Modeling and Querying Evidential Databases: First Steps towards a Strong Representation System

Fatma Ezzahra Bousnina
Mouna Chebbah Mohamed Anis Bach Tobji Allel HadjAli
Boutheina Ben Yaghlane

May 18^{th}, 2017

Outline

Outline

Imperfect Information

Imperfection

- Uncertainty
- Imprecision

Imperfect Information

Imperfection

- Uncertainty
- Imprecision

Example

- John has at least two children and I am not sure about it
  $\Rightarrow$ Imprecise and Uncertain

Imperfect Information

Imperfection

- Uncertainty
- Imprecision

Example

- John has at least two children and I am not sure about it
  $\Rightarrow$ Imprecise and Uncertain
- John has at least two children and I am sure about it
  $\Rightarrow$ Imprecise and Certain

Imperfect Information

Imperfection

- Uncertainty
- Imprecision

Example

- John has at least two children and I am not sure about it
  $\Rightarrow$ Imprecise and Uncertain
- John has at least two children and I am sure about it
  $\Rightarrow$ Imprecise and Certain
- John has three children and I am not sure about it
  $\Rightarrow$ Precise and Uncertain

Imperfect Information

Imperfection

- Uncertainty
- Imprecision

Example

- John has at least two children and I am not sure about it
  $\Rightarrow$ Imprecise and Uncertain
- John has at least two children and I am sure about it
  $\Rightarrow$ Imprecise and Certain
- John has three children and I am not sure about it
  $\Rightarrow$ Precise and Uncertain
- John has three children and I am sure about it
  $\Rightarrow$ Precise and Certain

Theories and database models of Imperfection

Theories and database models of Imperfection



- Probability Theory
  [Laplace, 1812]

Theories and database models of Imperfection

- Probability Theory [Laplace, 1812]
- Possibility Theory [Zadeh, 1978]
- Fuzzy sets Theory [Zadeh,1965]
- Evidence Theory [Dempster, 1967; Shafer, 1976]

Probabilistic Databases

ID	*Weather*	*Probability*
1	rainy	0.6
2	sunny	0.4

(R.Cavallo, M.Pittarelli,1987)

Possibilistic Databases

ID	*Weather*	*Possibility*
1	rainy	1
2	sunny	0.4

[H.Prade, C.testemale, 1984]

Evidential Databases

ID	*Weather*	*Confidence Level*
1	Rainy 0.5 Sunny 0.5	[0.3 1]
2	{Sunny ; Rainy} 1	[0.5 0.5]

[S.K.Lee, 1992]

Theory of Evidence

- Theory of evidence is also called theory of belief functions or Dempster-Shafer theory.

Theory of Evidence

- Theory of evidence is also called theory of belief functions or Dempster-Shafer theory.
- Introduced by Dempster in 1967 and formalized by Shafer in 1976.

Theory of Evidence

- Theory of evidence is also called theory of belief functions or Dempster-Shafer theory.
- Introduced by Dempster in 1967 and formalized by Shafer in 1976.
- Popularized by Smets in 1988 with the Transferable Belief Model (TBM).

Theory of Evidence

- Theory of evidence is also called theory of belief functions or Dempster-Shafer theory.
- Introduced by Dempster in 1967 and formalized by Shafer in 1976.
- Popularized by Smets in 1988 with the Transferable Belief Model (TBM).
- Theory of evidence provides an explicit representation of *uncertainty*, *imprecision* and *inconsistency*.

Evidential Databases

Evidential Databases

An *Evidential Database* (EDB) on its *compact form* has N tuples and D attributes. The value of an attribute c for an object l is called *evidential value*, V_{lc}.

ID	Disease	Symptom	CL
1	Diabetes 1	Fatigue 0.4 {Fatigue, Nausea} 0.6	[0.3 ; 0.9]

Evidential Databases

Mass Function

A mass function, m, is a mapping from 2^{Θ} to [0;1]. The basic belief mass of an hypothesis x, $m(x)$ represents the degree of truth about that hypothesis x such that:

$$\sum_{x \subseteq \Theta} m^{\Theta}(x) = 1$$

ID	Disease	Symptom	CL
1	Diabetes 1	Fatigue 0.4 {Fatigue, Nausea} 0.6	[0.3 ; 0.9]

Evidential Databases

Confidence Level

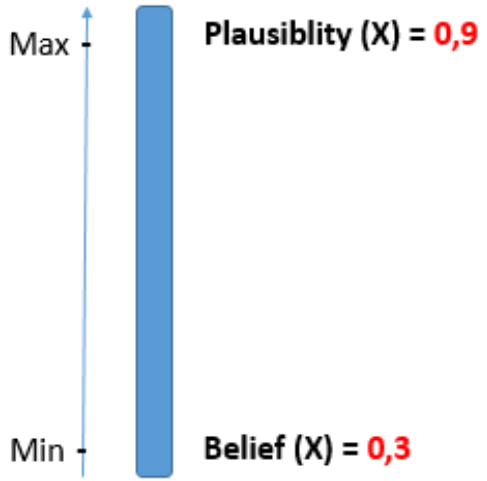
A confidence level CL is a measure that quantifies the degree of belief bel and plausibility pl about the existence of each tuple in the database, $CL = \{[bel, pl] | bel, pl \in [0, 1]; bel \leq pl\}$.

