Believe It or Not –

Adding belief annotations
to databases

Wolfgang Gatterbauer, Magda Balazinska, Nodira Khoussainova, and Dan Suciu

University of Washington
http://db.cs.washington.edu/beliefDB/
High-level overview

• DBMS: manage consistent data
• Applications need inconsistent DM
  – Scientific databases
  – Internet community databases
• Community DBMS: manage inconsistent views

• This work: Belief databases
  – manage data and curation
  – grounded in modal and default logic
  – implemented on top of relational model

reason: disagreement!
Agenda

- Motivating example
- Logic foundations
- Relational implementation
- Discussion
Motivating application

  - volunteer contribute animal observations
  - one person curates the database

### Observations

<table>
<thead>
<tr>
<th>id</th>
<th>uid</th>
<th>species</th>
<th>date</th>
<th>location</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Alice</td>
<td>Crow</td>
<td>06-14-08</td>
<td>Lake Placid</td>
<td>found feathers</td>
</tr>
</tbody>
</table>

**problem: does not scale!**

### Sightings (S)

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>species</th>
<th>date</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Crow</td>
<td>06-14-08</td>
<td>Lake Placid</td>
</tr>
</tbody>
</table>

### Comments (C)

<table>
<thead>
<tr>
<th>cid</th>
<th>comment</th>
<th>sid</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>found feathers</td>
<td>s2</td>
</tr>
</tbody>
</table>
1. Distinct database instances

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>species</th>
<th>date</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Crow</td>
<td>06-14-08</td>
<td>Lake Placid</td>
</tr>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Raven</td>
<td>06-14-08</td>
<td>Lake Placid</td>
</tr>
</tbody>
</table>

D1: Belief worlds: individually consistent, mutually possibly inconsistent
1. Distinct database instances

BeliefSQL

\[ I: Alice \text{ believes that she saw a crow.} \]
insert into `BELIEF` 'Alice' Sightings values ('s2','Alice','Crow','06-14-08','Lake Placid')

\[ I: Bob \text{ believes that she actually saw a raven.} \]
insert into `BELIEF` 'Bob' Sightings values ('s2','Alice','Raven','06-14-08','Lake Placid')

Q: Who believes something different than Alice and what?
select U2.name, S1.species, S2.species from Users as U,
`BELIEF` 'Alice' Sightings as S1,
`BELIEF` U.uid Sightings as S2,
where S1.sid = S2.sid
and S1.species <> S2.species
A: \{('Bob', 'Crow', 'Raven')\}
2. Open world assumption

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>species</th>
<th>date</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Crow</td>
<td>06-14-08</td>
<td>Lake Placid</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>species</th>
<th>date</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Raven</td>
<td>06-14-08</td>
<td>Lake Placid</td>
</tr>
</tbody>
</table>

Adapted key constraints!

D2: Model incomplete knowledge with explicit negative beliefs
2. Open world assumption

I: Carol does not believe that Alice saw a crow nor a raven.

insert into BELIEF ‘Carol’ not Sightings values (‘s2’,’Alice’,’Crow’,’06-14-08’,’Lake Placid’)

insert into BELIEF ‘Carol’ not Sightings values (‘s2’,’Alice’,’Raven’,’06-14-08’,’Lake Placid’)
2. Open world assumption

Q: Who disagrees with a sighting from '06-14-08' that Alice believes?

A: \{('Bob', 'Crow'), ('Carol', 'Crow')\}

```sql
select U.name, S1.species
from Users as U,
    BELIEF 'Alice' Sightings as S1,
    BELIEF U.uid not Sightings as S2
where S1.sid = S2.sid
and S1.uid = S2.uid
and S1.species = S2.species
and S1.date = '06-14-08'
and S2.date = '06-14-08'
and S1.location = S2.location
```
3. Higher-order beliefs

D3: Beliefs about other user’s beliefs: allow discussion between users
3. Higher-order beliefs

I: According to Bob, Alice believes that the feathers of the sighted animal were plain black.

insert into BELIEF ‘Bob’ BELIEF ‘Alice’ Comments values (‘c1’, ‘plain black feathers’, ‘s2’)
3. Higher-order beliefs

**Q:** Which comments does Alice believe according to Bob, which he does not believe himself?

**A:** \{'c1', 'plain-black feathers'\}

```sql
select C1.cid, C1.comment
from BELIEF 'Bob' BELIEF 'Alice' Comments as C1,
    BELIEF 'Bob' not Comments as C2
where C1.cid = C2.cid
    and C1.comment = C2.comment
    and C1.sid = C2.sid
```

