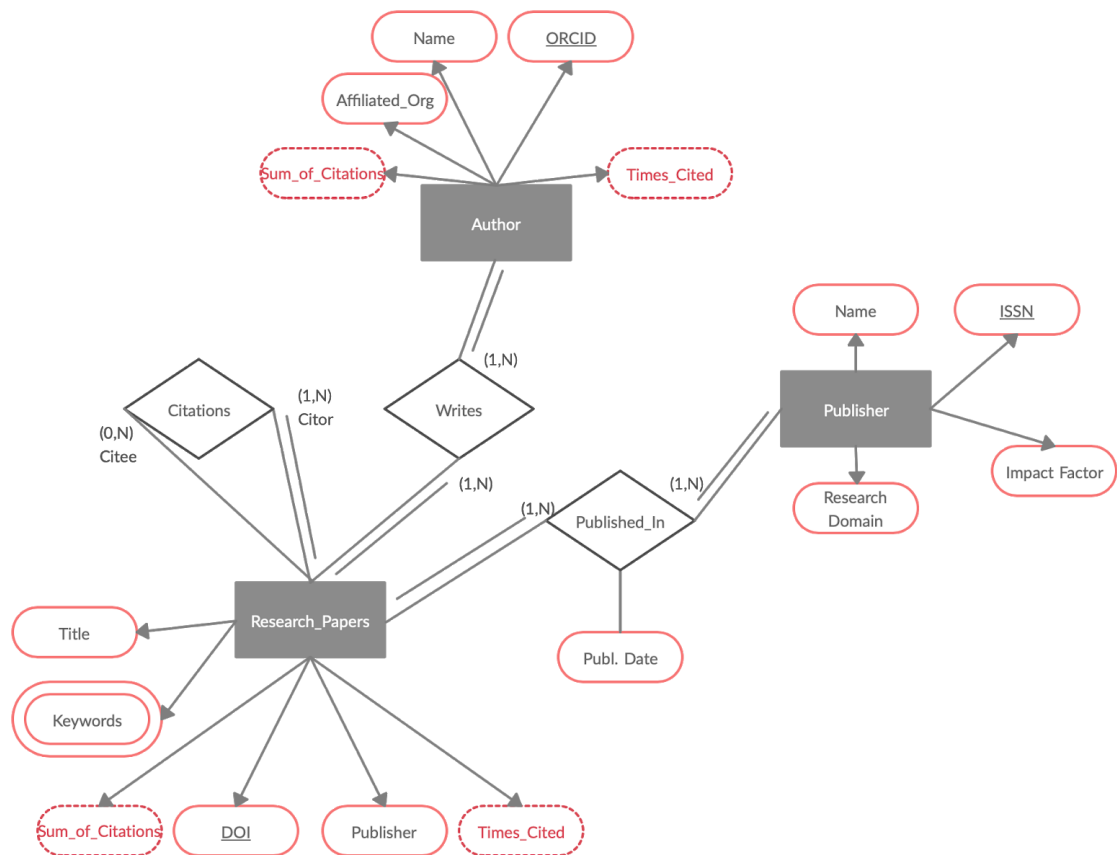# DSC 461: Milestone 2

DatSci Werx

Team #39

Tristan De Alwis (tdealwis); All parts contributed; 50%

Weiran Lin (wlin25); All parts contributed; 50%

Task A : Draw a ER diagram

Assumptions: there are 3 relationship types in our ER diagram

Each Author writes a research paper. This is M:N relationship: for research paper entity type, it is total participation meaning that all research paper should have author, the (min, max) constraint for research paper participation is (1, N) meaning that a research paper can have any number of author; for author entity type, it is total participation meaning that all author should have a published paper or else we would not include him in our database, the (min, max) constraint for author participation (1, N) meaning that an author can have any number of research papers.

Each research paper is published by a publisher. This is N:1 relationship: for research paper entity type, it is total participation meaning that all research paper should have publisher, the (min, max) constraint for research paper participation is (1, 1) meaning that a research paper can only have one publisher; for publisher entity type, it is total participation meaning that all publisher should have published research paper, the (min, max) constraint for publisher participation is (1, N) meaning that publisher can publish any number of research paper;

Research paper is cited by research paper. This is M:N relationship: for research paper entity type as citee, it is partial participation meaning that not all research paper can be cited by another paper, the (min, max) constraint for research paper participation is (0, N) meaning that a research paper can be cited

from 0 to any number of times; for research paper entity type as citor, it is total participation meaning that all research paper should citing another paper in its reference section, the (min, max) constraint for author participation (1, N) meaning that an research paper can cite any number of another research papers.