ID	Disease	Symptom	CL
1	Diabetes 1	Fatigue 0.4 {Fatigue, Nausea} 0.6	[0.3 ; 0.9]

Belief Function

The *belief function*, *bel*, is the minimal degree of belief given to an hypothesis x.

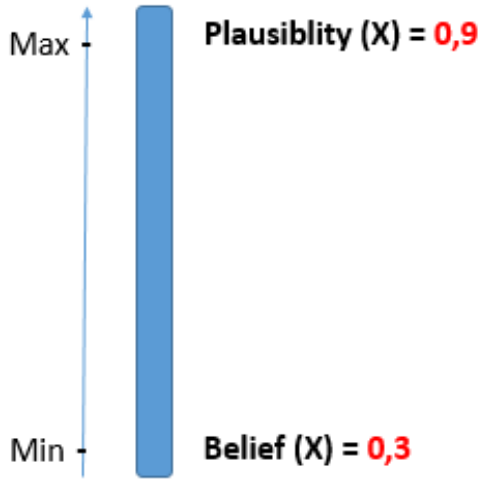
$$bel(x) = \sum_{y,x \subseteq \Theta : y \subseteq x} m(y)$$



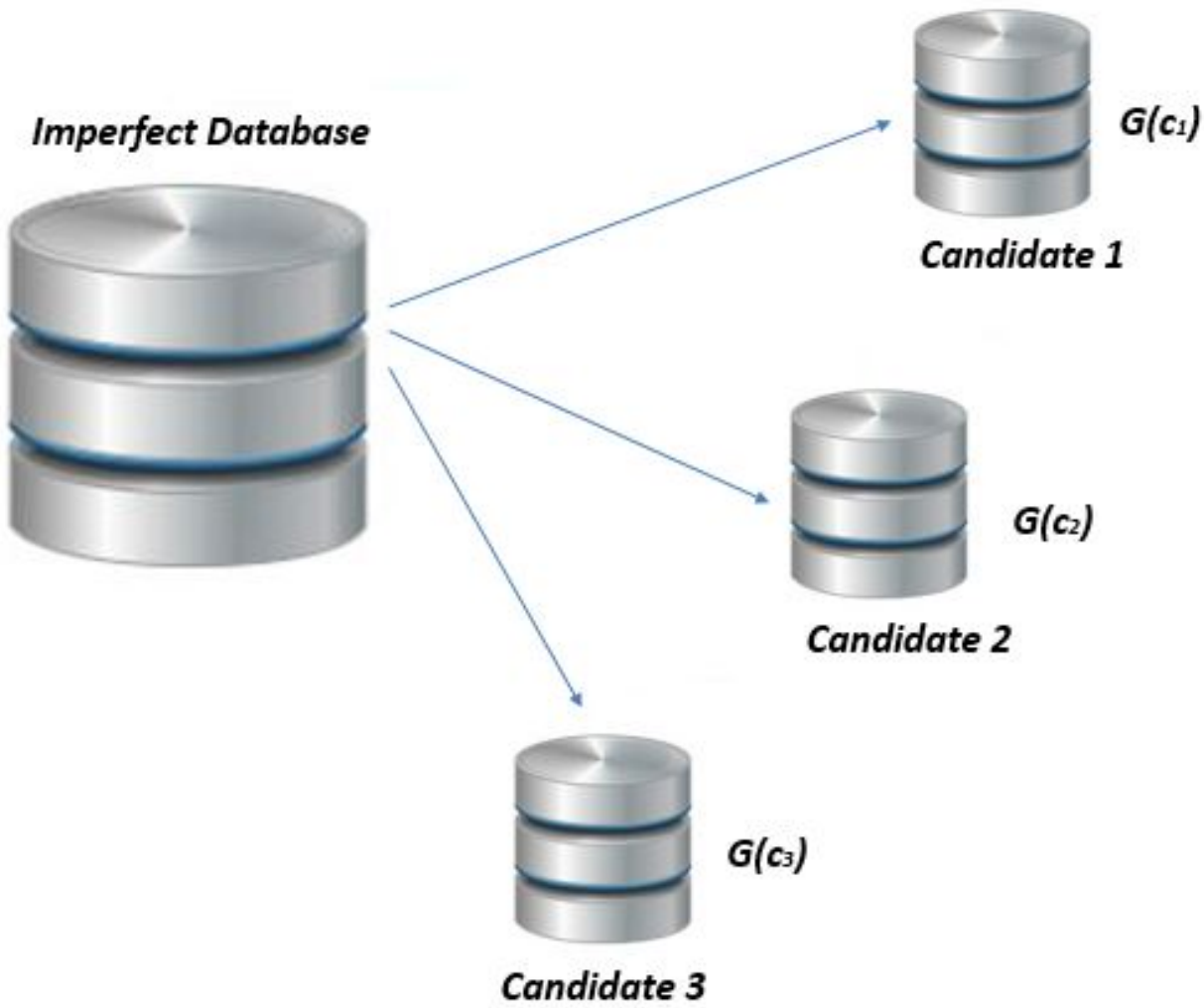
Plausibility Function

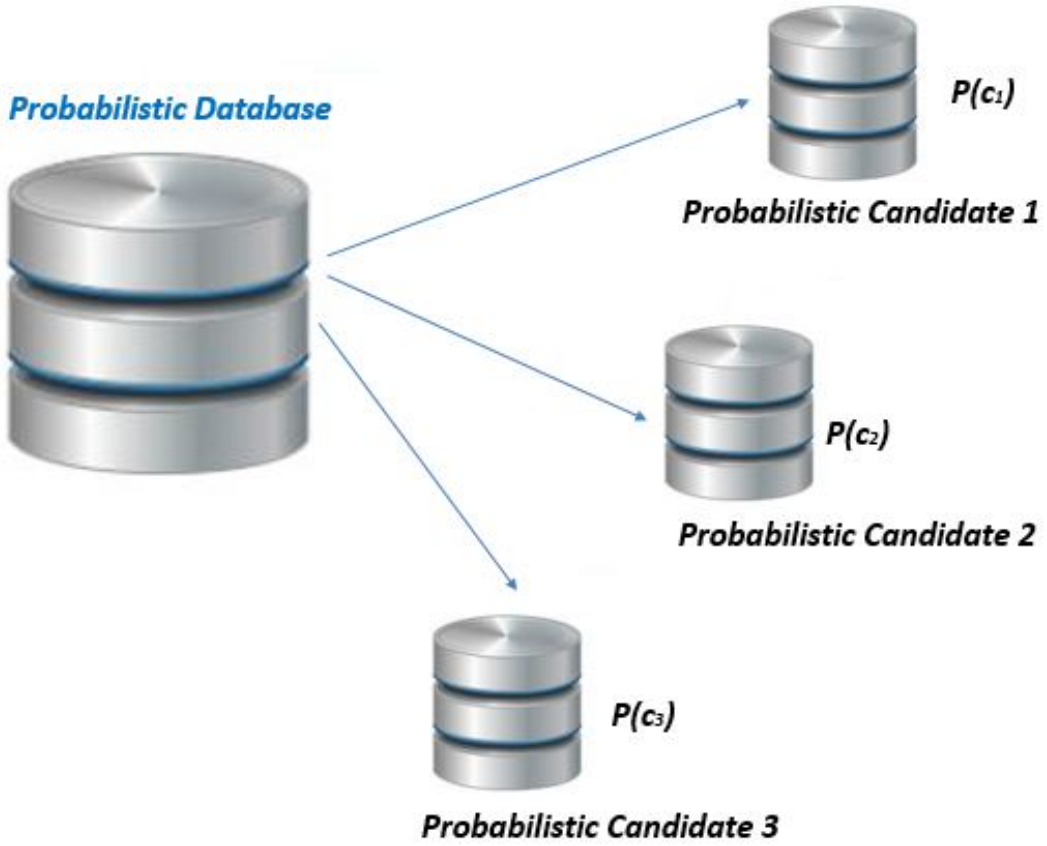
The *plausibility function*, *pl*, is the maximal amount of belief on the hypothesis x.

