Thematic Ranking of Object Summaries for Keyword Search

DKE 2018

GEORGIOS J. FAKAS
Department of Information Technology,
Uppsala University,
Sweden

YILUN CAI
Lenovo Group Limited
Hong Kong

ZHI CAI
College of Computer Science
Beijing University of Technology
Beijing, China

NIKOS MAMOULIS
Department of Computer Science
The University of Hong Kong
Hong Kong
Outline

1. Motivation
2. Background & Related work
3. Themtiac Size-$l$ OSs
4. Approaches
5. Evaluation Results
6. Conclusion & Future Work
1.1 Object Summaries

- Relational Databases are everywhere: Web, Desktops etc.
- Social graphs are also everywhere!
- Difficult to retrieve information about a **Data Subject (DS)** unless you know very well:
  - SQL and
  - Schema details etc.
- There is a need for keyword search facilities analogous to Web

Select *
From Employees, Orders, Shippers
Where Employees.ID=Orders.ID
  AND Orders.Shipper=Shippers.ID
  AND Name="Leverling"
1.1 Object Summaries

Query Search: **Faloutsos**
Web Search Result: ranked set of links and snippets
1.1 Object Summaries

Query Search: **Faloutsos**
Web Search Result: ranked set of links and snippets
1.1 Object Summaries

Query Search: **Faloutsos**

Web Search Result: ranked set of links and snippets

Query Search: **Faloutsos**

OS Result: set of OSs and size-$l$ OSs.
1.1 Object Summaries

Query Search: Faloutsos
Web Search Result: ranked set of links and snippets

Query Search: Faloutsos
OS Result: set of OSs and size-1 OSs.
Outline

1. Motivation
2. Background & Related work
3. Themtiac Size-$l$ OSs
4. Approaches
5. Evaluation Results
6. Conclusion & Future Work
2.1 Object Summaries

**OS Generation - Methodology**

- **t<sup>DS</sup>** a central tuple containing the Kw; tuples around t<sup>DS</sup> contain additional information about the Data Subject.
- **R<sup>DS</sup>** the corresponding central Relation; similarly Relations around contain additional information.

KW-ID = “Janet Leverling”

---

[Fakas, DKE, 2011]
2.1 Object Summaries

OS Generation - Methodology

KW-ID = “Janet Leverling”

- **T<sup>DS</sup>** a central tuple containing the Kw; tuples around T<sup>DS</sup> contain additional information about the Data Subject.
- **R<sup>DS</sup>** the corresponding central Relation; similarly Relations around contain additional information.

[Fakas, DKE, 2011]
2.1 Object Summaries

**OS Generation - Methodology**

KW-ID = “Janet Leverling”

- \(t^{DS}\) a central tuple containing the \(Kw\); tuples around \(t^{DS}\) contain additional information about the Data Subject.

- \(R^{DS}\) the corresponding central Relation; similarly Relations around contain additional information.

[Fakas, DKE, 2011]
2.1 Object Summaries

**OS Generation - Methodology**

KW-ID = “Janet Leverling”

### Territories

<table>
<thead>
<tr>
<th>TerritoryID</th>
<th>TerritoryDescription</th>
<th>RegionID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rockville</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Oceanboro</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Atlanta</td>
<td>4</td>
</tr>
</tbody>
</table>

### Region

<table>
<thead>
<tr>
<th>RegionID</th>
<th>RegionDescription</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eastern</td>
</tr>
<tr>
<td>2</td>
<td>Southern</td>
</tr>
</tbody>
</table>

### Employees

<table>
<thead>
<tr>
<th>EmployeeID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Title</th>
<th>TitleOfCntry</th>
<th>Address</th>
<th>ReportsTo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Davis</td>
<td>Nancy</td>
<td>Sales Representative</td>
<td></td>
<td>507 - 20th Ave, Apt. 1A 2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Fuller</td>
<td>Andrew</td>
<td>Vice President, Sales</td>
<td></td>
<td>409 E. Capital Way</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Levering</td>
<td>Janet</td>
<td>Sales Representative</td>
<td></td>
<td>712 Rose Bay Dr.</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Peacock</td>
<td>Margaret</td>
<td>Sales Representative</td>
<td></td>
<td>610 Old Redwood Rd.</td>
<td>2</td>
</tr>
</tbody>
</table>

