE
Raw (city, state, country) tuple of assignee
- **Country** STRING NULLABLE
Country Code derived from field ‘Geography’
- **State** STRING NULLABLE
State Code derived from field ‘Geography’ (if in U.S.)
- **City** STRING NULLABLE
City Name derived from field ‘Geography’
- **Sequence** STRING NULLABLE
Order of appearance (0 means the first assignee, 1 means the second assignee, . . . , etc)

erudite-marker-539:JEMS16.citation

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O’Reagan, D., Yeh, W., Zang, G., Fleming, L. “Machine learning and natural language processing applied to the patent corpus.” Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

“UCB Fung Institute Patent Data” by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15
Rows	174,205,746
Size	10.4 GB

Schema

View in BigQuery

- `id` INTEGER NULLABLE
System generated
 - `PatentNo_citing` STRING NULLABLE
Patent number (citing)
 - `CountryCodeOrNPL_cited` STRING NULLABLE
U.S. or a foreign country code or NPL (non-patent literature) of the cited art
 - `PatentNoOrNPL_cited` STRING NULLABLE
Patent number of non-patent literature (cited)
 - `sequence` STRING NULLABLE
Order of appearance (0 means the first cited art, 1 means the second cited art, ..., etc)
-

erudite-marker-539:JEMS16.citation_self

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. "Machine learning and natural language processing applied to the patent corpus." Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

"UCB Fung Institute Patent Data" by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15
Rows	1,667,637
Size	20.1 MB

Schema

View in BigQuery

- `PatentNo` STRING NULLABLE
Patent number
 - `Self_Citation_Flag` STRING NULLABLE
Backward prior art cites to the same pdpass
-

erudite-marker-539:JEMS16.cpc

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. "Machine learning and natural language processing applied to the patent corpus." Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

"UCB Fung Institute Patent Data" by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15
Rows	65,896,459
Size	3.7 GB

Schema

View in BigQuery

- **id** INTEGER NULLABLE
System generated
- **PatentNo** STRING NULLABLE
Patent number
- **Type** STRING NULLABLE
CPC
- **CPC_Full** STRING NULLABLE
Full CPC
- **CPC_Layer_1** STRING NULLABLE
CPC top layer 1: before {space}
- **CPC_Layer_2** STRING NULLABLE
CPC top layer 2: before {slash}
- **Sequence** STRING NULLABLE
Order of appearance (0 means the first CPC, 1 means the second CPC, ..., etc)

erudite-marker-539:JEMS16.inventor__disambiguated__2

Old table version 2, schema skipped.

erudite-marker-539:JEMS16.inventor__disambiguated__3

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. "Machine learning and natural language processing applied to the patent corpus." Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

“UCB Fung Institute Patent Data” by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

same table to inventor_disambiguated_2, except for data type differences easier for table joining

Stat	Value
Last updated	2018-02-23
Rows	13,345,776
Size	492.6 MB

Schema

View in BigQuery

- PatentNo STRING NULLABLE
 - InventorFullname STRING NULLABLE
 - InventorID STRING NULLABLE
-

erudite-marker-539:JEMS16.inventor__raw

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. “Machine learning and natural language processing applied to the patent corpus.” Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

“UCB Fung Institute Patent Data” by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15
Rows	14,745,325
Size	1.2 GB

Schema

View in BigQuery

- PatentNo STRING NULLABLE
Patent number
- Sequence STRING NULLABLE
Order of appearance (0 means the first inventor, 1 means the second inventor, ..., etc)
- FullName STRING NULLABLE

Full name (in form of Last Name {semicolon} First Name {single space} Middle Name)

- **LastName** STRING NULLABLE

Last Name

- **FirstMiddleName** STRING NULLABLE

First Name {single space} Middle Name

- **Geography** STRING NULLABLE

Raw (city, state, country) tuple of assignee

- **Country** STRING NULLABLE

Country Code derived from field 'Geography'

- **State** STRING NULLABLE

State Code derived from field 'Geography' (if in U.S.)

- **City** STRING NULLABLE

City Code derived from field 'Geography'

erudite-marker-539:JEMS16.patent_metadata_2

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. "Machine learning and natural language processing applied to the patent corpus." Forthcoming at Journal of Economics and Management Strategy.

Additional links:

o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>

o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

"UCB Fung Institute Patent Data" by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15
Rows	6,492,363
Size	5.4 GB

Schema

View in BigQuery

- **App1No** STRING NULLABLE

Application number

- **App1Date** STRING NULLABLE

Application date

- **PatentNo** STRING NULLABLE

Patent number

- **IssueDate** STRING NULLABLE

Patent issue date or grant date

- **FamilyID** STRING NULLABLE joins on **family__id**
Patent Family ID derived from USPTO HTML page of the focal patent
- **LawFirm** STRING NULLABLE
Agent / Law Firm / Correspondent
- **AssistExaminer** STRING NULLABLE
Assistant examiner
- **PrimaryExaminer** STRING NULLABLE
Primary examiner
- **Title** STRING NULLABLE
Patent title
- **Abstract** STRING NULLABLE
Patent abstract
- **GovernmentInterests** STRING NULLABLE
Full text statement acknowledging U.S. government supports (if any)

Join columns

FamilyID

joins to patents-public-data:patents.publications::family__id (87.45%, 5,677,428 rows)

Key	Percent	Rows	Sample values
all	87.45%	5,677,428	['41164314', '45348360', '46277349', '25524495', '44708394']

#standardSQL

```
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  ARRAY_AGG(first.FamilyID IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `erudite-marker-539.JEMS16.patent_metadata_2` AS first
LEFT JOIN (
  SELECT family_id AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.patents.publications`
  GROUP BY 1
) AS second ON first.FamilyID = second.second_column
```

joins from patents-public-data:patents.publications::family__id (25.67%, 25,206,642 rows)

erudite-marker-539:JEMS16.patent__novelty

Accompanying materials to

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G., W.S. Lueck, O'Reagan, D., Yeh, W., Zang, G., Fleming, L. "Machine learning and natural language processing applied to the patent corpus." Forthcoming at Journal of Economics and Management Strategy.

Additional links:

- o Inventor disambiguation golden file: <http://fung-storage.coe.berkeley.edu/disambig.golden.list.txt>
- o Inventor social network: <http://fung-storage.coe.berkeley.edu/inventors/>

“UCB Fung Institute Patent Data” by the University of California: Berkeley is licensed under a Creative Commons Attribution 4.0 International license.

Stat	Value
Last updated	2018-02-15
Rows	2,816,425
Size	90.7 MB

Schema

[View in BigQuery](#)

- **PatentNo** STRING NULLABLE

Patent number

- **Word** STRING NULLABLE

New word (unigram)

- **CurrentUse** STRING NULLABLE

Number of occurrence of the new word in the focal patent

- **FutureUse** STRING NULLABLE

Number of appearances of the new word in subsequent patents (up until Dec 31, 2014)