$$pl(x) = \sum_{y, x \subseteq \Theta : x \cap y \neq \varnothing} m(y)$$

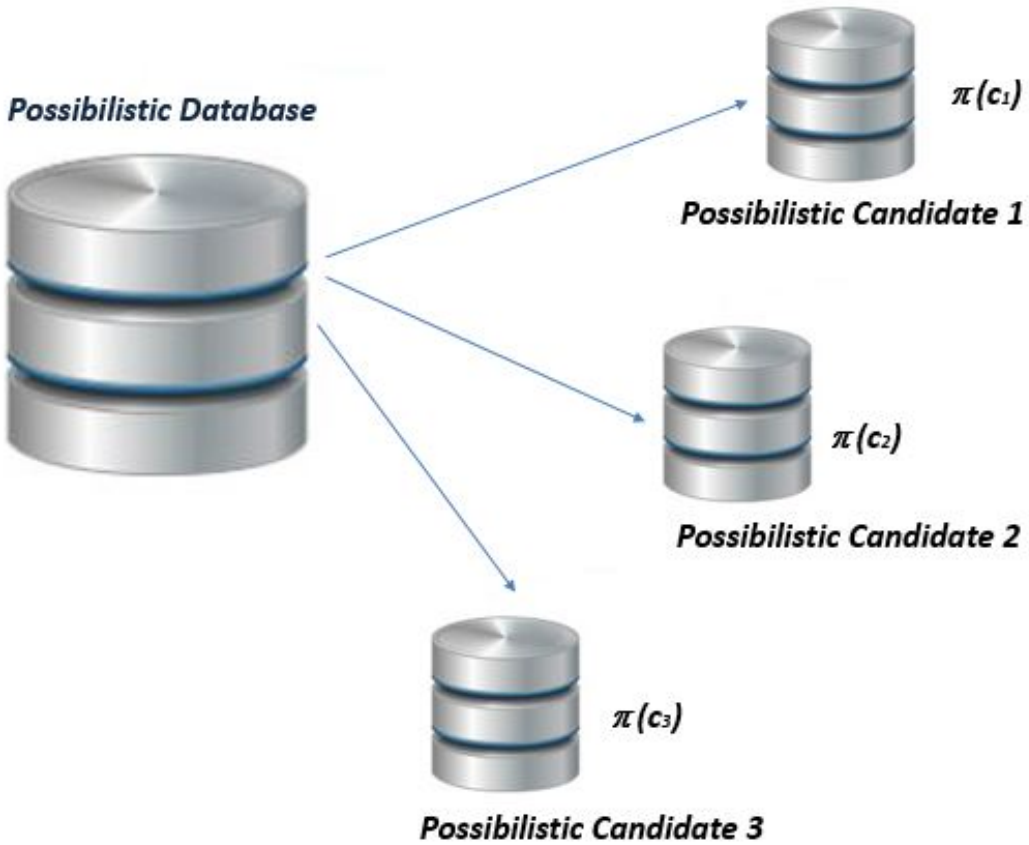


Outline


Imperfect Database
G(c1)
Candidate 1
G(c2)
Candidate 2
G(c3)
Candidate 3



[Abiteboul, Kanellakis and Grahne, 1991]



[Bosc, Liétard and Pivert, 2002]


Evidential Database


?