### Orders

<table>
<thead>
<tr>
<th>OrderID</th>
<th>CustomerID</th>
<th>EmployeeID</th>
<th>ProductID</th>
<th>Quantity</th>
<th>Discount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1073</td>
<td>3</td>
<td>1</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1009</td>
<td>4</td>
<td>2</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

### Customers

<table>
<thead>
<tr>
<th>CustomerID</th>
<th>ContactName</th>
<th>Address</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eron</td>
<td>Roland</td>
<td>505-5999</td>
</tr>
</tbody>
</table>

### Shippers

<table>
<thead>
<tr>
<th>ShipmentID</th>
<th>CompanyName</th>
<th>Address</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Federal Shipping</td>
<td>505-5999</td>
<td>555-1193</td>
</tr>
</tbody>
</table>

### Products

<table>
<thead>
<tr>
<th>ProductID</th>
<th>Quantity</th>
<th>Discount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

### Suppliers

<table>
<thead>
<tr>
<th>SupplierID</th>
<th>ContactName</th>
<th>Address</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robert</td>
<td>505-5999</td>
<td>New York</td>
</tr>
</tbody>
</table>

### Categories

<table>
<thead>
<tr>
<th>CategoryID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Beverages</td>
</tr>
</tbody>
</table>

---

**OS for “Janet Leverling”**

- Employees
- Territories
- Region
- Orders
- Customers
- CustomerDemographics
- CustomerCustomerDemo
- Shippers
- Order Details
- Products
- Suppliers
- Categories
2.3 Motivation

Ranking of Size-1 OSs

Query: *identifying keyword*: Chen

For the DBLP dataset, there 1,982 OSs, i.e. 1,982 authors having the name “Chen”.

Using Authoritative ranking, Peter Chen will always be ranked first because of his many citations. This is ineffective for users who search for a DS that does not have the best importance scores.

In view of this, in this paper, we propose the thematic ranking of OSs, where thematic keywords are also input by the user.
2.3 Motivation

Thematic Ranking of Size-l OSs

**Query:** identifying *keyword*: Chen

**thematic keyword:** Mining

The additional thematic keyword makes 'Ming-Syan Chen' previal, since his OS contains 'Mining' many times.

<table>
<thead>
<tr>
<th>Author: Ming-Syan Chen</th>
<th>[1.00, 0.45]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper: A robust and efficient clustering algorithm based...</td>
<td></td>
</tr>
<tr>
<td>Co-Author: Cheng-Ru Lin. Conf.: KDD. Year: 2002</td>
<td></td>
</tr>
<tr>
<td>Paper: Distributed data <em>mining</em> in a chain store... [0.98, 0.16]</td>
<td></td>
</tr>
<tr>
<td>Co-Author: Philip S. Yu, ... Conf.: KDD. Year: 2002.</td>
<td></td>
</tr>
<tr>
<td>Paper: <em>Mining</em> Relationship between Triggering...[0.98, 0.15]</td>
<td></td>
</tr>
<tr>
<td>Co-Author: Philip S. Yu, ... Conf.: SDM. Year: 2002.</td>
<td></td>
</tr>
<tr>
<td>Paper: DOMISA: DOM-Based infor... <em>Mining</em>... [0.98, 0.18]</td>
<td></td>
</tr>
<tr>
<td>Co-Author: Hung-Yu Kao, ... Conf.: SDM, Year: 2004.</td>
<td></td>
</tr>
<tr>
<td>Paper: Efficient ... <em>Mining</em> for Association Rules...[0.98, 0.19]</td>
<td></td>
</tr>
<tr>
<td>Co-Author: Philip S. Yu, ... Conf.: CIKM Year: 1995.</td>
<td></td>
</tr>
<tr>
<td>Cited by: Parallel <em>Mining</em> of Association... [0.93, 0.36]</td>
<td></td>
</tr>
<tr>
<td>Cited by: Data <em>Mining</em>: An Overview from... [0.93, 0.82]</td>
<td></td>
</tr>
<tr>
<td>Cited by: Dynamic Load Balan... <em>Mining</em>... [0.93, 0.16]</td>
<td></td>
</tr>
<tr>
<td>Cited by: Parallel <em>Mining</em> Algorithms for G... [0.93, 0.16]</td>
<td></td>
</tr>
<tr>
<td>……..</td>
<td></td>
</tr>
</tbody>
</table>
Outline

