Introduction
IR and Databases
The Logic View

Retrieval

- DB: given query $q$, find objects $o$ with $o \rightarrow q$
- IR: given query $q$, find documents $d$ with high values of $P(d \rightarrow q)$
- DB is a special case of IR!
  (in a certain sense)
IR and Databases
The Logic View

Retrieval
- DB: given query $q$, find objects $o$ with $o \rightarrow q$
- IR: given query $q$, find documents $d$ with high values of $P(d \rightarrow q)$
- DB is a special case of IR! (in a certain sense)

This tutorial: Focusing on the logic view
- Inference
- Vague predicates
- Query language expressiveness
Inference

- IR with the Relational Model
- The Probabilistic Relational Model
- Interpretation of probabilistic weights
- Extensions
  - Disjoint events
  - Relational Bayes
  - Probabilistic rules
Relational Model
Projection

<table>
<thead>
<tr>
<th>DOCNO</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ir</td>
</tr>
<tr>
<td>1</td>
<td>db</td>
</tr>
<tr>
<td>2</td>
<td>ir</td>
</tr>
<tr>
<td>3</td>
<td>db</td>
</tr>
<tr>
<td>3</td>
<td>oop</td>
</tr>
<tr>
<td>4</td>
<td>ir</td>
</tr>
<tr>
<td>4</td>
<td>ai</td>
</tr>
<tr>
<td>5</td>
<td>db</td>
</tr>
<tr>
<td>5</td>
<td>oop</td>
</tr>
</tbody>
</table>

**Projection:** what is the collection about?

$\text{topic}(T) :- \text{index}(D,T)$. 
### Relational Model

#### Selection

<table>
<thead>
<tr>
<th>DOCNO</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ir</td>
</tr>
<tr>
<td>1</td>
<td>db</td>
</tr>
<tr>
<td>2</td>
<td>ir</td>
</tr>
<tr>
<td>3</td>
<td>db</td>
</tr>
<tr>
<td>3</td>
<td>oop</td>
</tr>
<tr>
<td>4</td>
<td>ir</td>
</tr>
<tr>
<td>4</td>
<td>ai</td>
</tr>
<tr>
<td>5</td>
<td>db</td>
</tr>
<tr>
<td>5</td>
<td>oop</td>
</tr>
</tbody>
</table>

**Selection:** which documents are about IR?

\[
\text{aboutir}(D) :- \text{index}(D, \text{ir}).
\]
Join: who writes about IR?

`irauthor(A):- index(D,ir) & author(D,A).`
Bridging IR and DBs
Inference
IR with the Relational Model

Relational Model
Union

index

<table>
<thead>
<tr>
<th>DOCNO</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ir</td>
</tr>
<tr>
<td>1</td>
<td>db</td>
</tr>
<tr>
<td>2</td>
<td>ir</td>
</tr>
<tr>
<td>3</td>
<td>db</td>
</tr>
<tr>
<td>3</td>
<td>oop</td>
</tr>
<tr>
<td>4</td>
<td>ir</td>
</tr>
<tr>
<td>4</td>
<td>ai</td>
</tr>
<tr>
<td>5</td>
<td>db</td>
</tr>
<tr>
<td>5</td>
<td>oop</td>
</tr>
</tbody>
</table>

Union: which documents are about IR or DB?

\[
\text{irordb}(D) :- \text{index}(D,\text{ir}). \\
\text{irordb}(D) :- \text{index}(D,\text{db}).
\]
**Relational Model**

**Difference**

<table>
<thead>
<tr>
<th>DOCNO</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ir</td>
</tr>
<tr>
<td>1</td>
<td>db</td>
</tr>
<tr>
<td>2</td>
<td>ir</td>
</tr>
<tr>
<td>3</td>
<td>db</td>
</tr>
<tr>
<td>3</td>
<td>oop</td>
</tr>
<tr>
<td>4</td>
<td>ir</td>
</tr>
<tr>
<td>4</td>
<td>ai</td>
</tr>
<tr>
<td>5</td>
<td>db</td>
</tr>
<tr>
<td>5</td>
<td>oop</td>
</tr>
</tbody>
</table>

**Difference:** which documents are about IR, but not DB?

\[ \text{irnotdb}(D) :\text{:- index}(D,\text{ir}) \& \text{not}(\text{index}(D,\text{db})). \]
The Probabilistic Relational Model

[Fuhr & Roelleke 97] [Suciu et al 11]

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>DOCNO</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>1</td>
<td>IR</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>DB</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>IR</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>DB</td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>OOP</td>
</tr>
<tr>
<td>0.9</td>
<td>4</td>
<td>IR</td>
</tr>
<tr>
<td>0.4</td>
<td>4</td>
<td>AI</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>DB</td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>OOP</td>
</tr>
</tbody>
</table>
The Probabilistic Relational Model

[Fuhr & Roelleke 97] [Suciu et al 11]

index 

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>DOCNO</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>1</td>
<td>IR</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>DB</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>IR</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>DB</td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>OOP</td>
</tr>
<tr>
<td>0.9</td>
<td>4</td>
<td>IR</td>
</tr>
<tr>
<td>0.4</td>
<td>4</td>
<td>AI</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>DB</td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>OOP</td>
</tr>
</tbody>
</table>

Which documents are about DB?

aboutdb(D) :- index(D,db).
The Probabilistic Relational Model

[Fuhr & Roelleke 97] [Suciu et al 11]

index

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>DOCNO</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>1</td>
<td>IR</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>DB</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>IR</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>DB</td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>OOP</td>
</tr>
<tr>
<td>0.9</td>
<td>4</td>
<td>IR</td>
</tr>
<tr>
<td>0.4</td>
<td>4</td>
<td>AI</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>DB</td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>OOP</td>
</tr>
</tbody>
</table>

aboutdb(D) :- index(D,db).