Generating Possible Worlds from an EDB: Example

EDB

ID	Disease	BloodType
1	Anemia 1	B 0.3 {B, O} 0.7
2	{Asthma, Flu} 1	A 1

Generating Possible Worlds from an EDB: Example

W_1

ID	Disease	BloodType
1	Anemia	B
2	Asthma	A

W_2

ID	Disease	BloodType
1	Anemia	B
2	Flu	A

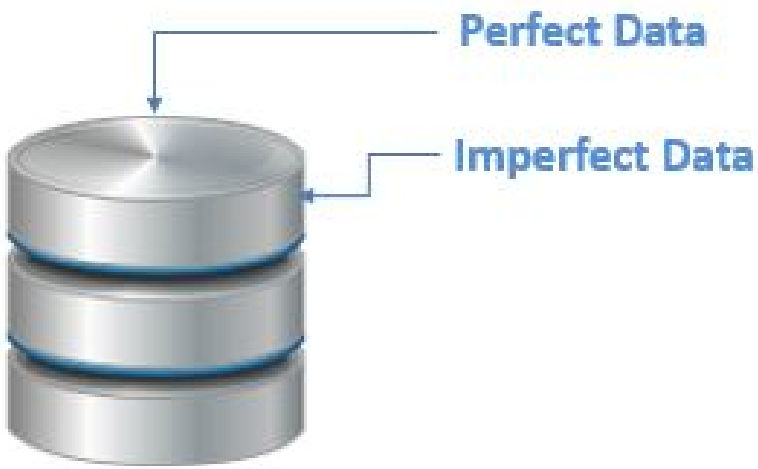
W_3

ID	Disease	BloodType
1	Anemia	O
2	Asthma	A

W_4

ID	Disease	BloodType
1	Anemia	O
2	Flu	A

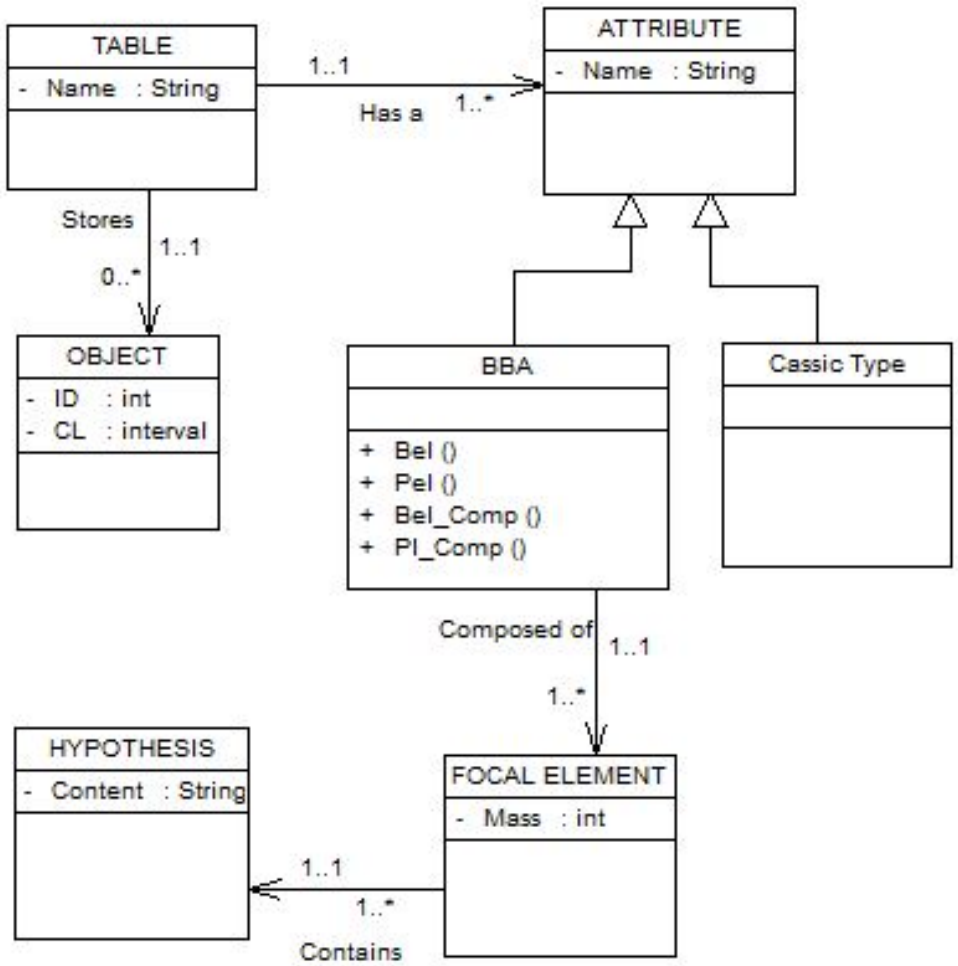
$$m(\{W_1; W_2\}) = 0.3 \qquad m(\{W_1; W_2; W_3; W_4\}) = 0.7$$



Evidential Database

