

DATASET DESCRIPTION

- 1) **Google Web Graph dataset:** This dataset can be downloaded from <https://snap.stanford.edu/data/web-Google.html>
This data was released by Google as a part of Google programming contest in 2002.

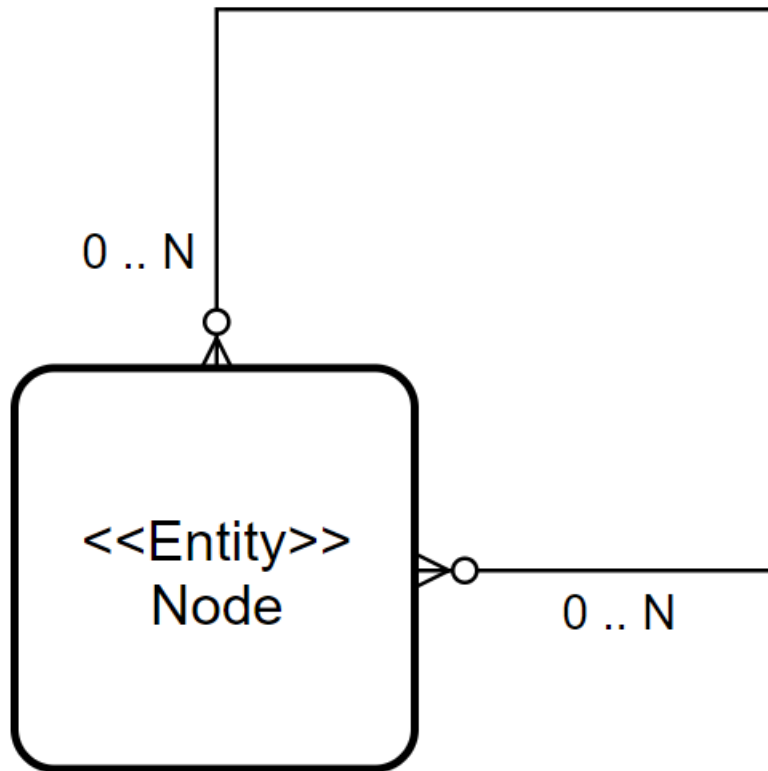


Figure 1: Google Web Graph ER diagram

2) **Yahoo dataset:** This dataset can be downloaded from

<https://finance.yahoo.com/quote/CSV/history?p=CSV>

We use this data to know the stock market effects on particular date or month or some year. We can download this data as per our selection criteria given in particular fields “data from-to”, we can customize the data using the time period and frequency.

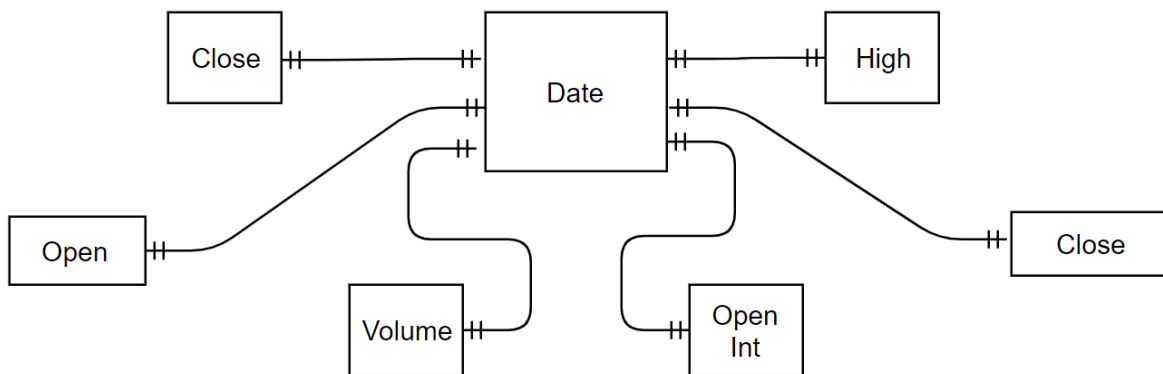


Figure 2: Yahoo dataset ER diagram

3) **MAG dataset:** This dataset can be downloaded from

<https://aminer.org/open-academic-graph>

This dataset comes from Microsoft Academic Graph and AMiner, consists of scientific publication records and their relationships with publications, a paper consists of several contents and each of their relationships.