Which documents are about DB?
The Probabilistic Relational Model

[Fuhr & Roelleke 97] [Suciu et al 11]

index

<table>
<thead>
<tr>
<th>β</th>
<th>DOCNO</th>
<th>TERM</th>
<th>aboutdb</th>
<th>aboutirdb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>1</td>
<td>IR</td>
<td>0.7</td>
<td>0.8*0.7</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>DB</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>IR</td>
<td>0.5</td>
<td>0.8*0.7</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>DB</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>OOP</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0.9</td>
<td>4</td>
<td>IR</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>4</td>
<td>AI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>DB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>OOP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Which documents are about DB?
aboutdb(D) :- index(D,db).

Which documents are about IR and DB?
aboutirdb(D) :- index(D,ir) & index(D,db).
## Extensional vs. intensional semantics

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>DOC</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>d1</td>
<td>ir</td>
</tr>
<tr>
<td>0.5</td>
<td>d1</td>
<td>db</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>S</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>d2</td>
<td>d1</td>
</tr>
</tbody>
</table>

\[
q(D) \leftarrow \text{about}(D, \text{ir}) \& \text{about}(D, \text{db}).
\]

\[
\text{about}(D,T) \leftarrow \text{docTerm}(D,T).
\]

\[
\text{about}(D,T) \leftarrow \text{link}(D,D1) \& \text{about}(D1,T)
\]

Problem: “improper treatment of correlated sources of evidence” [Pearl 88] → extensional semantics only correct for tree-shaped inference structures
## Extentional vs. intensional semantics

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>DOC</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>d1</td>
<td>ir</td>
</tr>
<tr>
<td>0.5</td>
<td>d1</td>
<td>db</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>S</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>d2</td>
<td>d1</td>
</tr>
</tbody>
</table>

- `about(D,T) :- docTerm(D,T).`
- `about(D,T) :- link(D,D1) & about(D1,T)`
- `q(D) :- about(D,ir) & about(D,db).`

**extentional semantics:**

Weight of derived fact as function of weights of subgoals:

$$P(q(d2)) = P(about(d2,ir)) \cdot P(about(d2,db)) = (0.7 \cdot 0.9) \cdot (0.7 \cdot 0.5)$$
### Extensional vs. intensional semantics

<table>
<thead>
<tr>
<th>docterm</th>
<th>link</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>DOC</td>
</tr>
<tr>
<td>0.5</td>
<td>d1</td>
</tr>
<tr>
<td></td>
<td>d1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>link</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>d2</td>
</tr>
<tr>
<td>d1</td>
</tr>
</tbody>
</table>

\[
\text{about}(D,T) := \text{docTerm}(D,T).
\]

\[
\text{about}(D,T) := \text{link}(D,D1) \land \text{about}(D1,T)
\]

\[
\text{q}(D) := \text{about}(D,\text{ir}) \land \text{about}(D,\text{db}).
\]

**extensional semantics:**

weight of derived fact as function of weights of subgoals

\[
P(q(d2)) = P(\text{about}(d2,\text{ir})) \cdot P(\text{about}(d2,\text{db})) = (0.7 \cdot 0.9) \cdot (0.7 \cdot 0.5)
\]

**Problem**

“improper treatment of correlated sources of evidence” [Pearl 88] → extensional semantics only correct for tree-shaped inference structures
Intensional semantics

weight of derived fact as function of weights of underlying ground facts
Intensional semantics

weight of derived fact as function of weights of underlying ground facts

**Method:** Event keys and event expressions

<table>
<thead>
<tr>
<th>docterm</th>
<th>link</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>( \kappa )</td>
</tr>
<tr>
<td>0.9</td>
<td>( d_T(d_1,ir) )</td>
</tr>
<tr>
<td>0.5</td>
<td>( d_T(d_1,db) )</td>
</tr>
</tbody>
</table>
Intensional semantics

weight of derived fact as function of weights of underlying ground facts

Method: Event keys and event expressions

docterm

<table>
<thead>
<tr>
<th>β</th>
<th>κ</th>
<th>DOC</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>dT(d1,ir)</td>
<td>d1</td>
<td>ir</td>
</tr>
<tr>
<td>0.5</td>
<td>dT(d1,db)</td>
<td>d1</td>
<td>db</td>
</tr>
</tbody>
</table>

link

<table>
<thead>
<tr>
<th>β</th>
<th>κ</th>
<th>S</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>l(d2,d1)</td>
<td>d2</td>
<td>d1</td>
</tr>
</tbody>
</table>

?- docTerm(D,ir) & docTerm(D,db).

gives

d1 \[dT(d1,ir) \& dT(d1,db)]
Intensional semantics

weight of derived fact as function of weights of underlying ground facts

**Method:** Event keys and event expressions

docterm

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\kappa$</th>
<th>DOC</th>
<th>TERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>$dT(d1,ir)$</td>
<td>$d1$</td>
<td>$ir$</td>
</tr>
<tr>
<td>0.5</td>
<td>$dT(d1,db)$</td>
<td>$d1$</td>
<td>$db$</td>
</tr>
</tbody>
</table>

link

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\kappa$</th>
<th>S</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>$l(d2,d1)$</td>
<td>$d2$</td>
<td>$d1$</td>
</tr>
</tbody>
</table>

