Dataset Descriptions

Use this document to provide descriptions of the entities and attributes of datasets in the IntelComp data like.

Please also include information on the provenance of the dataset and how often it can be updated

[The metadata schemata can be found here.](https://docs.google.com/spreadsheets/d/19nx4mn0dfc--cMAIusK5L5UvjuQVgByh8KPZR28LM7g/edit?usp=sharing)

# 1st round

## UC3M - SemanticScholar

Provenance: [SemanticScholar](https://api.semanticscholar.org/corpus/download/)

Update frequency: every few months

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| papers | Each of the publications. | parquet |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| id | Unique S.S. identifier. | string |  |
| title |  | string |  |
| paperAbstract |  | string |  |
| s2Url | URL to S2 research paper. | string |  |
| pdfUrls | URLs related. | array[string] |  |
| year | Publication year. | long |  |
| sources | Identifies papers sourced from DBLP and/or Medline. | array[string] |  |
| doi |  | string |  |
| doiUrl | DOI link. | string |  |
| pmid | PubMed identifier. | string |  |
| magId | Microsoft Academic Graph identifier. | string |  |
| fieldsOfStudy |  | array[string] |  |
| journalName |  | string |  |
| journalPages |  | string |  |
| journalVolume |  | string |  |
| venue |  | string |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| authors |  | parquet |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| id | Unique identifier. | string |  |
| name | Author’s name. | string |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| paper\_author | Paper – Authors relationship | parquet |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| paper\_id | S.S. paper identifier. | string | papers(id) |
| author\_id | S.S. author identifier. | string | authors(id) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| citations | Paper – Paper relationship | parquet |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| source | S.S. paper identifier. | string | papers(id) |
| dest | S.S. paper identifier. | string | papers(id) |