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>species</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Raven</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>species</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Crow</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cid</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>plain black feathers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cid</th>
<th>comment</th>
<th>sid</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>purple-black feathers</td>
<td>s2</td>
</tr>
</tbody>
</table>
3. Higher-order beliefs

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>species</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>Raven</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sid</th>
<th>uid</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>s2</td>
<td>Alice</td>
<td>plain black feathers</td>
</tr>
</tbody>
</table>

**Q:** Which comments does Alice believe according to somebody, which (s)he does not believe themself?

```sql
select U.name, C1.sid, C1.comment
from Users as U,
BELIEF U.uid BELIEF ‘Alice’ Comments as C1,
BELIEF U.uid not Comments as C2
where C1.cid = C2.cid
and C1.comment = C2.comment
and C1.sid = C2.sid
```

**A:** {{‘Bob’, ‘c1’, ‘plain-black feathers’}}
4. Message board assumption

D4: Default assumption: models a trusting attitude & avoids repeated inserts
4. Message board assumption

Q: Which animal sightings does Alice believe according to Bob, which he does not believe himself?

select S1.sid, S1.species
from BELIEF 'Bob' BELIEF 'Alice' Sightings as S1, BELIEF 'Bob' not Sightings as S2
where S1.sid = S2.sid
and S1.uid = S2.uid
and S1.species = S2.species
and S1.date = S2.date
and S1.location = S2.location

A: {('s2', 'Crow')}
What we have seen so far

- 4 Design decisions for Belief Database model
  - Distinct belief worlds
  - Open world assumption (OWA)
  - Higher-order beliefs
  - Message board assumption

- **BeliefSQL**
  - SQL + ‘BELIEF’ (+ ‘not’)
Agenda

• Motivating example

• Logic foundations

• Relational implementation

• Discussion
Logic foundations: Belief statements

Belief database $D = \{\varphi_1, \ldots, \varphi_n\}$

- Insert into $\text{BELIEF}$ ‘Alice’ S values (s2, ‘Alice’, ‘Crow’,…)

$$\varphi = \Box_w t^s$$

“Annotation of ground tuple”
Logic foundations: Entailment

One belief annotation:

\( D = \{ \varphi_1 \} \)

\( \varphi_1 = \square_{\text{Alice}} S^+ (\ldots 'Crow', \ldots) \)

More than one entailed belief:

\( D \models \square_{\text{Bob} \cdot \text{Alice}} S^+ (\ldots 'Crow', \ldots) \)
Logic foundations: Message board assumption

Message board assumption

If $D \models \Box_w t^s$

and $\Box_{u \cdot w} t^s$ consistent with $D$

then $D \models \Box_{u \cdot w} t^s$

Default logic

$\varphi : \Box_u \varphi$

$\Box_u \varphi$

non-monotonic reasoning!

$D$

Explicit beliefs (annotations)

$\tilde{D} \setminus D$

Implicit beliefs (assumptions)

$\tilde{D}$

Entailed beliefs (extension)

belief database “contains” more than the explicit belief annotations!
“Semi-formal” problem statement

INPUT
Belief statements
\[ i_1: \varphi_1 \]
\[ i_2: \varphi_2 \]
\[ ... \]
\[ i_n: \varphi_n \]

Adapted key constraints

Message board assumption
\[ \varphi: \Box u \varphi \]
\[ \Box u \varphi \]

OUTPUT

Belief Conjunctive Queries (BCQ)
\[ q(\bar{x}) := \Box w_1 R_1^{s_1}(\bar{x}_1), ..., \Box w_g R_g^{s_g}(\bar{x}_g) \]

Belief statements
\[ \varphi \]
\[ D \models \varphi \]

Belief statements
\[ \Box w_1...w_d R^+(x_1,...x_l) \]

Belief statements
\[ q(\bar{x}) := \Box w R_i^+(\bar{x}_i) \]
Agenda

• Motivating example
• Logic foundations
• Relational implementation
• Discussion
Canonical Kripke structure

**Belief statements***

- $i_1: s_1^+$
- $i_2: □_{Bob} s_1^-$
- $i_3: □_{Bob} s_2^-$
- $i_4: □_{Alice} s_2^+$
- $i_5: □_{Alice} c_1^+$
- $i_6: □_{Bob} s_2^+$
- $i_7: □_{Bob\cdot Alice} c_2^+$
- $i_8: □_{Bob} c_2^+$