?- docTerm(D,ir) & docTerm(D,db).
gives
d1 $[dT(d1,ir) \& dT(d1,db)]$ $0.9 \cdot 0.5 = 0.45$
Event keys and event expressions

\[
\begin{array}{c|c|c|c|c}
\beta & \kappa & \text{DOC} & \text{TERM} & \beta & \kappa & S & T \\
0.9 & dT(d1,ir) & d1 & ir & 0.7 & l(d2,d1) & d2 & d1 \\
0.5 & dT(d1,db) & d1 & db & & & & \\
\end{array}
\]

\[
\text{about}(D,T) :- \text{docTerm}(D,T).
\]
\[
\text{about}(D,T) :- \text{link}(D,D1) \& \text{about}(D1,T)
\]
\[
?\:- \text{about}(D,\text{ir}) \& \text{about}(D,\text{db}).
\]

\[
gives \quad d1 \ [dT(d1,ir) \& dT(d1,db)] \quad 0.9 \cdot 0.5 = 0.45
\]
\[
d2 \ [l(d2,d1) \& dT(d1,ir) \& l(d2,d1) \& dT(d1,db)] \quad 0.7 \cdot 0.9 \cdot 0.5 = 0.315
\]
Recursion

about(D,T) :- docTerm(D,T).
about(D,T) :- link(D,D1) & about(D1,T).

?- about(D,ir)
d1 [dT(d1,ir) | l(d1,d2) & l(d2,d3) & l(d3,d1) &
   dT(d1,ir) | ...] 0.900

d3 [l(d3,d1) & dT(d1,ir)] 0.720

d2 [l(d2,d3) & l(d3,d1) & dT(d1,ir)] 0.288

?- about(D,ir) & about(D,db)
d1 [dT(d1,ir) & dT(d1,db)] 0.450

d3 [l(d3,d1) & dT(d1,ir) & l(d3,d1) & dT(d1,db)] 0.360
Computation of probabilities for event expressions

1. Transformation of expression into disjunctive normal form
2. Application of sieve formula:
   - Simple case of 2 conjuncts: \( P(a \lor b) = P(a) + P(b) - P(a \land b) \)
Computation of probabilities for event expressions

1. transformation of expression into disjunctive normal form
2. application of sieve formula:
   - simple case of 2 conjuncts: $P(a \lor b) = P(a) + P(b) - P(a \land b)$
   - general case:
     $c_i$ – conjunct of event keys

\[
P(c_1 \lor \ldots \lor c_n) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{1 \leq j_1 < \ldots < j_i \leq n} P(c_{j_1} \land \ldots \land c_{j_i}).
\]

- $\rightsquigarrow$ exponential complexity
Computation of probabilities for event expressions

1. transformation of expression into disjunctive normal form
2. application of sieve formula:
   - simple case of 2 conjuncts: \( P(a \lor b) = P(a) + P(b) - P(a \land b) \)
   - general case:
     \[ P(c_1 \lor \ldots \lor c_n) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{1 \leq j_1 < \ldots < j_i \leq n} P(c_{j_1} \land \ldots \land c_{j_i}). \]

\( \Rightarrow \) exponential complexity

\( \Rightarrow \) use only when necessary for correctness
Computation of probabilities for event expressions

1. transformation of expression into disjunctive normal form
2. application of sieve formula:
   - simple case of 2 conjuncts: \( P(a \lor b) = P(a) + P(b) - P(a \land b) \)
   - general case:
     \( c_i \) – conjunct of event keys

\[
P(c_1 \lor \ldots \lor c_n) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{1 \leq j_1 < \ldots < j_i \leq n} P(c_{j_1} \land \ldots \land c_{j_i}).
\]

- \( \rightsquigarrow \) exponential complexity
- \( \rightsquigarrow \) use only when necessary for correctness
- see [Dalvi & Suciu 07]
Possible worlds semantics

0.9 \text{docTerm}(d1, \text{ir}).

P(W_1) = 0.9: \{\text{docTerm}(d1, \text{ir})\}

P(W_2) = 0.1: \{\}
0.6 docTerm(d1,ir). 0.5 docTerm(d1,db).

Possible interpretations:

$I_1$: $P(W_1) = 0.3$: \{docTerm(d1,ir)\}
$P(W_2) = 0.3$: \{docTerm(d1,ir), docTerm(d1,db)\}
$P(W_3) = 0.2$: \{docTerm(d1,db)\}
$P(W_4) = 0.2$: {} 

$I_2$: $P(W_1) = 0.5$: \{docTerm(d1,ir)\}
$P(W_2) = 0.1$: \{docTerm(d1,ir), docTerm(d1,db)\}
$P(W_3) = 0.4$: \{docTerm(d1,db)\}

$I_3$: $P(W_1) = 0.1$: \{docTerm(d1,ir)\}
$P(W_2) = 0.5$: \{docTerm(d1,ir), docTerm(d1,db)\}
$P(W_3) = 0.4$: {}
0.6 \text{docTerm}(d1, \text{ir}). 0.5 \text{docTerm}(d1, \text{db}).