1. Motivation
2. Background & Related work
3. Themtiac Size-\(l\) OSs
4. Approaches
5. Evaluation Results
6. Conclusion & Future Work
3 Themtiac Size-l OSs

**Definition:**
A query $Q$ comprises **two sets of keywords**, $Q=\langle q_1, q_2 \rangle$,
- $q_1$ is a set of **identifying keywords**
- $q_2$ is a set of **thematic keywords**

**Criteria:**
1. **global Importance**;
2. **IR-properties**; and
3. **Affinity**

---

**Author:** Ming-Syan Chen  
**Paper:** A robust and efficient clustering algorithm based...  
**Co-Author:** Cheng-Ru Lin.  
**Conf.:** KDD.  
**Year:** 2002

**Paper:** Distributed data mining in a chain store...  
**Co-Author:** Philip S. Yu, ...  
**Conf.:** KDD.  
**Year:** 2002.

**Paper:** Mining Relationship between Triggering...  
**Co-Author:** Philip S. Yu, ...  
**Conf.:** SDM.  
**Year:** 2002.

**Paper:** DOMISA: DOM-Based infor...  
**Co-Author:** Hung-Yu Kao, ...  
**Conf.:** SDM.  
**Year:** 2004.

**Paper:** Efficient ... Mining for Association Rules...  
**Co-Author:** Philip S. Yu, ...  
**Conf.:** CIKM  
**Year:** 1995.

**Cited by:** Parallel Mining of Association...  
**Cited by:** Data Mining: An Overview from...  
**Cited by:** Dynamic Load Balan.. Mining...  
**Cited by:** Parallel Mining Algorithms for G...
3 Themtiac Size-l OSs

**Definition:**
A query $Q$ comprises two sets of keywords, $Q=\langle q_1, q_2 \rangle$,
- $q_1$ is a set of identifying keywords
- $q_2$ is a set of thematic keywords

**Criteria:**
1. *global Importance*;  
2. IR-properties; and 
3. Affinity

$$score(O, q_1) = Im(t^{D_5}),$$

$$score(O, q_2) = \frac{\sum_{t \in O} s(t, q_2)}{1 - \alpha + \alpha \cdot \frac{dl(O)}{avdl(OS)}}$$

$$score(O, Q) = score_1(O, q_1) \cdot score_2(O, q_2)$$

where

$$s(t, q_2) = \sum_{w \in q_2} (1 + \ln(1 + \ln(tf_w(t)))) \cdot \ln(idf_w) \cdot li(t),$$

$$idf_w = \frac{N_{OS} + 1}{df_w(OS)},$$

$$li(t) = Af(t) \cdot Im(t).$$
1. Motivation
2. Related work
3. Themtiac Size-$l$ OSs
4. Approaches
5. Evaluation Results
6. Conclusion & Future Work
4.1 Problem reformulation

We reformulate our OSs ranking problem as a *top-k* Group By join problem (*kGBJ*).

Considering two selection operations on $R^{DS}$ and $R^{TH}$ then we get $R^{DS}(q_1)$ and $R^{TH}(q_2)$.

