SQL, SQL, SQL

CSE 444 section

October 7, 2010

Today

- Basic SQL review
- Practice with grouping and aggregation

Document index database

Author (aid, name)

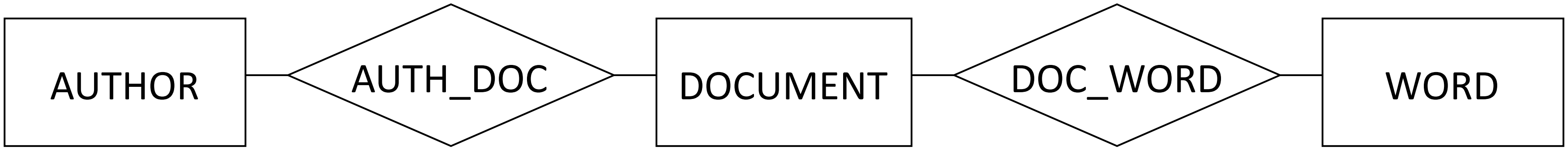
Auth_Doc (aid, did)

Document (did, title, year)

Doc_Word (did, word)

Word (word)

Underlined = key (unique identifier for a tuple)



Warm-up exercises

- Authors whose last name is “Crick”
- All documents written in 2000 or later
- Names and years of all documents from earliest to latest

Using more than one table

Who wrote this paper?

“Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid” (1953)

Authors of double-helix paper

Word count of double-helix paper

Today

- Basic SQL review
- Practice with grouping and aggregation

Find authors who wrote ≥ 20 docs

Find authors who wrote ≥ 20 docs

This could work:

```
SELECT name
FROM Author a
WHERE 20 <= (SELECT COUNT(*) FROM
Auth_Doc ad WHERE ad.aid = a.aid)
```

Find authors who wrote ≥ 20 docs

Use grouping to eliminate the subquery:

```
SELECT name
FROM Author a, Auth_Doc ad
WHERE a.aid = ad.aid
GROUP BY a.aid, a.name
HAVING COUNT(*) >= 20
```

Find authors who wrote ≥ 20 docs

Use grouping to eliminate the subquery:

```
SELECT name
FROM Author a, Auth_Doc ad
WHERE a.aid = ad.aid
GROUP BY a.aid, a.name
HAVING COUNT(*) >= 20
```

← One row per (a.aid, a.name) pair

Find authors who wrote ≥ 20 docs

Use grouping to eliminate the subquery:

```
SELECT name
FROM Author a, Auth_Doc ad
WHERE a.aid = ad.aid
GROUP BY a.aid, a.name
HAVING COUNT(*) >= 20
```

← Only groups that
combine ≥ 20
tuples will match

Find authors who wrote ≥ 20 docs

Use grouping to eliminate the subquery:

```
SELECT name
FROM Author a, Auth_Doc ad
WHERE a.aid = ad.aid
GROUP BY a.aid, a.name
HAVING COUNT(*) >= 20
```

← If aid is the key, why group by name?

If we deleted a.name…

ERROR: Column 'name' is invalid in the select list because it is not contained in either an aggregate function or the GROUP BY clause.

Finding literate authors

How can we find authors who use more than 10,000 distinct words?

Authors who use > 10,000 words

```sql
SELECT name
FROM Author a, Auth_Doc ad,
     Doc_Word dw
WHERE a.aid = ad.aid AND ad.did = dw.did
GROUP BY a.aid, a.name
HAVING COUNT(DISTINCT word) > 10000
```

Authors who use > 10,000 words

```
SELECT name
FROM Author a, Auth_Doc ad,
          Doc_Word dw
WHERE a.aid = ad.aid AND ad.did = dw.did
GROUP BY a.aid, a.name
HAVING COUNT(DISTINCT word) > 10000
```

→ What does DISTINCT mean within COUNT?

More examples

- For each author, give the total number of words in all documents he has (co-)written.
- For each author, give the average length in words of his documents.
- Give the author with the longest average documents.

Total word count by author

Average word count by author

Wordiest-on-average author

Try these at home

- All words used by at least 10 authors
- The most frequently used word
- The longest document
- Authors of the longest document