The non existence of an implementation of an Evidential Database in the literature

Evidential Database Design



Implementation

SQL3 implementation:

- Facilitate the complex structure of an evidential database
- Optimize the I/O cost
- Accelerate the information extraction thanks to the indexes

Implementation

Table: Contribution of caches to queries re-execution

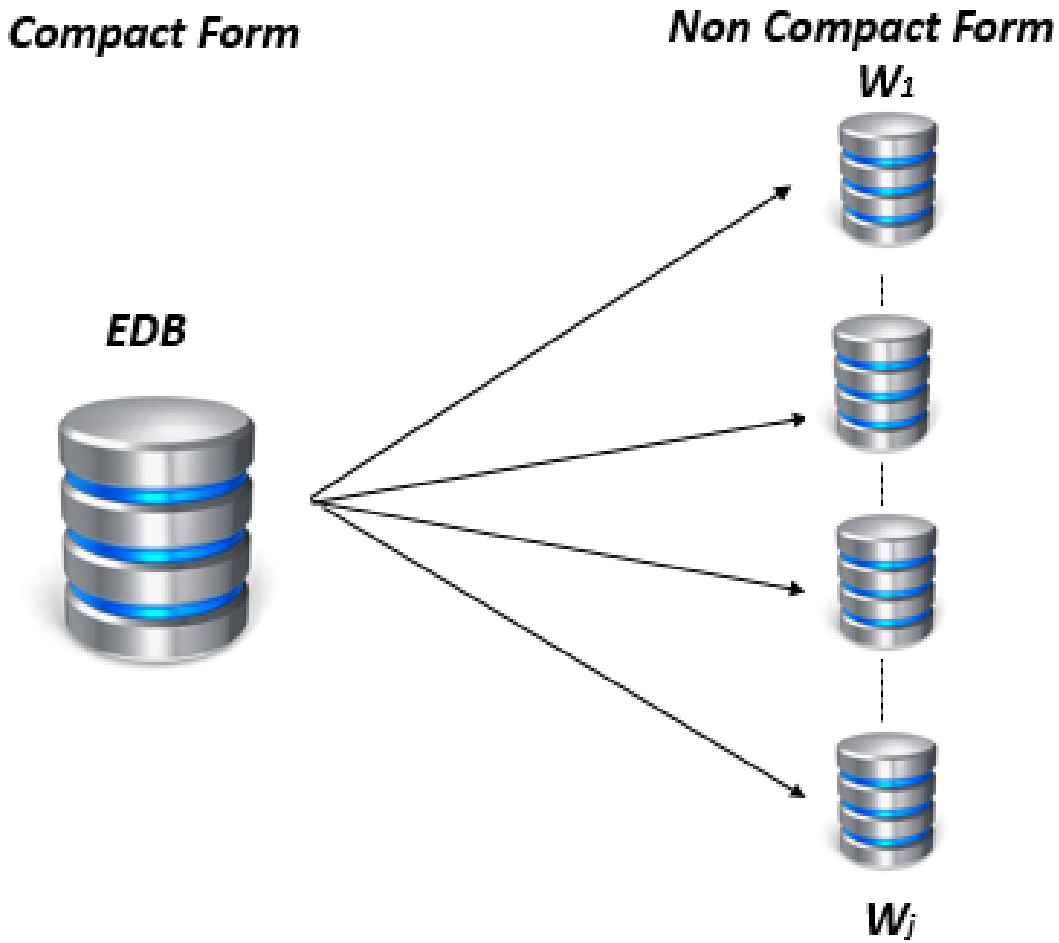
Database size	**First execution time (s)**	**Next executions' times (s)**
1000	0.2	0.03
5000	0.8	0.04
10 000	4.1	0.06
50 000	5.8	0.12
70 000	6.4	0.16
100 000	17	1.2