![Diagram showing the DBLP Author $G^{DS}$ (Affinity).]

<table>
<thead>
<tr>
<th>Identifying Keywords ($q_1$)</th>
<th>Frequency in DBLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>David</td>
<td>4,235</td>
</tr>
<tr>
<td>Chen</td>
<td>1,982</td>
</tr>
<tr>
<td>Wang</td>
<td>1,778</td>
</tr>
<tr>
<td>Alan</td>
<td>660</td>
</tr>
<tr>
<td>John</td>
<td>3,717</td>
</tr>
<tr>
<td>Nick</td>
<td>179</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thematic Keywords ($q_2$)</th>
<th>Frequency in DBLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining</td>
<td>2,961</td>
</tr>
<tr>
<td>System</td>
<td>32,253</td>
</tr>
<tr>
<td>Logic</td>
<td>65</td>
</tr>
<tr>
<td>Data</td>
<td>22,500</td>
</tr>
</tbody>
</table>
4.1 Baseline: Bi-Directional approach

**BD approach**

As in a query optimizer, given the sizes of $R_{DS}(q_1)$ and $R_{TH}(q_2)$, the estimation of the optimal **meeting point** is done with the help of statistics.

**Meeting point Examples**
4.2 Top-

$k$ Bi-Directional approach

$kBD$ approach

Rationale of this approach is to avoid the entire BD traversal and processing of our input (i.e. of $R^{DS}(q_1)$ and $R^{TH}(q_2)$).

We achieve this by estimating upper and lower bounds for each OS and by managing them in descending order of their upper bounds in a max-heap.
1. Motivation
2. Background & Related work
3. Themtiac Size-$l$ OSs
4. Approaches
5. Evaluation Results
6. Conclusion & Future Work
5. Experimental Evaluation

Effectiveness

<table>
<thead>
<tr>
<th>$k$</th>
<th>Precision (=Recall)</th>
<th>Ranking Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>92.0%</td>
<td>0.84</td>
</tr>
<tr>
<td>10</td>
<td>96.5%</td>
<td>0.92</td>
</tr>
<tr>
<td>15</td>
<td>98.8%</td>
<td>0.96</td>
</tr>
<tr>
<td>20</td>
<td>100%</td>
<td>0.98</td>
</tr>
<tr>
<td>25</td>
<td>100%</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Precision(Recall) and Ranking Correlation
5. Experimental Evaluation

Efficiency

Efficiency of $BD$ and $kBD$
for Various Values of $k$
1. Motivation
2. Background & Related work
3. Themtiac Size-l OSs
4. Approaches
5. Evaluation Results
6. Conclusion & Future Work
6.1 Conclusion & Future Work

*Contributions*

- The formal definition of *thematic ranking object summaries* for keyword search.

- The an efficient *top-k group-by join* algorithm.

- Applications: Google, Google Desktop, DBMS, etc.
Thank you

Questions!
4.2 Top-$k$ Bi-Directional approach

$kBD$ approach

Upper and lower bounds of OSs are calculated as follows:

$$LB(O) = \ln(O) \cdot \sum_{j=1}^{l} n_j \cdot s(t_j)$$

$$UB_1(O, Q) = LB(O) + \ln(O) \cdot (M - \sigma) \cdot s(t_{l+1})$$

tighter upper bound

$$UB_2(O, Q) = LB(O) + \ln(O) \left( \sum_{j=1}^{d} s(t_{l+j}) \cdot m + s(t_{l+d+1}) \cdot r \right)$$

$$d = \lfloor \gamma/m \rfloor, \quad r = \gamma \mod m$$

further tighten

$$UB_3(O, Q) = LB(O) + \ln(O) \left( \sum_{j=1}^{d} m \cdot s(t_{a_j}) + r \cdot s(t_{a_{d+1}}) \right)$$
4.2 Top-\(k\) Bi-Directional approach