Task B: Relational Database Design using ER-to-Relational Mapping

1. ER-to-Relational Mapping Algorithm
   Steps
   1. Mapping of Regular Entity Types
      a. Research paper relation with attributes (<u>DOI</u>, Title, Publisher
      b. Author relation with attributes(<u>ORCID</u>, name, affiliated organization
      c. Publisher relation with attibutes(<u>ISSN</u>, name, impact factor, research domain)

   2. Mapping of Weak Entity Types
      a. Not applicable

   3. Mapping of Binary 1:1 Relationship Types
      a. Not applicable

   4. Mapping of Binary 1:N Relationship Types
      a. Research paper relation with attributes (<u>DOI</u>, Title, Publisher_ISSN, Publication Date)
   5. Mapping of Binary M:N Relationship Types
      a. Writes relation with attributes(<u>Author ORCID</u>, <u>Paper DOI</u>)
      b. Citations relation with attributes(<u>Citor Paper DOI</u>, <u>Citee Paper DOI</u>)
      c. Research paper relation with attributes (<u>DOI</u>, Title, Publisher_ISSN, Publ_Date, Cited Times, Citing times)
      d. Author relation with attributes(<u>ORCID</u>, Name, Affiliated Organization,
      e. Sum of Cited times, sum of citing times)

   6. Mapping of Multivalued Attributes
      a. Keywords Relation with attributes(<u>Paper DOI</u>, <u>Keyword)</u>

   7. Mapping of N-ary Relational Types
      a. Not Applicable

   8. Options for Mapping Specialization or Generalization
      a. Not Applicable

   9. Mapping of Union Types(Categories)
      a. Not Applicable

Research Paper

DOI	Title	Publisher_ISSN	Publ_Date	Sum_of_Citations	Times_Cited

Author

ORCID	Name	Affiliated_Org	Sum_of_Citations	Times_Cited

Publisher

ISSN	Name	impact_factor	Research_Domain

Writes

Author_ORCID	Paper_DOI

Citations

Citor_DOI	Citee_DOI

Keywords

Paper_DOI	Keyword
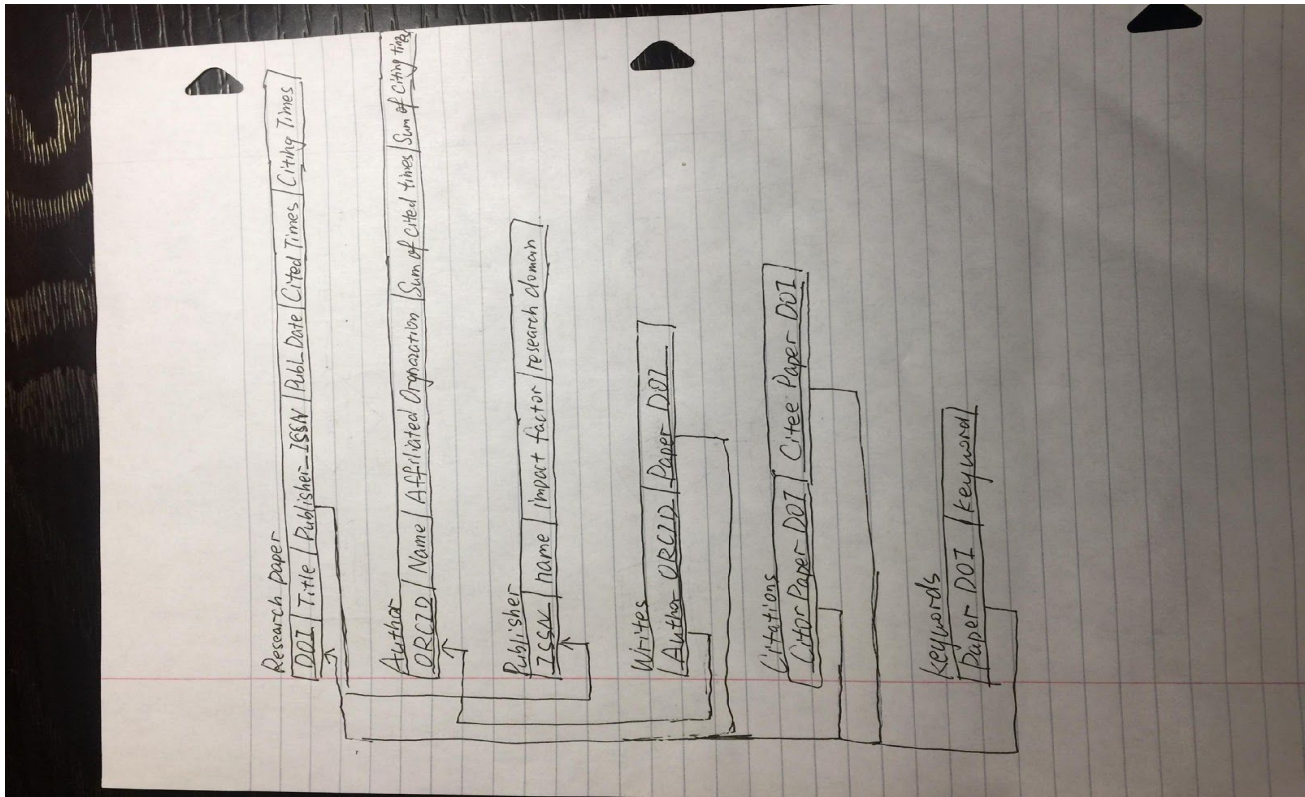
Relation Name	ER diagram components
Research Paper	E(Research Paper)+R(Published_In)
Author	E(Author)
Publisher	E(Publisher)
Writes	R(Writes)
Citations	R(Citations)
Keywords	A(Keywords)