\textbf{Possible interpretations:}

\begin{align*}
I_1: & \quad P(W_1) = 0.3: \{\text{docTerm}(d1, \text{ir})\} \\
& \quad P(W_2) = 0.3: \{\text{docTerm}(d1, \text{ir}), \text{docTerm}(d1, \text{db})\} \\
& \quad P(W_3) = 0.2: \{\text{docTerm}(d1, \text{db})\} \\
& \quad P(W_4) = 0.2: \{\} \\
I_2: & \quad P(W_1) = 0.5: \{\text{docTerm}(d1, \text{ir})\} \\
& \quad P(W_2) = 0.1: \{\text{docTerm}(d1, \text{ir}), \text{docTerm}(d1, \text{db})\} \\
& \quad P(W_3) = 0.4: \{\text{docTerm}(d1, \text{db})\} \\
I_3: & \quad P(W_1) = 0.1: \{\text{docTerm}(d1, \text{ir})\} \\
& \quad P(W_2) = 0.5: \{\text{docTerm}(d1, \text{ir}), \text{docTerm}(d1, \text{db})\} \\
& \quad P(W_3) = 0.4: \{\} \\
\end{align*}

\textbf{probabilistic logic:}
0.1 \leq P(\text{docTerm}(d1, \text{ir}) \& \text{docTerm}(d1, \text{db})) \leq 0.5
0.6 \text{docTerm}(d1,ir). 0.5 \text{docTerm}(d1,db).

Possible interpretations:

$l_1$: $P(W_1) = 0.3$: \{docTerm(d1,ir)\}
$P(W_2) = 0.3$: \{docTerm(d1,ir), \text{docTerm}(d1,db)\}
$P(W_3) = 0.2$: \{docTerm(d1,db)\}
$P(W_4) = 0.2$: \{

$l_2$: $P(W_1) = 0.5$: \{docTerm(d1,ir)\}
$P(W_2) = 0.1$: \{docTerm(d1,ir), \text{docTerm}(d1,db)\}
$P(W_3) = 0.4$: \{docTerm(d1,db)\}

$l_3$: $P(W_1) = 0.1$: \{docTerm(d1,ir)\}
$P(W_2) = 0.5$: \{docTerm(d1,ir), \text{docTerm}(d1,db)\}
$P(W_3) = 0.4$: \{

probabilistic logic:

$0.1 \leq P(\text{docTerm}(d1,ir) \& \text{docTerm}(d1,db)) \leq 0.5$

probabilistic Datalog with independence assumptions:

$P(\text{docTerm}(d1,ir) \& \text{docTerm}(d1,db)) = 0.3$
Disjoint events

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>City</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>Paris</td>
<td>France</td>
</tr>
<tr>
<td>0.2</td>
<td>Paris</td>
<td>Texas</td>
</tr>
<tr>
<td>0.1</td>
<td>Paris</td>
<td>Idaho</td>
</tr>
</tbody>
</table>

Interpretation: $P(W_1) = 0.7$: \{cityState(paris, france)\}

$P(W_2) = 0.2$: \{cityState(paris, texas)\}

$P(W_3) = 0.1$: \{cityState(paris, idaho)\}
Disjoint events

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>City</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>Paris</td>
<td>France</td>
</tr>
<tr>
<td>0.2</td>
<td>Paris</td>
<td>Texas</td>
</tr>
<tr>
<td>0.1</td>
<td>Paris</td>
<td>Idaho</td>
</tr>
</tbody>
</table>

**Interpretation:**

$P(W_1) = 0.7$: \{cityState(paris, france)\}

$P(W_2) = 0.2$: \{cityState(paris, texas)\}

$P(W_3) = 0.1$: \{cityState(paris, idaho)\}
[Roelleke et al. 07]

Role of the relational Bayes: Generation of a probabilistic database

Non-probabilistic database  Bayes  Probabilistic database
### Relational Bayes

#### Example: $P(\text{Nationality} \mid \text{City})$

<table>
<thead>
<tr>
<th>nationality_and_city</th>
<th>nationality_city</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nationality</td>
<td>City</td>
</tr>
<tr>
<td>&quot;British&quot;</td>
<td>&quot;London&quot;</td>
</tr>
<tr>
<td>&quot;British&quot;</td>
<td>&quot;London&quot;</td>
</tr>
<tr>
<td>&quot;British&quot;</td>
<td>&quot;London&quot;</td>
</tr>
<tr>
<td>&quot;Scottish&quot;</td>
<td>&quot;London&quot;</td>
</tr>
<tr>
<td>&quot;French&quot;</td>
<td>&quot;London&quot;</td>
</tr>
<tr>
<td>&quot;German&quot;</td>
<td>&quot;Hamburg&quot;</td>
</tr>
<tr>
<td>&quot;Danish&quot;</td>
<td>&quot;Hamburg&quot;</td>
</tr>
<tr>
<td>&quot;British&quot;</td>
<td>&quot;Hamburg&quot;</td>
</tr>
<tr>
<td>&quot;German&quot;</td>
<td>&quot;Dortmund&quot;</td>
</tr>
<tr>
<td>&quot;Turkish&quot;</td>
<td>&quot;Dortmund&quot;</td>
</tr>
<tr>
<td>&quot;Scottish&quot;</td>
<td>&quot;Glasgow&quot;</td>
</tr>
</tbody>
</table>

1. $\# P(\text{Nationality} \mid \text{City})$:
2. $\text{nationality}_\text{city } \text{SUM}(\text{Nat, City}) :$ =
3. $\text{nationality}_\text{and}_\text{city (Nat, City) | (City)}$;
Relational Bayes
Example: $P(t|d)$