**kBD approach**

**Algorithm 1. kBD Algorithm**

```
kBD \((R^{DS}(q_1), R^{Th}(q_2), k)\)
1: \(H := \emptyset\);
2: \(L^{Th} := R^{Th}(q_2)\);
3: sort tuples in \(L^{Th}\) in descending order of their \(s(\cdot)\) scores;
4: for each \(O\) w.r.t. \(t^{DS}\) in \(R^{DS}(q_1)\) do
5: \(LB(O):=0; UB(O):=\text{calcub } O;\)
6: insert \(O\) into \(H\) with priority \(UB(O)\);
7: while \(k > 0 \land H\) is not empty do
8: pop \(O_{cur}\) from \(H\);
9: \(O_{next} := H.\text{top}()\);
10: if \(LB(O_{cur}) \geq UB(O_{next})\) then
11: report \(O_{cur}\) as a result;
12: \(k := k - 1\);
13: else
14: \(t_i := \text{the next tuple in } L^{Th}\) can join with \(O_{cur}\);
15: \(n := \text{join } O_{cur}, t_i\);
16: \(LB(O_{cur}) := LB(O_{cur}) + n.\text{Join}(O_{cur}) \cdot s(t_i, q_2)\);
17: \(UB(O_{cur}) := \text{calcub } O_{cur}\);
18: push \(O_{cur}\) back into \(H\) with priority \(UB(O_{cur})\);
```

Calculating the Upper Bound Scores of an OS \(O\) \((In(O) = 1.0, M = 13, m = 4)\)
### 4.2 Top-\(k\) Bi-Directional approach

**kBD approach**

#### Algorithm 1. kBD Algorithm

\[
kBD (R^{DS}(q_1), R^{Th}(q_2), k)
\]

1: \(H:=\emptyset;\)
2: \(L^{Th}:=R^{Th}(q_2);\)
3: sort tuples in \(L^{Th}\) in descending order of their \(s(\cdot)\) scores;
4: for each \(O\) w.r.t. \(t^{DS}\) in \(R^{DS}(q_1)\) do
5: \(LB(O):=0; UB(O):=\text{CALCUB } O;\)
6: insert \(O\) into \(H\) with priority \(UB(O)\);
7: while \(k > 0 \land H\) is not empty do
8: pop \(O_{\text{cur}}\) from \(H;\)
9: \(O_{\text{next}}:=H.\text{top}();\)
10: if \(LB(O_{\text{cur}}) \geq UB(O_{\text{next}})\) then
11: report \(O_{\text{cur}}\) as a result;
12: \(k:=k-1;\)
13: else
14: \(t_i:=\text{the next tuple in } L^{Th}\) can join with \(O_{\text{cur}};\)
15: \(n:=\text{JOIN } O_{\text{cur}}, t_i;\)
16: \(LB(O_{\text{cur}}):=LB(O_{\text{cur}})+n.\text{-Ind}(O_{\text{cur}})\cdot s(t_i, q_2);\)
17: \(UB(O_{\text{cur}}):=\text{CALCUB } O_{\text{cur}};\)
18: push \(O_{\text{cur}}\) back into \(H\) with priority \(UB(O_{\text{cur}});\)

#### The kBD Algorithm for \(k = 1\)

<table>
<thead>
<tr>
<th>OS</th>
<th>(UB(.))</th>
<th>(LB(.))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O_1)</td>
<td>8.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_2)</td>
<td>6.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_3)</td>
<td>10.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_4)</td>
<td>5.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_5)</td>
<td>4.0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(H=(O_3, O_1, O_2, O_4, O_5), O_{\text{cur}}=O_3, O_{\text{next}}=O_1;\)
(a) Initialization

<table>
<thead>
<tr>
<th>OS</th>
<th>(UB(.))</th>
<th>(LB(.))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O_1)</td>
<td>8.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_2)</td>
<td>6.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_3)</td>
<td>7.0</td>
<td>6.5</td>
</tr>
<tr>
<td>(O_4)</td>
<td>5.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_5)</td>
<td>4.0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(H=(O_1, O_3, O_2, O_4, O_5), O_{\text{cur}}=O_1, O_{\text{next}}=O_3;\)
(b) Iteration 1