**Message board assumption**

- $\phi: □_i \phi$
- $□_i \phi$

* Running example from the paper
### Relational representation

#### Sightings_INTERNAL

<table>
<thead>
<tr>
<th>tid</th>
<th>sid</th>
<th>uid</th>
<th>species</th>
<th>date</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1.1</td>
<td>s1</td>
<td></td>
<td>Carol Bald eagle</td>
<td>06-14-08</td>
<td>Lake Forest</td>
</tr>
<tr>
<td>s1.2</td>
<td>s1</td>
<td></td>
<td>Carol Fish eagle</td>
<td>06-14-08</td>
<td>Lake Forest</td>
</tr>
<tr>
<td>s2.1</td>
<td>s2</td>
<td></td>
<td>Alice Crow</td>
<td>06-14-08</td>
<td>Lake Placid</td>
</tr>
<tr>
<td>s2.2</td>
<td>s2</td>
<td></td>
<td>Alice Raven</td>
<td>06-14-08</td>
<td>Lake Placid</td>
</tr>
</tbody>
</table>

#### Comments_INTERNAL

<table>
<thead>
<tr>
<th>tid</th>
<th>cid</th>
<th>comment</th>
<th>sid</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1.1</td>
<td>c1</td>
<td>found feathers</td>
<td>s2</td>
</tr>
<tr>
<td>c2.1</td>
<td>c2</td>
<td>plain black feathers</td>
<td>s2</td>
</tr>
<tr>
<td>c2.2</td>
<td>c2</td>
<td>purple-black feathers</td>
<td>s2</td>
</tr>
</tbody>
</table>

#### Sightings_V

<table>
<thead>
<tr>
<th>wid</th>
<th>tid</th>
<th>sid</th>
<th>s</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>#0</td>
<td>s1.1</td>
<td>s1</td>
<td>+</td>
<td>y</td>
</tr>
<tr>
<td>#1</td>
<td>s1.1</td>
<td>s1</td>
<td>+</td>
<td>n</td>
</tr>
<tr>
<td>#2</td>
<td>s1.1</td>
<td>s1</td>
<td>-</td>
<td>y</td>
</tr>
<tr>
<td>#2</td>
<td>s1.2</td>
<td>s1</td>
<td>-</td>
<td>y</td>
</tr>
<tr>
<td>#2</td>
<td>s2.2</td>
<td>s2</td>
<td>+</td>
<td>y</td>
</tr>
<tr>
<td>#3</td>
<td>s1.1</td>
<td>s1</td>
<td>+</td>
<td>n</td>
</tr>
<tr>
<td>#3</td>
<td>s2.1</td>
<td>s2</td>
<td>+</td>
<td>n</td>
</tr>
</tbody>
</table>

#### E

<table>
<thead>
<tr>
<th>wid1</th>
<th>uid</th>
<th>wid2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#0</td>
<td>Alice</td>
<td>#1</td>
</tr>
<tr>
<td>#0</td>
<td>Bob</td>
<td>#2</td>
</tr>
<tr>
<td>#0</td>
<td>Carol</td>
<td>#0</td>
</tr>
<tr>
<td>#1</td>
<td>Bob</td>
<td>#2</td>
</tr>
<tr>
<td>#1</td>
<td>Carol</td>
<td>#0</td>
</tr>
<tr>
<td>#2</td>
<td>Alice</td>
<td>#3</td>
</tr>
<tr>
<td>#2</td>
<td>Carol</td>
<td>#0</td>
</tr>
<tr>
<td>#3</td>
<td>Bob</td>
<td>#2</td>
</tr>
<tr>
<td>#3</td>
<td>Carol</td>
<td>#0</td>
</tr>
</tbody>
</table>

#### Comments_V

<table>
<thead>
<tr>
<th>wd</th>
<th>tid</th>
<th>cid</th>
<th>s</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>c1.1</td>
<td>c1</td>
<td>+</td>
<td>y</td>
</tr>
<tr>
<td>#2</td>
<td>c2.2</td>
<td>c2</td>
<td>+</td>
<td>y</td>
</tr>
<tr>
<td>#3</td>
<td>c1.1</td>
<td>c1</td>
<td>+</td>
<td>n</td>
</tr>
<tr>
<td>#3</td>
<td>c2.1</td>
<td>c2</td>
<td>+</td>
<td>y</td>
</tr>
</tbody>
</table>

