**use** hcibib;

*-- Set Difference - all authors who have not written in CHI*

*-- Start with authors in CHI*

**SELECT DISTINCT** author\_id, author\_name

**FROM** author

**JOIN** article\_author **USING** (author\_id)

**JOIN** article **USING** (article\_id)

**JOIN** proceedings **USING** (proc\_id)

**JOIN** conf\_series **USING** (cs\_id)

**WHERE** cs\_hb\_key = **'CHI'**

**ORDER BY** author\_name;

*-- What if we <>?*

**SELECT DISTINCT** author\_id, author\_name

**FROM** author

**JOIN** article\_author **USING** (author\_id)

**JOIN** article **USING** (article\_id)

**JOIN** proceedings **USING** (proc\_id)

**JOIN** conf\_series **USING** (cs\_id)

**WHERE** cs\_hb\_key <> **'CHI'**

**ORDER BY** author\_name;

*-- what this did: this is people who published in not-CHI*

*-- NOT poeple who didn't publish in CHI*

*-- Remember 'IN'? It can be negated*

**SELECT DISTINCT** author\_id, author\_name

**FROM** author

**WHERE** author\_id **NOT IN** (

**SELECT** author\_id

**FROM** article\_author

**JOIN** article **USING** (article\_id)

**JOIN** proceedings **USING** (proc\_id)

**JOIN** conf\_series **USING** (cs\_id)

**WHERE** cs\_hb\_key = **'CHI'**

)

**ORDER BY** author\_name;

*-- Another way: outer join \*with\* subselect*

**SELECT DISTINCT** a.author\_id, author\_name

**FROM** author a

**LEFT OUTER JOIN** (

**SELECT** author\_id

**FROM** article\_author

**JOIN** article **USING** (article\_id)

**JOIN** proceedings **USING** (proc\_id)

**JOIN** conf\_series **USING** (cs\_id)

**WHERE** cs\_hb\_key = **'CHI'**

) ca **ON** a.author\_id = ca.author\_id

**WHERE** ca.author\_id **IS NULL**

**ORDER BY** author\_name;

*-- A new friend: EXISTS*

**SELECT DISTINCT** author\_id, author\_name

**FROM** author a

**WHERE NOT** *EXISTS* (

**SELECT** *\**

**FROM** article\_author

**JOIN** article **USING** (article\_id)

**JOIN** proceedings **USING** (proc\_id)

**JOIN** conf\_series **USING** (cs\_id)

**WHERE** cs\_hb\_key = **'CHI' AND** author\_id = a.author\_id

)

**ORDER BY** author\_name;

*-- All authors who have published in AH*

*-- There does not exist a year of AH for which the author did not have a paper*

*-- First, a helper view:*

**CREATE VIEW** author\_proc\_articles

**AS SELECT** author\_id, cs\_id, proc\_id, *COUNT*(article\_id)

**FROM** article\_author

**JOIN** article **USING** (article\_id)

**JOIN** proceedings **USING** (proc\_id)

**GROUP BY** author\_id, cs\_id, proc\_id;

**SELECT DISTINCT** author\_id, author\_name

**FROM** author

**JOIN** author\_proc\_articles apa1 **USING** (**author\_id**)

**JOIN** conf\_series cs **USING** (**cs\_id**)

**WHERE** cs\_hb\_key = **'AH'**

**AND NOT** *EXISTS* (

**SELECT proc\_id FROM** proceedings inp

**WHERE** inp.cs\_id = cs.cs\_id

**AND NOT** *EXISTS* (

**SELECT** *\** **FROM** author\_proc\_articles apa2

**WHERE** apa1.**author\_id** = apa2.**author\_id**

**AND** apa2.**proc\_id** = inp.proc\_id

)

);

*-- Simplify*

**SELECT DISTINCT** author\_id, author\_name

**FROM** author

**JOIN** author\_proc\_articles apa1 **USING** (**author\_id**)

**WHERE NOT** *EXISTS* (

**SELECT proc\_id FROM** proceedings inp

**JOIN** conf\_series cs **USING** (**cs\_id**)

**WHERE** cs.cs\_hb\_key = **'AH'**

**AND NOT** *EXISTS* (

**SELECT** *\** **FROM** author\_proc\_articles apa2

**WHERE** apa1.**author\_id** = apa2.**author\_id**

**AND** apa2.**proc\_id** = inp.proc\_id

)

);

*-- Fast way - counting!*

*-- Simplify*

**SELECT** author\_id, author\_name, *COUNT*(**DISTINCT proc\_id**) **AS** nprocs

**FROM** author

**JOIN** author\_proc\_articles apa1 **USING** (**author\_id**)

**JOIN** conf\_series **USING** (**cs\_id**)

**WHERE** cs\_hb\_key = **'AH'**

**GROUP BY** author\_id, author\_name

**HAVING** nprocs = (**SELECT** *COUNT*(**DISTINCT proc\_id**)

**FROM** proceedings **JOIN** conf\_series **USING** (cs\_id)

**WHERE** cs\_hb\_key = **'AH'**);

**SELECT** *COUNT*(**DISTINCT** proc\_id)

**FROM** proceedings **JOIN** conf\_series **USING** (cs\_id)

**WHERE** cs\_hb\_key = **'AH'**;

*-- INTERSECT stuff*

**SELECT** book\_id, **'az' AS source**

**FROM** az\_ratings **JOIN** isbn\_book\_id ib **ON** (asin = isbn)

**UNION DISTINCT**

**SELECT** book\_id, **'bx' AS source**

**FROM** bx\_ratings **JOIN** isbn\_book\_id ib **USING** (isbn)

**ORDER BY** book\_id;

*-- INTERSECT -- all books in both data sets*

**SELECT DISTINCT** book\_id

**FROM** isbn\_book\_id

**JOIN** az\_ratings **ON** (asin = isbn)

**JOIN** bx\_ratings **USING** (isbn);

**SELECT** 5 **IS DISTINCT FROM null**;

**SELECT NULL IS DISTINCT FROM null**;