<table>
<thead>
<tr>
<th>term</th>
<th>DocId</th>
</tr>
</thead>
<tbody>
<tr>
<td>sailing</td>
<td>doc1</td>
</tr>
<tr>
<td>boats</td>
<td>doc1</td>
</tr>
<tr>
<td>sailing</td>
<td>doc2</td>
</tr>
<tr>
<td>boats</td>
<td>doc2</td>
</tr>
<tr>
<td>sailing</td>
<td>doc2</td>
</tr>
<tr>
<td>east</td>
<td>doc3</td>
</tr>
<tr>
<td>coast</td>
<td>doc3</td>
</tr>
<tr>
<td>sailing</td>
<td>doc3</td>
</tr>
<tr>
<td>sailing</td>
<td>doc4</td>
</tr>
<tr>
<td>boats</td>
<td>doc5</td>
</tr>
</tbody>
</table>

$p_{t,d}$ SPACE(Term, DocId) :-
- term(Term, DocId)
- (DocId);

| $P(t|d)$ | Term | DocId |
|----------|------|-------|
| 0.50     | sailing | doc1  |
| 0.50     | boats   | doc1  |
| 0.33     | sailing | doc2  |
| 0.33     | boats   | doc2  |
| 0.33     | sailing | doc2  |
| 0.33     | east    | doc3  |
| 0.33     | coast   | doc3  |
| 0.33     | sailing | doc3  |
| 1.00     | sailing | doc4  |
| 1.00     | boats   | doc5  |

$p_{t,d}$ SUM(Term, DocId) :-
- term(Term, DocId)
- (DocId);

| $P(t|d)$ | Term | DocId |
|----------|------|-------|
| 0.50     | sailing | doc1  |
| 0.50     | boats   | doc1  |
| 0.67     | sailing | doc2  |
| 0.33     | boats   | doc2  |
| 0.33     | east    | doc3  |
| 0.33     | coast   | doc3  |
| 0.33     | sailing | doc3  |
| 1.00     | sailing | doc4  |
| 1.00     | boats   | doc5  |
Probabilistic rules
Rules for deterministic facts:

\[
\begin{align*}
0.7 \text{ } \text{likes-sports}(X) & :\ - \text{ man}(X). \\
0.4 \text{ } \text{likes-sports}(X) & :\ - \text{ woman}(X). \\
\text{man}(\text{peter}).
\end{align*}
\]
Probabilistic rules
Rules for deterministic facts:

0.7 \(\text{likes-sports}(X) :\neg \text{man}(X)\).
0.4 \(\text{likes-sports}(X) :\neg \text{woman}(X)\).

\text{man}(peter).

**Interpretation:**
\[ P(W_1) = 0.7: \{\text{man}(peter), \text{likes-sports}(peter)\} \]
\[ P(W_2) = 0.3: \{\text{man}(peter)\} \]
Probabilistic rules
Rules for uncertain facts:

# gender is disjoint on the first attribute
0.7 l-sports(X) :- gender(X,male).
0.4 l-sports(X) :- gender(X,female).
0.5 gender(X,male) :- human(X).
0.5 gender(X,female) :- human(X).
human(jo).

Interpretation:
\[ P(W_1) = 0.35: \{ \text{gender(jo,male), l-sports(jo)} \} \]
\[ P(W_2) = 0.15: \{ \text{gender(jo,male)} \} \]
\[ P(W_3) = 0.20: \{ \text{gender(jo,female), l-sports(jo)} \} \]
\[ P(W_4) = 0.30: \{ \text{gender(jo,female)} \} \]

?- l-sports(jo)
\[ P(W_1) + P(W_3) = 0.55 \]
Probabilistic rules
Rules for uncertain facts:

# gender is disjoint on the first attribute
0.7 l-sports(X) :- gender(X,male).
0.4 l-sports(X) :- gender(X,female).
0.5 gender(X,male) :- human(X).
0.5 gender(X,female) :- human(X).
human(jo).

Interpretation:
\[ P(W_1) = 0.35: \{ \text{gender(jo,male), l-sports(jo)} \} \]
\[ P(W_2) = 0.15: \{ \text{gender(jo,male)} \} \]
\[ P(W_3) = 0.20: \{ \text{gender(jo,female), l-sports(jo)} \} \]
\[ P(W_4) = 0.30: \{ \text{gender(jo,female)} \} \]

?- l-sports(jo)
Probabilistic rules
Rules for uncertain facts:

# gender is disjoint on the first attribute
0.7 l-sports(X) :- gender(X,male).
0.4 l-sports(X) :- gender(X,female).
0.5 gender(X,male) :- human(X).
0.5 gender(X,female) :- human(X).
human(jo).

**Interpretation:**

\[ P(W_1) = 0.35: \{ \text{gender}(\text{jo},\text{male}), \text{l-sports}(\text{jo}) \} \]
\[ P(W_2) = 0.15: \{ \text{gender}(\text{jo},\text{male}) \} \]
\[ P(W_3) = 0.20: \{ \text{gender}(\text{jo},\text{female}), \text{l-sports}(\text{jo}) \} \]
\[ P(W_4) = 0.30: \{ \text{gender}(\text{jo},\text{female}) \} \]

?- l-sports(jo) \quad \quad P(W_1) + P(W_3) = 0.55
sameauthor(D1,D2) :- author(D1,X) & author(D2,X).
0.5 link(D1,D2) :- refer(D1,D2).
0.2 link(D1,D2) :- sameauthor(D1,D2).