#### D

<table>
<thead>
<tr>
<th>wid</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>#0</td>
<td>0</td>
</tr>
<tr>
<td>#1</td>
<td>1</td>
</tr>
<tr>
<td>#2</td>
<td>1</td>
</tr>
<tr>
<td>#3</td>
<td>2</td>
</tr>
</tbody>
</table>

#### S

<table>
<thead>
<tr>
<th>wid1</th>
<th>wid2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>#0</td>
</tr>
<tr>
<td>#2</td>
<td>#0</td>
</tr>
<tr>
<td>#3</td>
<td>#1</td>
</tr>
</tbody>
</table>
Example Translation of a Belief CQ (BCQ)

Q: Who disagrees with a sighting from ’06-14-08’ that Alice believes?

BeliefSQL

<table>
<thead>
<tr>
<th>select</th>
<th>U.name, S1.species</th>
</tr>
</thead>
<tbody>
<tr>
<td>from</td>
<td>Users as U,</td>
</tr>
<tr>
<td></td>
<td>BELIEF ‘Alice’ Sighting as S1,</td>
</tr>
<tr>
<td></td>
<td>BELIEF U.uid not Sighting as S2</td>
</tr>
<tr>
<td>where</td>
<td>S1.sid = S2.sid</td>
</tr>
<tr>
<td>and</td>
<td>S1.uid = S2.uid</td>
</tr>
<tr>
<td>and</td>
<td>S1.species = S2.species</td>
</tr>
<tr>
<td>and</td>
<td>S1.date = ’06-14-08’</td>
</tr>
<tr>
<td>and</td>
<td>S2.date = ’06-14-08’</td>
</tr>
<tr>
<td>and</td>
<td>S1.location = S2.location</td>
</tr>
</tbody>
</table>

Translation into SQL

```sql
select E1.uid as uid1, V.tid, V.sid, R.uid, R.species, R.date, R.location, V.s into T2
from E as E1, Sightings_V as V, Sightings_STAR as R
where E1.wid1=0
and V.wid=E1.wid2
and V.tid=R.tid
and E1.uid='1';

select E1.uid as uid1, V.tid, V.sid, R.uid, R.species, R.date, R.location, V.s into T1
from E as E1, Sightings_V as V, Sightings_STAR as R
where E1.wid1=0
and V.wid=E1.wid2
and V.tid=R.tid;

select T1.uid1, T2.species
from T1 as T1, T2 as T2
where T1.sid=T2.sid
and (T1.s=0 and T1.uid=T2.uid and T1.species=T2.species and T1.date='06-14-08' and T1.location=T2.location) or
(T1.s=1 and (T1.uid<>T2.uid or T1.species<>T2.species or T1.date<>'06-14-08' or T1.location<>T2.location))
and T2.s=1
and T2.date='06-14-08';

drop table T2;
drop table T1;
```

\[ q(x,y) \vdash \\square_{\text{Alice}} S^+(u,v,y,'06-14-08',z), \\square_x S^-(u,v,y,'06-14-08',z) \]
Agenda

• Motivating example
• Logic foundations
• Relational implementation
• Discussion
Experiments

Size

Relative overhead \( \rho := \frac{|R^*|}{n} \)

\[ \rho = O(m^{d_{\text{max}}}) \]

In theory: e.g. 100 users, max. depth 2

\( \rho \rightarrow 10,000 \)

Experiments: \( \rho \rightarrow 21 - 1,009 \)

Size not limitation of semantics, but of relational implementation!

Time

Depends on type of query (3 types in paper)

Q1: \(~0.1\) s
Q2: \(~0.4\) s
Q3: \(~4.5\) s

Considerable speed-up to come!
Inspirations and related work (excerpt)

• Annotations
  – Buneman et al. [ICDT 2001 / ICDT 2007]
  – Bhagwat et al. [VLDBJ 2005], Geerts et al. [ICDE 2006]
  – Srivastava & Velegrakis [SIGMOD 2007]

• Modal logic
  – Fagin et al. [1995]
  – Calvanese et al. [IS 2008]
  – Nguyen [LJ-IGPL 2008]

• Uncertain / incomplete information
  – Sarma et al. [ICDE 2006]
  – Green & Tannen [IEEE Data Eng. 2006]
  – Dalvi & Suciu [PODS 2007]

• Inconsistency / key violations
  – Arenas et al. [PODS 1999]
  – Fuxman et al. [SIGMOD 2005]