<table>
<thead>
<tr>
<th>OS</th>
<th>(UB(.))</th>
<th>(LB(.))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O_1)</td>
<td>7.5</td>
<td>5.0</td>
</tr>
<tr>
<td>(O_2)</td>
<td>6.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_3)</td>
<td>7.0</td>
<td>6.5</td>
</tr>
<tr>
<td>(O_4)</td>
<td>5.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_5)</td>
<td>4.0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(H=(O_1, O_3, O_2, O_4, O_5), O_{\text{cur}}=O_1, O_{\text{next}}=O_3;\)
(c) Iteration 2

<table>
<thead>
<tr>
<th>OS</th>
<th>(UB(.))</th>
<th>(LB(.))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O_1)</td>
<td>6.2</td>
<td>6.0</td>
</tr>
<tr>
<td>(O_2)</td>
<td>6.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_3)</td>
<td>7.0</td>
<td>6.5</td>
</tr>
<tr>
<td>(O_4)</td>
<td>5.0</td>
<td>0</td>
</tr>
<tr>
<td>(O_5)</td>
<td>4.0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(H=(O_3, O_1, O_2, O_4, O_5), O_{\text{cur}}=O_3, O_{\text{next}}=O_1;\)
(d) Iteration 3
4.3 Multiple thematic relations

*Holistic Top-k BD (HkBD) algorithm*

Given $j$ thematic relations $R_1^{Th}, ..., R_j^{Th}$, we can extend analogously the original $kBD$ algorithm by defining appropriate upper and lower bound scores for each $DS$. We can easily see that the sum of the upper (resp. lower) bound scores of all join paths (denote as $JP$) is the upper (resp. lower) bound score of an OS, namely:

$$UB^H(O, Q) = \sum_{JP_i} UB^{JP_i}(O, Q),$$

$$LB^H(O, Q) = \sum_{JP_i} LB^{JP_i}(O, Q),$$

where $JP_i$ ranges over all thematic paths and $UB^{JP_i}(\cdot)$ (resp.$UB^{JP_i}(\cdot)$) is the upper (resp. lower) bound score of $O$. 
### 4.3 Multiple thematic relations

**Holistic Top-k BD (HkBD) algorithm**

We force the meeting point to the common $G^D_S$ prefix which is shared by all paths, then we can compute the **join result once** and reuse if later for the other paths;

*HkBD* approach is advantageous in this aspect over the *HBD* algorithm, as it facilitates reuse of join results.

<table>
<thead>
<tr>
<th>ID</th>
<th>Path Name</th>
<th>20%</th>
<th>DS</th>
<th>20%</th>
<th>Th</th>
</tr>
</thead>
<tbody>
<tr>
<td>$JP_1$</td>
<td>DBLP: Author-Paper-ConfYear-Conference</td>
<td>68K</td>
<td></td>
<td>593</td>
<td></td>
</tr>
<tr>
<td>$JP_2$</td>
<td>DBLP: Author-Paper-PaperCitedBy</td>
<td>68K</td>
<td></td>
<td>103K</td>
<td></td>
</tr>
<tr>
<td>$JP_3$</td>
<td>DBLP: Author-Paper-PaperCites</td>
<td>68K</td>
<td></td>
<td>103K</td>
<td></td>
</tr>
<tr>
<td>$JP_4$</td>
<td>TPC-H: Cust.-Ord.-Lineitem-Partsupp-Part</td>
<td>30K</td>
<td></td>
<td>400K</td>
<td></td>
</tr>
<tr>
<td>$JP_5$</td>
<td>TPC-H: Cust.-Ord.-Lineitem-Partsupp-Supplier</td>
<td>30K</td>
<td></td>
<td>2K</td>
<td></td>
</tr>
</tbody>
</table>

**Examples of Join Paths**