?? link(D1,D2) :- refer(D1,D2) & sameauthor(D1,D2).

\[ P(l|r), P(l|s) \rightarrow P(l|r \land s)? \]
Rules for independent events
Modeling probabilistic inference networks

0.7 \text{link}(D1,D2) :- \text{refer}(D1,D2) \& \text{sameauthor}(D1,D2).
0.5 \text{link}(D1,D2) :- \text{refer}(D1,D2) \& \text{not}(\text{sameauthor}(D1,D2)).
0.2 \text{link}(D1,D2) :- \text{sameauthor}(D1,D2) \& \text{not}(\text{refer}(D1,D2)).

Probabilistic inference networks,
rules define link matrix
Vague Predicates

- The Logical View on Vague Predicates
- Vague Predicates in IR and Databases
- Probabilistic Modeling of Vague Predicates
Bridging IR and DBs

Vague Predicates

Motivating Example

"lcd tv 46 inch"

Showing 1 - 16 of 3,851 Results

Samsung LN46E550 46-Inch 1080p 60Hz LCD HDTV by Samsung

$879.99 Click for product details
Order in the next 5 hours and get it by Wednesday, Jan 16.
More Buying Choices
$463.80 used & new (14 offers)

Samsung LN46D550 46-Inch 1080p 60Hz LCD HDTV (Black) by Samsung

$899.99 $599.27
Only 15 left in stock - order soon.
More Buying Choices
$599.27 new (4 offers)
$490.00 used (10 offers)

Cheetah Mounts APTMM2B Flush Tilt Dual Hook (1.3" from wall) Flat Screen Cheetah

$40.99 $27.99
Order in the next 7 hours and get it by Wednesday, Jan 16.
More Buying Choices
$27.99 new (9 offers)
Vague Predicates
Motivating Example

"lcd tv 45inch"

Showing 1 - 16 of 2,617 Results

RCA 32LB45RQ 32-Inch Full 1080p 60Hz LCD HDTV by RCA
$228.38 used (4 offers)

RCA 42LB45RQ 42-Inch 1080p 60Hz LCD HDTV (Black) by RCA
$476.99
Only 1 left in stock - order soon.
More Buying Choices
$476.99 new (2 offers)
$333.67 used (3 offers)

RCA 22LB45RQD 22-Inch Full 1080p LCD/DVD Combo HDTV by RCA
$299.99 $219.99
Only 1 left in stock - order soon.
More Buying Choices
$188.99 new (3 offers)
$125.00 used (19 offers)
Propositional vs. Predicate Logic

Current IR systems are based on proposition logic (query term present/absent in document)
Propositional vs. Predicate Logic

- Current IR systems are based on proposition logic (query term present/absent in document)
- Similarity of values not considered
Propositional vs. Predicate Logic

- Current IR systems are based on proposition logic (query term present/absent in document)
- Similarity of values not considered
- but multimedia IR deals with similarity already
Propositional vs. Predicate Logic

- Current IR systems are based on proposition logic (query term present/absent in document)
- Similarity of values not considered
- but multimedia IR deals with similarity already
Propositional vs. Predicate Logic

- Current IR systems are based on proposition logic (query term present/absent in document)
- Similarity of values not considered
- but multimedia IR deals with similarity already
- $\Rightarrow$ transition from propositional to predicate logic necessary
Propositional vs. Predicate Logic

- Current IR systems are based on proposition logic (query term present/absent in document)
- Similarity of values not considered
- but multimedia IR deals with similarity already
- $\Rightarrow$ transition from propositional to predicate logic necessary
- $\Rightarrow$ Probabilistic databases / Datalog are already based on predicate logic!
Vague Predicates in Probabilistic Datalog

[Fuhr & Roelleke 97] [Fuhr 00]

- Example: Shopping 45 inch LCD TV
- Vague predicates as builtin predicates: \( X \approx Y \)
- query(D):- Category(D, tv) & type(D, lcd) & size(D, X) & \( \approx (X, 45) \)

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>42</td>
<td>45</td>
</tr>
<tr>
<td>0.8</td>
<td>43</td>
<td>45</td>
</tr>
<tr>
<td>0.9</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>1.0</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>0.9</td>
<td>46</td>
<td>45</td>
</tr>
<tr>
<td>0.8</td>
<td>47</td>
<td>45</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Data types and vague predicates in IR

Data type: domain + (vague) predicates

- Language (multilingual documents) / (language-specific stemming)
- Person names / “his name sounds like Jones”
- Dates / “about a month ago”
- Amounts / “orders exceeding 1 Mio $”
- Technical measurements / “at room temperature”
- Chemical formulas
"I am looking for a 45-inch LCD TV with
- wide viewing angle
- high contrast
- low price
- high user rating"

→ vague criteria are very frequent in end-user querying of fact databases

→ but no appropriate support in SQL
"I am looking for a 45-inch LCD TV with
- wide viewing angle
- high contrast
- low price
- high user rating"

→ vague criteria are very frequent in end-user querying of fact databases

→ but no appropriate support in SQL

vague conditions → similar to fuzzy predicates
learn vague predicates from feedback data

construct feature vector \( \vec{x}(q_i, d_i) \) from query value \( q_i \) and document value \( d_i \) (e.g. relative difference)

apply logistic regression
Expressiveness

- Retrieval Rules, Joins, Aggregations and Restructuring
- Expressiveness in XML Retrieval
about(D,T) :- docTerm(D,T).