• Peer-to-peer computing / collaborative data sharing
  – Bernstein et al. [WebDB 2002]
  – Ives et al. [SIGMOD record 2008]
Conclusions

- Proposed BELIEF databases
  - Goal: manage, curate inconsistent data

- Implementation
  - Logical foundations
  - Relational translation

- Current work
  - making it compact and fast
BACKUP
Relative overhead of relational representation

Bound for relative overhead $\frac{|R^*|}{n} = O(m^{d_{\max}})$

Measured relative overhead $\frac{|R^*|}{n}$ for $n = 10,000$ annotations, $m = 100$ users, uniform or Zipf user participation, and 3 distributions of annotation depth:

<table>
<thead>
<tr>
<th>Pr$[d = {0, 1, 2}]$</th>
<th>uniform</th>
<th>Zipf</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0.3, 0.3, 0.3]</td>
<td>1,009</td>
<td>130</td>
</tr>
<tr>
<td>[0.8, 0.19, 0.01]</td>
<td>162</td>
<td>68</td>
</tr>
<tr>
<td>[0.199, 0.8, 0.001]</td>
<td>26</td>
<td>21</td>
</tr>
</tbody>
</table>

Measured relative overhead $\frac{|R^*|}{n}$ for $m = 100$ users, uniform user participation, and 2 distributions of annotation depth:
Query types and execution times

1. **Query for content**: “What does Alice believe?” $d \in \{0, \ldots, 4\}$:
   
   $$ q_{1,d}(x, y) : = \Box_{w}S^{+}(x, - , y, - , - ) , \text{ with } \|w\| \in \{0, \ldots, 4\} $$

2. **Query for conflicts**: “Which animal sightings does Bob believe that Alice believes, which he does not believe himself?”
   
   $$ q_{2}(x, y) : = \Box_{2.1}S^{+}(x, z, y, u, v), \Box_{2}S^{-}(x, z, y, u, v) $$

3. **Query for users**: Who disagrees with any of Alice’s beliefs of sightings at Lake Placid?”
   
   $$ q_{3}(x):=\Box_{x}S^{-}(y, z, u, v, 'a'), \Box_{1}S^{+}(y, z, u, v, 'a') $$

Execution times and size of result sets for example queries executed over a belief database with 10,000 annotations and relative overhead 22.4.

<table>
<thead>
<tr>
<th>E(Time) [msec]</th>
<th>$q_{1,0}$</th>
<th>$q_{1,1}$</th>
<th>$q_{1,2}$</th>
<th>$q_{1,3}$</th>
<th>$q_{1,4}$</th>
<th>$q_{2}$</th>
<th>$q_{3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$(Time) [msec]</td>
<td>105</td>
<td>145</td>
<td>146</td>
<td>152</td>
<td>144</td>
<td>436</td>
<td>4473</td>
</tr>
<tr>
<td>Result size</td>
<td>1626</td>
<td>2816</td>
<td>2253</td>
<td>2061</td>
<td>1931</td>
<td>196</td>
<td>99</td>
</tr>
</tbody>
</table>
Belief Conjunctive Queries (BCQ)

Conjunctive Queries (CQ) in Datalog form:

\[ q(\bar{x}) : - R_1(\bar{x}_1), \ldots, R_g(\bar{x}_g) \]

Belief Conjunctive Queries (BCQ) in ”Modal Datalog” form:

\[ q(\bar{x}) : - \square_{\bar{w}_1} R_{1}^{s_1}(\bar{x}_1), \ldots, \square_{\bar{w}_g} R_{g}^{s_g}(\bar{x}_g) \]

\[ q_1 : \text{”Who disagrees with any sighting from ’06-14-08’ that Alice believes?”} \]

\[ q_1(x, y) : - \square_{\text{Alice}} S^+(u, v, y, ’06-14-08’, z), \square_x S^-(u, v, y, ’06-14-08’, z) \]

\[ q_1(D) =\{’Bob’, ’bald eagle’\},\{’Bob’, ’crow’\} \]
Revisiting the semantics / the user

Standard relational model

(1) SQL

Conflicts in belief worlds:
OWA, keys, ML, DA

(2) BeliefSQL

BELIEF 'Alice' (..., 'eagle', ...)

BELIEF 'Bob' BELIEF 'Alice'
(..., 'black feathers', ...)

BELIEF 'Alice' ASSERTS (..., 'eagle', ...)

(3) ?

-> Structured discourse

-> 'Bob' SUGGESTS that the ASSUMPTION (...,'black feathers',...) has led 'Alice' to her original observation