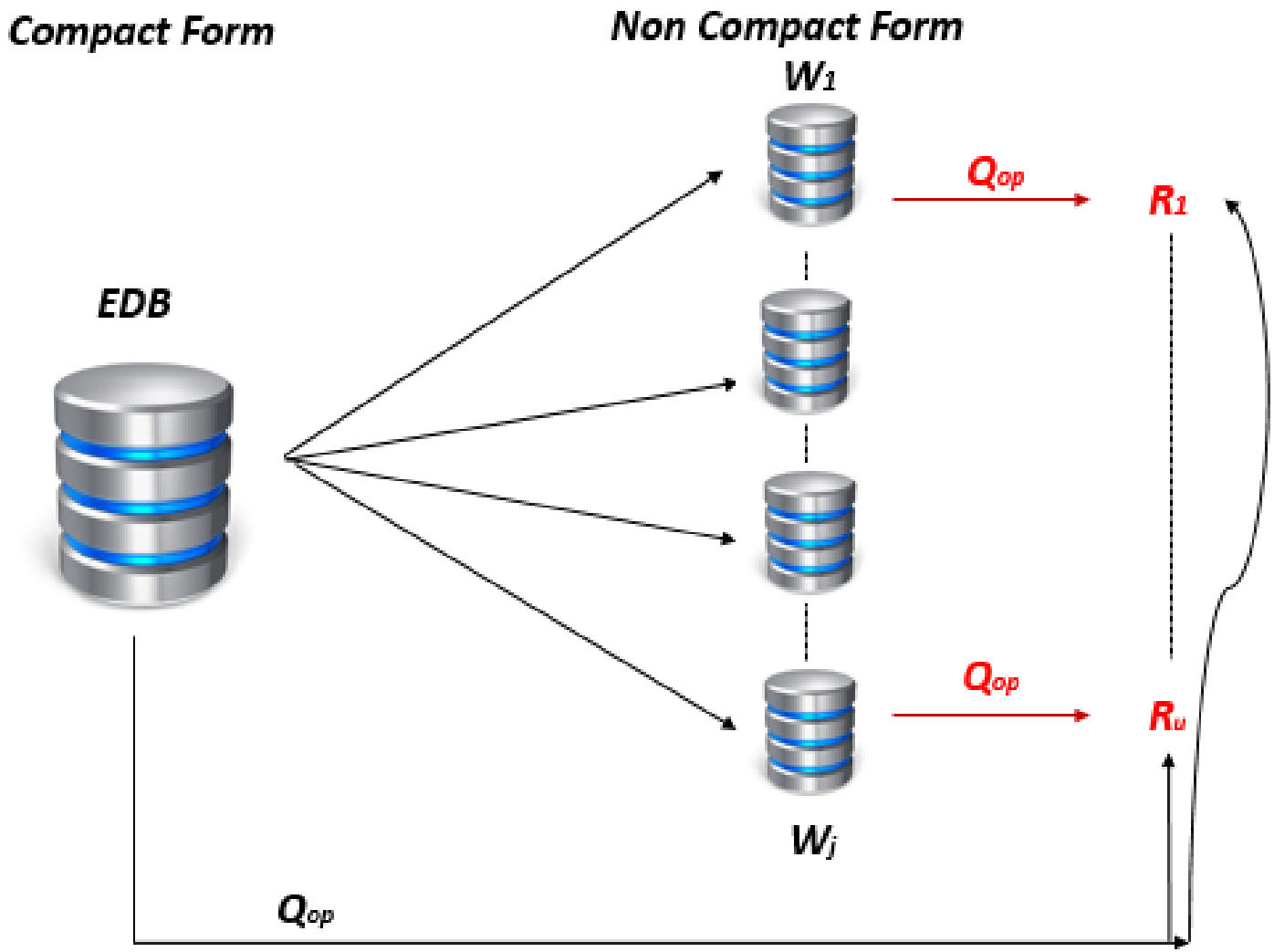
Outline

$\Longrightarrow$ The compact form is the only feasible model in practice but the non-compact form is **fundamental** to prove if the model is a **Strong Representation System**.