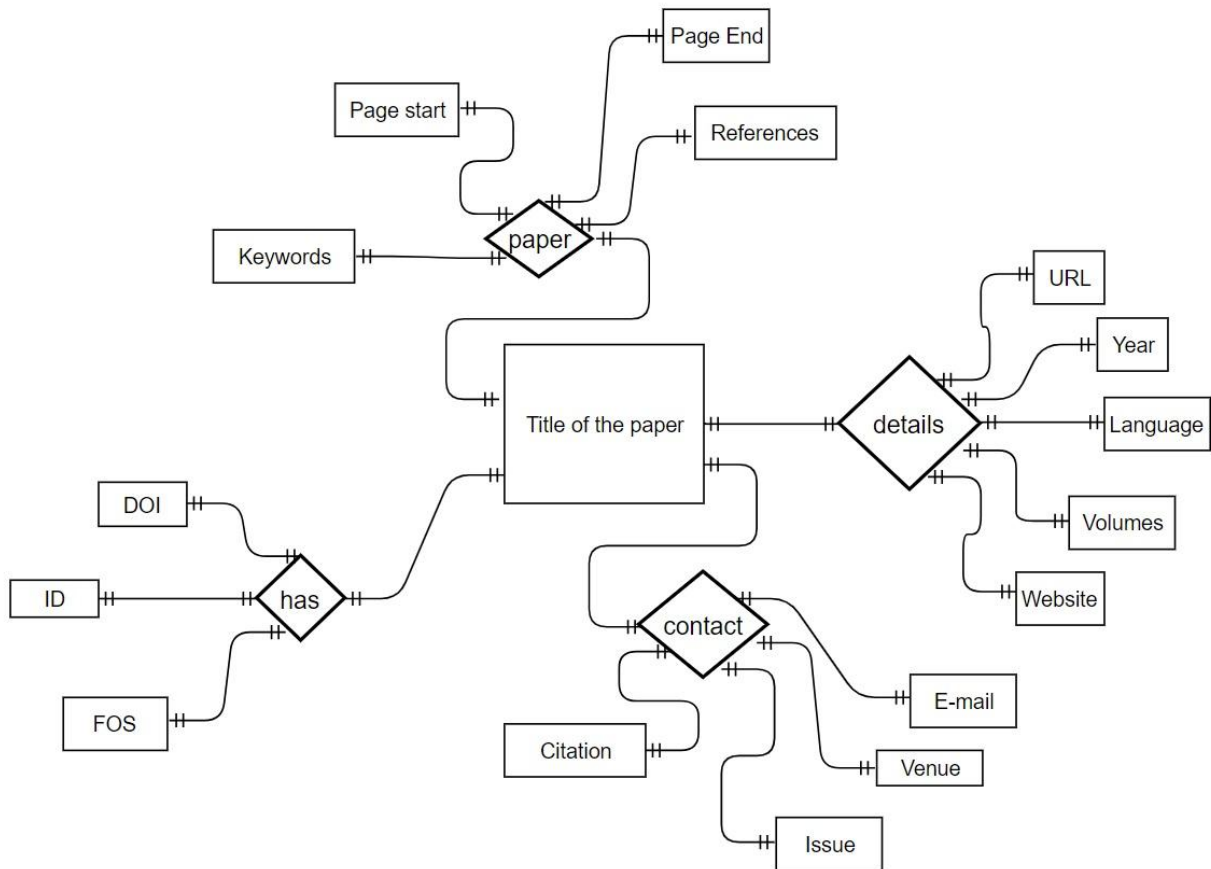


Figure 3: MAG dataset ER diagram

- 4) **TPC-H dataset:** This dataset can be downloaded from <https://relational.fit.cvut.cz/dataset/TPCH>

It is a decision support benchmark and also popular one for comparing database vendors. This benchmark is a standard model for all the industries who try to sell their goods or distribute their products all over the world like food distribution, car rentals, parts and suppliers.