2. Schema of Database

    1.1.1   CREATE TABLE Research Paper(
          DOI   CHAR   NOT NULL
          #Digital Object Identifier serves as an identifier for each research paper, usually takes the format of "prefex/suffix"
          Title   CHAR   NOT NULL
          Publisher_ISSN   CHAR(9)
          Publ_Date   Date
          Cited_Times   INT   DEFAULT 0
          #Cited_Times is an aggregated function derived from Citations relation to calculate the number of each research paper as citee
          Citing_Times   INT   DEFAULT 0
          #Citing_Times is an aggregated function derived from Citations relation to calculate the number of each research paper as citor
          PRIMARY KEY (DOI)
          FOREIGN KEY (Publisher_ISSN) REFERENCES Publisher(ISSN)
               ON DELETE SET NULL   ON UPDATE CASCADE);

    1.1.2   CREATE TABLE Author(
          ORCID   CHAR(15)   NOT NULL
          #Open Researcher and Contributor ID serves as an identifier for each researcher, usually takes the format of "xxxx-xxxx-xxxx-xxxx"
          Name   CHAR   NOT NULL
          Affiliated Organization CHAR
          Sum_of_Cited_Times   INT   DEFAULT 0
          #Sum_of_Cited_Times is an aggregated function derived from Writes relation and Citations relation to calculate the total number of research paper as citee written by certain researcher.
          Sum_of_Citing_Times   INT   DEFAULT 0
          # Sum_of_Citing_Times is an aggregated function derived from Writes relation and Citations relation to calculate the total number of research paper as citor written by certain researcher.
          PRIMARY KEY (ORCID));

    1.1.3   CREATE TABLE Publisher(
          ISSN   CHAR(9)   NOT NULL
          #International Standard Serial Number serves as an identifier for each publisher, usually takes the format of "xxxx-xxxx"
          Name   CHAR   NOT NULL
          Impact Factor FLOAT

Journal impact factor (JIF) of an academic journal is a scientometric index that reflects the yearly average number of citations that articles published in the last two years in a given journal received.
Research Domain CHAR
PRIMARY KEY (ISSN));

1.1.4 CREATE TABLE Writes(
Author ORCID   CHAR(15)   NOT NULL
Paper DOI   CHAR   NOT NULL
PRIMARY KEY (Author ORCID, Paper DOI)
FOREIGN KEY (Author ORCID) REFERENCES Author(ORCID)
  ON DELETE CASCADE   ON UPDATE CASCADE
FOREIGN KEY (Paper DOI) REFERENCES Research Paper(DOI)
        ON DELETE CASCADE   ON UPDATE
CASCADE);

1.1.5 CREATE TABLE Citations(
Citor Paper DOI   CHAR   NOT NULL
Citee Paper DOI   CHAR   NOT NULL
PRIMARY KEY (Citor Paper DOI, Citee Paper DOI)
FOREIGN KEY (Citor Paper DOI) REFERENCES Research Paper(DOI)
  ON DELETE CASCADE   ON UPDATE CASCADE
FOREIGN KEY (Citee Paper DOI) REFERENCES Research Paper(DOI)
        ON DELETE CASCADE   ON UPDATE
CASCADE);

1.1.6 CREATE TABLE Keywords(
Paper DOI   CHAR   NOT NULL
Keyword   CHAR
PRIMARY KEY (Paper DOI)
FOREIGN KEY (Paper DOI) REFERENCES Research Paper(DOI)
  ON DELETE CASCADE   ON UPDATE CASCADE);