Consider document linking / anchor text
about(D,T) :- link(D1,D), about(D1,T).

Consider term hierarchy
about(D,T) :- subconcept(T,T1) & about(D,T1).

Field-specific term weighting
0.9 docTerm(D,T) :- occurs(D,T,title).
0.5 docTerm(D,T) :- occurs(D,T,body).
Expressiveness

Formulating Retrieval Rules

about(D,T) :- docTerm(D,T).

consider document linking / anchor text
about(D,T) :- link(D1,D), about(D1,T).
Expressiveness

Formulating Retrieval Rules

about(D,T) :- docTerm(D,T).

consider document linking / anchor text
about(D,T) :- link(D1,D),about(D1,T).

consider term hierarchy
about(D,T) :- subconcept(T,T1) & about(D,T1).
Expressiveness

Formulating Retrieval Rules

about(D,T) :- docTerm(D,T).

consider document linking / anchor text
about(D,T) :- link(D1,D),about(D1,T).

consider term hierarchy
about(D,T) :- subconcept(T,T1) & about(D,T1).

field-specific term weighting
0.9 docTerm(D,T) :- occurs(D,T,title).
0.5 docTerm(D,T) :- occurs(D,T,body).
IR authors:

\[
\text{irauthor}(N) :\equiv \text{about}(D, \text{ir}) \land \text{author}(D, N).
\]
IR authors:

irauthor(N):- about(D,ir) & author(D,N).

Smith’s IR papers cited by Miller

?- author(D,smith) & about(D,ir) &
   author(D1,miller) & cites(D,D1).
Who are the major IR authors?

<table>
<thead>
<tr>
<th>β</th>
<th>DNO</th>
<th>TERM</th>
<th>author</th>
<th>DNO</th>
<th>NAME</th>
<th>irauthor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1</td>
<td>ir</td>
<td></td>
<td>1</td>
<td>smith</td>
<td>0.98 smith</td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>db</td>
<td></td>
<td>2</td>
<td>miller</td>
<td>0.6 miller</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>ir</td>
<td></td>
<td>3</td>
<td>smith</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>ir</td>
<td></td>
<td>3</td>
<td>smith</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>3</td>
<td>ai</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

irauthor(A):- index(D,ir) & author(D,A).
Who are the major IR authors?

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>DNO</th>
<th>TERM</th>
<th>$\beta$</th>
<th>DNO</th>
<th>NAME</th>
<th>$\beta$</th>
<th>DNO</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1</td>
<td>ir</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>db</td>
<td>1</td>
<td>1</td>
<td>smith</td>
<td>0.98</td>
<td>smith</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>ir</td>
<td>2</td>
<td>2</td>
<td>miller</td>
<td>0.6</td>
<td>miller</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>ir</td>
<td>3</td>
<td>3</td>
<td>smith</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>3</td>
<td>ai</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\text{irauthor}(A) :- \ \text{index}(D,\text{ir}) \ \& \ \text{author}(D,A)$.

Aggregation through projection!
Who are the major IR authors?

<table>
<thead>
<tr>
<th>β</th>
<th>DNO</th>
<th>TERM</th>
<th>DNO</th>
<th>NAME</th>
<th>irauths</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1</td>
<td>ir</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>db</td>
<td></td>
<td>smith</td>
<td>1.7</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>ir</td>
<td>1</td>
<td>smith</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>ir</td>
<td>2</td>
<td>miller</td>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
<td>3</td>
<td>ai</td>
<td>3</td>
<td>smith</td>
<td></td>
</tr>
</tbody>
</table>

**Aggregation through summing:**

irauth(D,A):- index(D,ir) & author(D,A).

irauths SUM(Name) :- irdbauth(Doc,Name) | (Name)
Expressiveness in XML Retrieval

[Fuhr & Lalmas 07]

Content Typing

Object Types

Data Types

Text only

Structure

Nested structure

Named fields

XPath

XQuery
Expressiveness in XML Retrieval

[Fuhr & Lalmas 07]

Content Typing

Object Types

Data Types

Text only

Structure

Nested structure

Named fields

XPath

XQuery
XML structure: 1. Nested Structure

- XML document as hierarchical structure
- Retrieval of elements (subtrees)
- Typical query language does not allow for specification of structural constraints
- Relevance-oriented selection of answer elements: return the most specific relevant elements
XML structure: 2. Named Fields

- Reference to elements through field names only
- Context of elements is ignored (e.g. author of article vs. author of referenced paper)
- Post-Coordination may lead to false hits (e.g. author name – author affiliation)

Example: Dublin Core

```xml
  <dc:title>Generic Algebras</dc:title>
  <dc:creator>A. Smith (ESI), B. Miller (CMU)</dc:creator>
  <dc:subject>Orthogonal group, Symplectic group</dc:subject>
  <dc:date>2001-02-27</dc:date>
  <dc:format>application/postscript</dc:format>
  <dc:source>ESI preprints</dc:source>
  <dc:language>en</dc:language>
</oai_dc:dc>
```
/document/chapter[about(.//heading, XML) AND about(.//section/*, syntax)]
XML structure: 3. XPath

/document/chapter[about(.)/heading, XML) AND about(.)/section/*,syntax)]]
XML structure: 3. XPath (cont’d)

- Full expressiveness for navigation through document tree (+links)
  - Parent/child, ancestor/descendant
  - Following/preceding, following-sibling, preceding-sibling
  - Attribute, namespace