Strong Representation System

A model is a *strong representation system* (SRS), when the result of querying *the compact form* is *equivalent* to the result of querying the set of *its possible worlds (the non-compact form)*.


Compact Form
Non Compact Form
W_1
EDB
W_j


Compact Form
Non Compact Form
W_1
Q_{op}
R_1
EDB
Q_{op}
R_u
W_j
Q_{op}

Querying Possible Worlds

Definition

Let Q be the query processed on each possible world W_j. Querying each possible world W_j, gives a possible answer $R = Q(W_j)$

Querying Possible Worlds

Definition

Let Q be the query processed on each possible world W_j. Querying each possible world W_j, gives a possible answer $R = Q(W_j)$

Confidence Level

For each tuple t belonging to the set of responses $Q(W_u)$, a confidence level CL is calculated. t.CL = [bel(S), pl(S)]; $S = \{W_u | t \in Q(W_u)\}$

Querying Possible Worlds

Definition

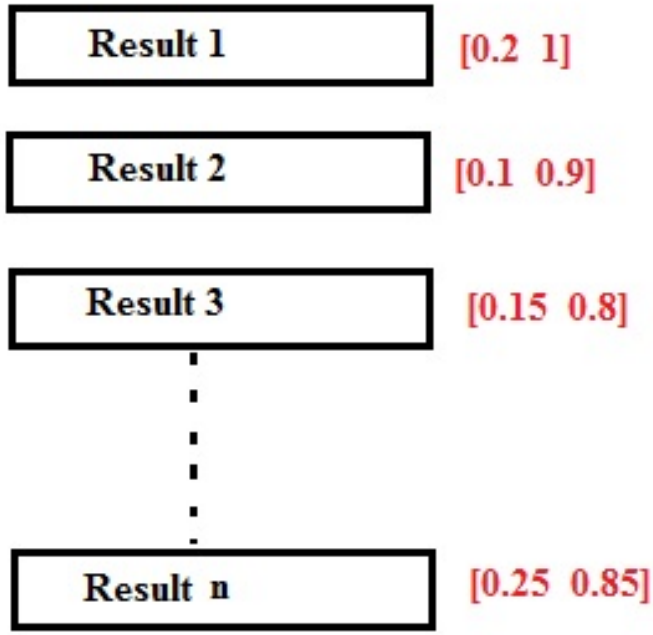
Let Q be the query processed on each possible world W_j. Querying each possible world W_j, gives a possible answer $R = Q(W_j)$

Confidence Level

For each tuple t belonging to the set of responses $Q(W_u)$, a confidence level CL is calculated. t.CL = [bel(S), pl(S)]; $S = \{W_u | t \in Q(W_u)\}$

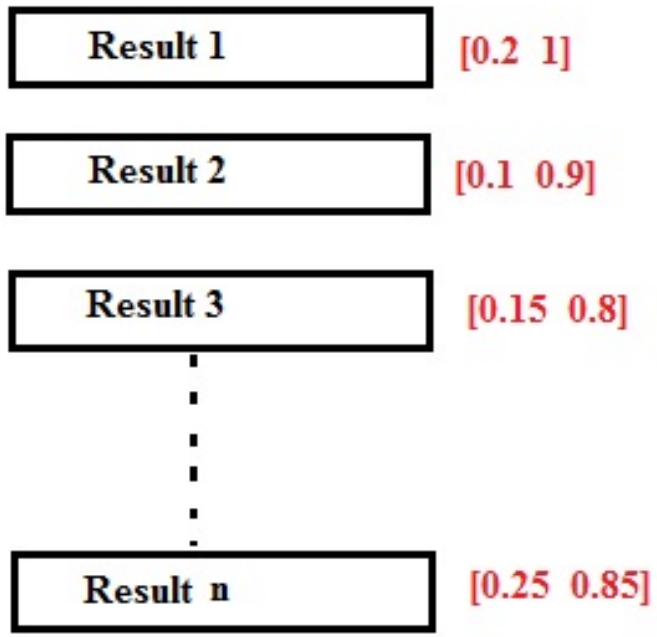
- Extended Evidential Select Operator
- Extended Evidential Project Operator

Evidential Database



Which are the k best results ?

How to rank them ?



- Top-k queries also known as Ranking queries.

- Top-k queries also known as Ranking queries.
- Introduced in the multimedia systems by Fagin in 1996.

- Top-k queries also known as Ranking queries.
- Introduced in the multimedia systems by Fagin in 1996.
- Order queries' results in order to give only the most interesting answers.

