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** |
| S2papers | Papers. | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| paperID | Unique identifier. | serial |  |
| S2paperID | Unique S.S. identifier. Primary key. | char(40) |  |
| title |  | text |  |
| lowertitle | Lowercased title. | text |  |
| paperAbstract |  | text |  |
| s2Url | URL to S2 research paper. | text |  |
| pdfUrls | URLs related. | text |  |
| year | Publication year. | smallint |  |
| isDBLP | Identifies papers sourced from DBLP. | boolean |  |
| isMEDLINE | Identifies papers sourced from Medline. | boolean |  |
| doi |  | varchar(128) |  |
| doiUrl | DOI link. | varchar(256) |  |
| pmid | PubMed identifier. | int |  |
| magId | Microsoft Academic Graph identifier. | bigint |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| paperLemas | Research fields or subjects. | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| S2paperID | S.S. paper identifier. | char(40) | S2papers(S2paperID) |
| lemas | List of lemas | text |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| S2authors |  | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| authorID | Unique identifier. | serial |  |
| S2authorID | Unique S.S. identifier. Primary key. | int |  |
| orcidID | Unique ORCID identifier. | varchar(20) |  |
| orcidGivenName | ORCID name. | varchar(40) |  |
| orcidFamilyName |  | varchar(100) |  |
| scopusID | Unique Scopus identifier. | bigint |  |
| name | Author’s name. | varchar(256) |  |
| influentialCitationCount |  | smallint |  |
| ESP\_affiliation |  | boolean |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| paperAuthor | Paper – Authors relationship | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| paperAuthorID | Unique identifier. Primary key. | serial |  |
| S2paperID | S.S. paper identifier. | char(40) | S2papers(S2paperID) |
| S2authorID | S.S. author identifier. | int | S2authors(S2authorID) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| S2fields | Research fields or subjects. | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| fieldID | Unique identifier. Primary key. | serial |  |
| fieldName |  | varchar(32) |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| paperField | Paper – Field relationship | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| paperFieldID | Unique identifier. Primary key. | serial |  |
| S2paperID | S.S. paper identifier. | char(40) | S2papers(S2paperID) |
| fieldID | Field identifier. | int | S2fields(fieldID) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| S2venues | Research venues. | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| venueID | Unique identifier. Primary key. | serial |  |
| venueName |  | varchar(320) |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| paperVenue | Paper – Venue relationship | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| paperVenueID | Unique identifier. | serial |  |
| S2paperID | S.S. paper identifier. | char(40) | S2papers(S2paperID) |
| venueID | Venue identifier. | int | S2venues(venueID) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| S2journals | Research journals. | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| journalID | Unique identifier. Primary key. | serial |  |
| journalName |  | varchar(320) |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Entities** | | | |
| **name** | **description / assumptions** | **format** | **foreign key / relation** |
| paperJournal | Paper – Venue relationship | PostgreSQL |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| paperJournalID | Unique identifier. Primary key. | serial |  |
| S2paperID | S.S. paper identifier. | char(40) | S2papers(S2paperID) |
| journalID | Journal identifier. | int | S2journals(journalID) |
| journalVolume | Paper publication volume. | varchar(300) |  |
| journalPages | Paper publication pages. | varchar(100) |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | | | |
| **name** | **description / assumptions** | **format** | **foreign key/relation** |
| citationID | Unique identifier. | serial |  |
| S2paperID1 | S.S. paper identifier. | char(40) | S2papers(S2paperID) |
| S2paperID2 | S.S. paper identifier. | char(40) | S2papers(S2paperID) |
| isInfluential |  | boolean |  |
| MethodIntent | Citation type. | boolean |  |
| BackgrIntent | Citation type. | boolean |  |
| ResultIntent | Citation type. | boolean |  |