- Selection of arbitrary elements/subtrees
  (but answer can be only a single element of the originating document)
Higher expressiveness, especially for database-like applications:

- Joins (trees → graphs)
- Aggregations
- Constructors for restructuring results
Higher expressiveness, especially for database-like applications:

- Joins (trees → graphs)
- Aggregations
- Constructors for restructuring results

Example: List each publisher and the average price of its books

FOR $p$ IN distinct(document("bib.xml")//publisher)
LET $a :=$ avg(document("bib.xml")//book[publisher = $p]/price)
RETURN
  <publisher>
  <name> $p/text() </name>
  <avgprice> $a </avgprice>
  </publisher>
XML content typing

- Text only
- Data Types
- Object Types

Content Typing

Structure

XQuery

XPath

Named fields

Nested structure
XML content typing: 1. Text

<book>
<author>John Smith</author>
<title>XML Retrieval</title>
<chapter>
    <heading>Introduction</heading>
    This text explains all about XML and IR.
</chapter>
<chapter>
    <heading>XML Query Language XQL</heading>
    <section>
        <heading>Examples</heading>
    </section>
    <section>
        <heading>Syntax</heading>
        Now we describe the XQL syntax.
    </section>
</chapter>
</book>
XML content typing: 2. Data Types

- Data type: domain + (vague) predicates (see above)
- Close relationship to XML Schema, but
  - XMLS supports syntactic type checking only
  - No support for vague predicates
Object types: Persons, Locations, Dates, ...  

Pablo Picasso (October 25, 1881 - April 8, 1973) was a Spanish painter and sculptor. In Paris, Picasso entertained a distinguished coterie of friends in the Montmartre and Montparnasse quarters, including André Breton, Guillaume Apollinaire, and writer Gertrude Stein.

To which other artists did Picasso have close relationships? Did he ever visit the USA?

Named entity recognition methods allow for automatic markup of object types.

Object types support increased precision.
XML content typing
Tag semantics modelled as hierarchies

Object type hierarchies

Person

Scientist  Artist

Physicist  Chemist  Poet  Actor  Singer

Role hierarchies

Creator

Author  Editor
XML content typing
Tag semantics modelled in OWL
Further Concepts
Further Concepts

4-valued (probabilistic) logics

Supported concepts

- conflicting knowledge
- open + closed world assumptions
Further Concepts
4-valued (probabilistic) logics

Supported concepts
- conflicting knowledge
- open + closed world assumptions

Applications
- 4-valued probabilistic Datalog [Fuhr & Roelleke 98]
- POOL: Probabilistic Object-Oriented Logic [Lalmas et al. 02]
- POLAR: Retrieval with Annotations [Frommholz & Fuhr 06]
- POLIS: Information summarization [Forst et al. 07]
IR Systems vs. DBMS

Application

DBMS

DB

IRS

Collection
IR Systems vs. DBMS
IR Systems vs. DBMS

Separation between IRS and IR application?
Towards an IRMS
Towards an IRMS

Application

SQL

DBMS

DB

Application

IRMS

Collection
Towards an IRMS

Application \(\rightarrow\) SQL \(\rightarrow\) DBMS

Application \(\rightarrow\) IR Query Language \(\rightarrow\) IRMS

DB

Collection
Conclusion
Conclusion

Inference

- Probabilistic relational model supports integration of IR+DB
- Probabilistic Datalog as powerful inference mechanism
- Allows for formulating retrieval strategies as logical rules
Conclusion

Inference

- Probabilistic relational model supports integration of IR+DB
- Probabilistic Datalog as powerful inference mechanism
- Allows for formulating retrieval strategies as logical rules

Vague predicates

- Natural extension of IR methods to attribute values
- Vague predicates can be learned from feedback data
- Transition from propositional to predicate logic
Conclusion

Inference
- Probabilistic relational model supports integration of IR+DB
- Probabilistic Datalog as powerful inference mechanism
- Allows for formulating retrieval strategies as logical rules

Vague predicates
- Natural extension of IR methods to attribute values
- Vague predicates can be learned from feedback data
- Transition from propositional to predicate logic

Expressive query language
- Joins
- Aggregations
- (Re)structuring of results
Don’t program search engines, design them

http://www.spinque.com/
Dalvi, N. N.; Suciu, D.
(2007).
Efficient query evaluation on probabilistic databases.

Forst, J. F.; Tombros, A.; Roelleke, T.
(2007).
POLIS: A Probabilistic Logic for Document Summarisation.

Frommholz, I.; Fuhr, N.
Probabilistic, Object-oriented Logics for Annotation-based Retrieval in Digital Libraries.
References II

Fuhr, N.; Lalmas, M.
(2007).
Advances in XML retrieval: the INEX initiative.

Fuhr, N.; Rölleke, T.
(1997).
A Probabilistic Relational Algebra for the Integration of Information Retrieval and Database Systems.

Fuhr, N.; Rölleke, T.
HySpirit – a Probabilistic Inference Engine for Hypermedia Retrieval in Large Databases.


References

Pearl, J.
(1988).
Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.
Morgan Kaufman, San Mateo, California.

Rölleke, T.; Wu, H.; Wang, J.; Azzam, H.
(2007).
Modelling retrieval models in a probabilistic relational algebra with a new operator: the relational Bayes.
The International Journal on Very Large Data Bases (VLDB) 17(1), pages 5–37.

Suciu, D.; Olteanu, D.; Ré, C.; Koch, C.
(2011).
Probabilistic Databases.