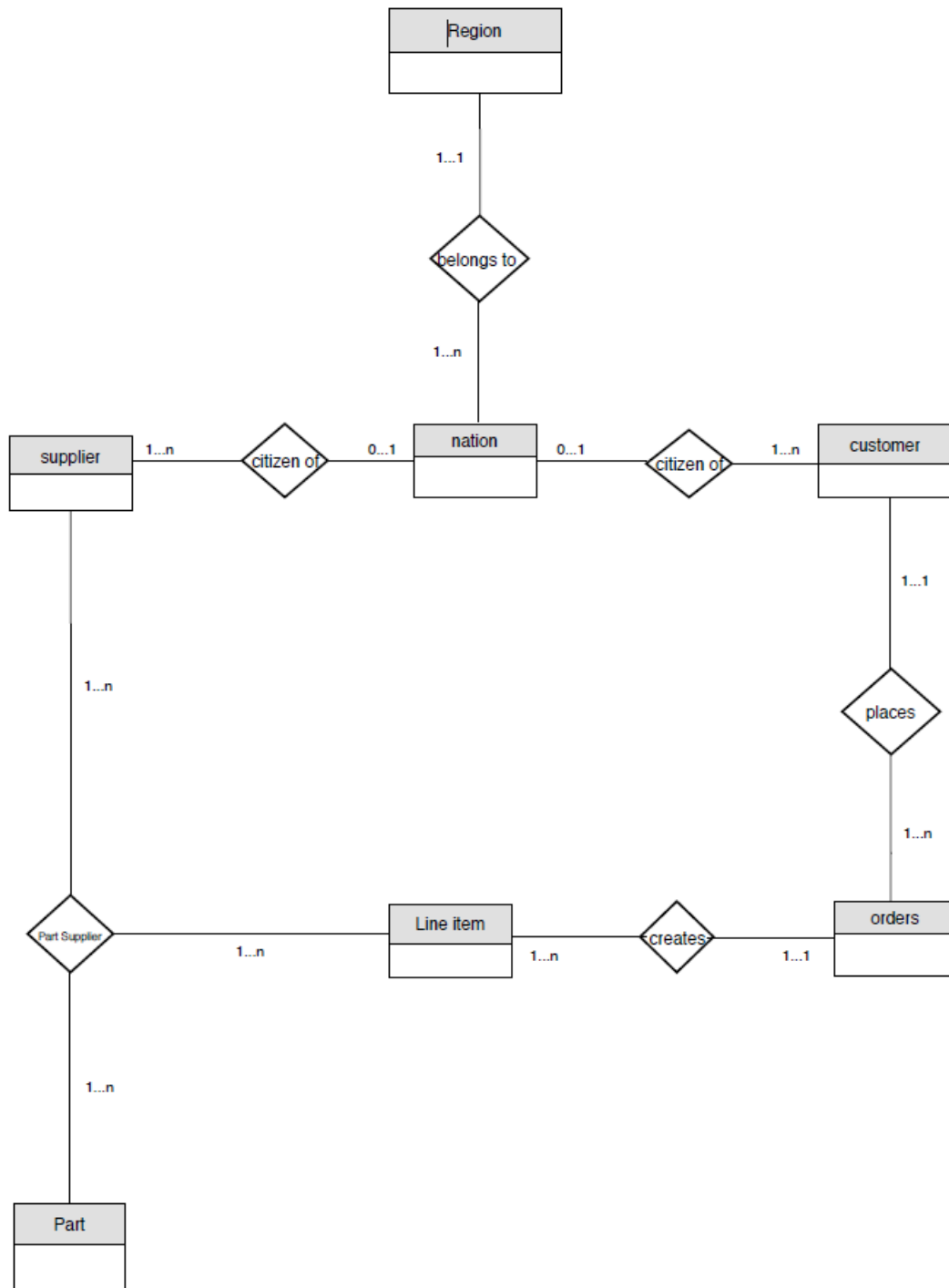


Figure 4: TPC-H ER diagram

Tabular Information:

This Table gives the information of which dataset has been used (artificial or original data), file format and its size, and the no. of Nodes and Edges.

DATASET	ARTIFICIAL/REAL	FILE FORMAT	SIZE	NODES	EDGES
TPC-H	Real	CSV	2 GB	-	-
Google Web Graph	Real	Text	71.8 MB	875713	5105039
MAG dataset	Real	CSV	100 GB	384 (for a file of 48 records)	336 (for a file of 48 records)
Yahoo financial dataset	Real	CSV	115 MB	9800 (for a file of 1400 records)	8400 (for a file of 1400 records)

Table 1: Dataset information