- Top-k queries also known as Ranking queries.
- Introduced in the multimedia systems by Fagin in 1996.
- Order queries' results in order to give only the most interesting answers.
- Top-k queries order answers using a score function that returns the k important answers with the highest scores.











H-index






H-index







H-index


ID	BookRate	CL
1	b_1 0.3 $\{b_2, b_3\}$ 0.7	[0.5 ; 1]
2	b_2 0.5 b_4 0.5	[0.3 ; 0.8]
3	$\{b_1, b_2, b_3\}$ 1	[1 ; 1]
4	b_3 1	[0.5; 0.9]

Top-k Query

The top-2 most appreciated books for readers ?

- Introduced Evidential Score

item	EvidentialScore
b_1	R_1=[0.0375 ; 0.325]
b_2	R_2=[0.0375 ; 0.525]
b_3	R_3= [0.125 ; 0.65]
b_4	R_4= [0.0375 ; 0.1]

- Adopted Preference Degree Equation

item	EvidentialScore
b_1	R_1=[0.0375 ; 0.325]
b_2	R_2=[0.0375 ; 0.525]
b_3	R_3= [0.125 ; 0.65]
b_4	R_4= [0.0375 ; 0.1]

Final Ranking

$b_3 \succ b_2 \succ b_1 \succ b_4$.
The Top-2 appreciated books are :

- b_3 with a confidence level [0.125 ; 0.65]
- b_2 with a confidence level [0.0375 ; 0.525]

- Defined A new Semantic for Imperfect Databases

Etop-2 Semantics

Books b_3 and b_2 are the 2 most appreciated credible answers from the set of results.

Outline

Conclusion

- We introduced the possible worlds' model of an Evidential Database.

Conclusion

- We introduced the possible worlds' model of an Evidential Database.
- We implemented the Evidential Database model on its compact form.

Conclusion

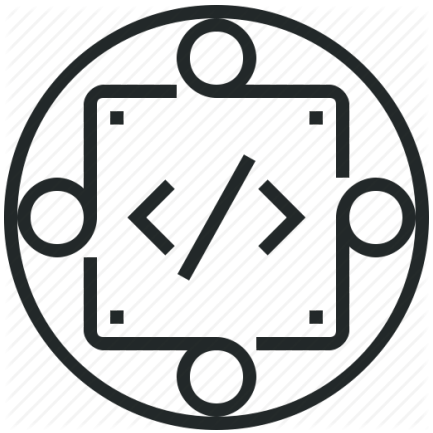
- We introduced the possible worlds' model of an Evidential Database.
- We implemented the Evidential Database model on its compact form.
- We introduced two extended relational evidential queries: Selection and Projection.

Conclusion

- We introduced the possible worlds' model of an Evidential Database.
- We implemented the Evidential Database model on its compact form.
- We introduced two extended relational evidential queries: Selection and Projection.
- We introduced a new imperfect top-k query, called Evidential Top-k Query.

Future Works

- Implementation of Etop-k queries.



Future Works

- Proof of the Strong Representation System for Evidential Databases.

A		B	
a_1	1	$\{b_1, b_2\}$	0.2
		b_1	0.8
$\{a_1, a_2\}$	1	b_2	1

A	B
a_1	b_1
a_1	b_2

A	B
a_1	b_1
a_2	b_2

A	B
a_1	b_2
a_1	b_2

A	B
a_1	b_2
a_2	b_2

Fatma Ezzahra Bousnina, Mohamed Anis Bach Tobji, Mouna Chabbeh, Ludovic Liétard, Boutheina Ben Yaghlane. A New Formalism for Evidential Databases, In procceedings of 22nd International Symposium for Intelligent Systems, LNCS, 9384, Lyon, pages 31-40, 21-23, October, 2015.

Fatma Ezzahra Bousnina, Sayda Elmi, Mohamed Anis Bach Tobji, Mouna Chabbeh, Allel HadjAli, Boutheina Ben Yaghlane. Object-Relational implementation of Evidential Databases, In procceedings of 1st International Conference on Digital Economy, IEEE, Carthage, pages 80-87, 21-23, April, 2016.

Fatma Ezzahra Bousnina, Mouna Chabbeh, Mohamed Anis Bach Tobji, Allel HadjAli, Boutheina Ben Yaghlane. On Top-k Queries over Evidential Data, In procceedings of 19th International Conference on Enterprise Information Systems, LNBIP, Porto, 26-29, April, 2017.

Fatma Ezzahra Bousnina, Sayda Elmi, Mouna Chabbeh, Mohamed Anis Bach Tobji, Allel HadjAli, Boutheina Ben Yaghlane. Skyline Operator over Tripadvisor Reviews within the Belief Functions Framework, In procceedings of 2nd International Conference on Digital Economy, LNBIP, Sidi Bousaid, 04-06, May, 2017.

Thank You