Using the Dempster-Shafer theory of evidence to resolve ABox inconsistencies

Andriy Nikolov
Victoria Uren
Enrico Motta
Anne de Roeck

KNOWLEDGE MEDIA KMi INSTITUTE

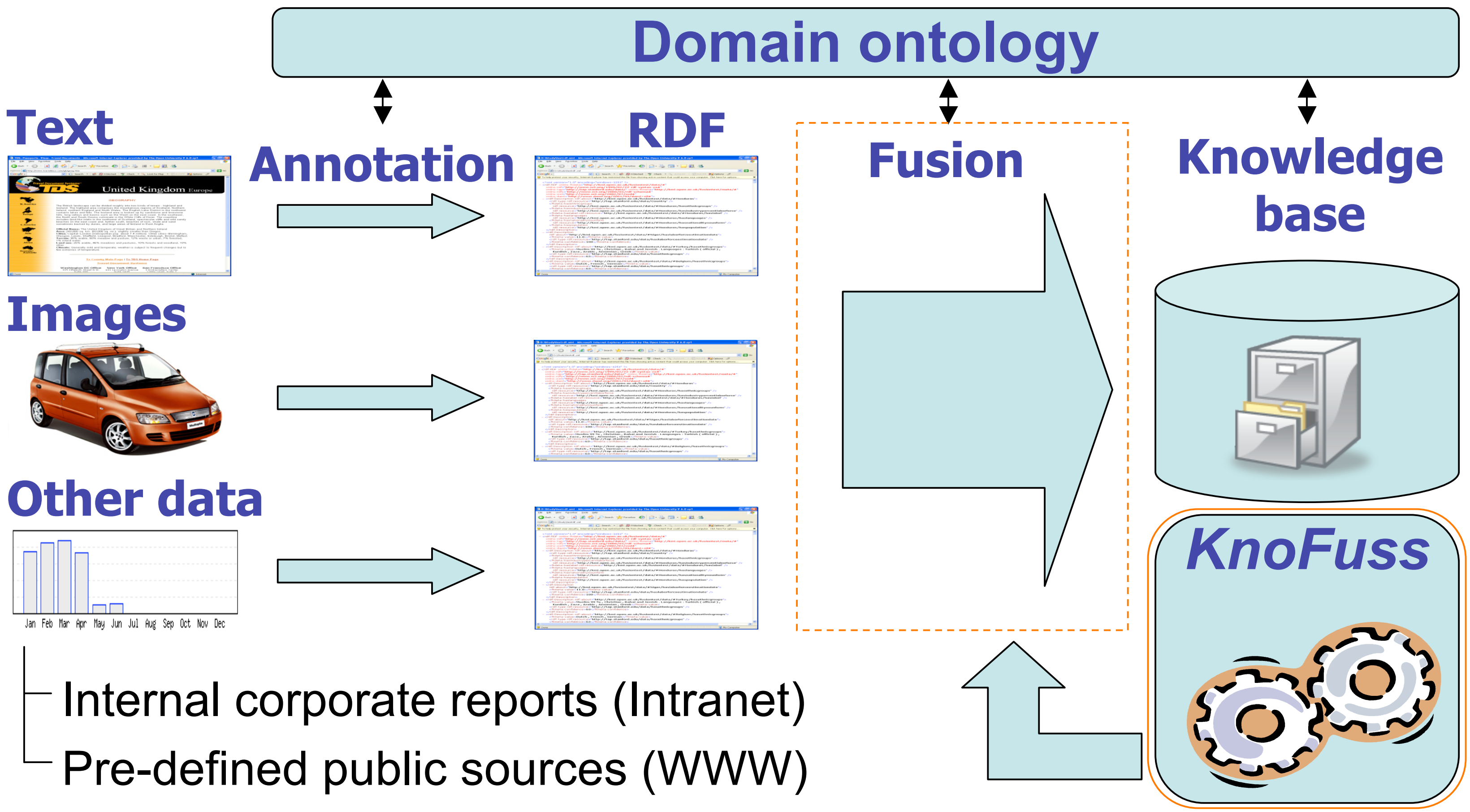
X-MEDIA

The Open University

Outline

- Motivation
- Algorithm description
- Limitations and future work

Knowledge fusion scenario



Fusion problems

- Ambiguity
  - Object identification (instance-level ontology matching)
- **Inconsistency**
  - Detecting and localizing conflicts
  - Processing conflicts (ranking the alternatives)
  - Use uncertainty
- Sources of uncertainty
  - Extraction errors
  - Obsolete data
  - Unreliable sources

conference1	rdfs:label	ISWC2007
conference2	rdfs:label	ISWC2007+ASWC2007
conference3	rdfs:label	ISC2007

conference1	rdfs:label	ISWC2007
	rdfs:label	ISWC2007+ASWC2007
conference3	rdfs:label	ISC2007

conference1	rdfs:label	ISWC2007	0.8
	rdfs:label	ISWC2007+ASWC2007	0.6
conference3	rdfs:label	ISC2007	

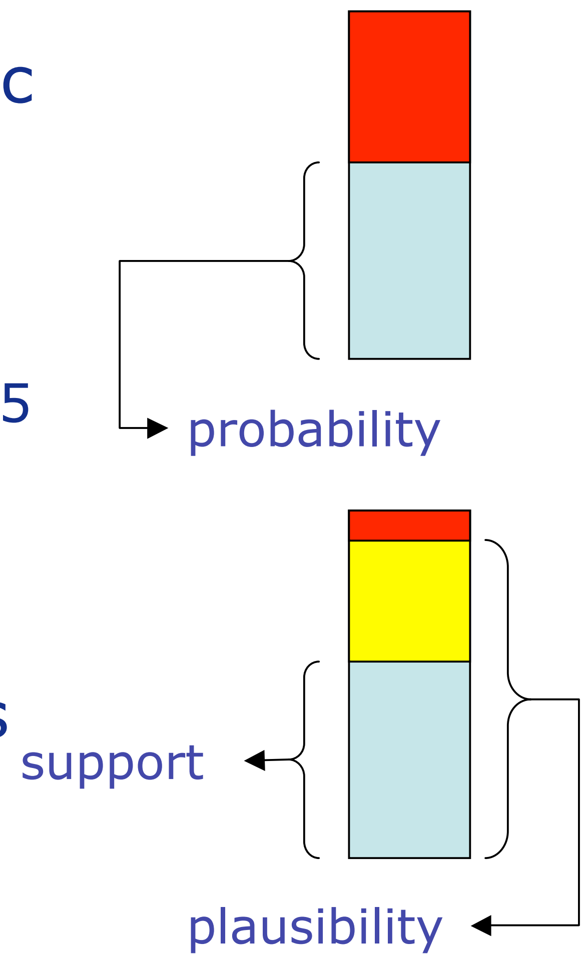
Motivation

- Fuzzy logic
  - Different interpretation:
    degree of vagueness
    vs degree of confidence
- Probability intervals [de Campos et al 2004]
  - Using *max* and *min* combination operators
  - Hard to represent cumulative evidence
- Bayesian probability
  - Appropriate but has disadvantages…

Dempster-Shafer theory of evidence

- Bayesian probability theory:
  Assigning probabilities to atomic alternatives:
  - p(true)=0.6 ! p(false)=0.4
  - Sometimes hard to assign
  - Negative bias:
  - Extraction uncertainty less than 0.5 ! negative evidence rather than insufficient evidence

Dempster-Shafer theory:
  Assigning confidence degrees (masses) to sets of alternatives
  - m({true}) = 0.6
  - m({false}) = 0.1
  - m({true;false})=0.3

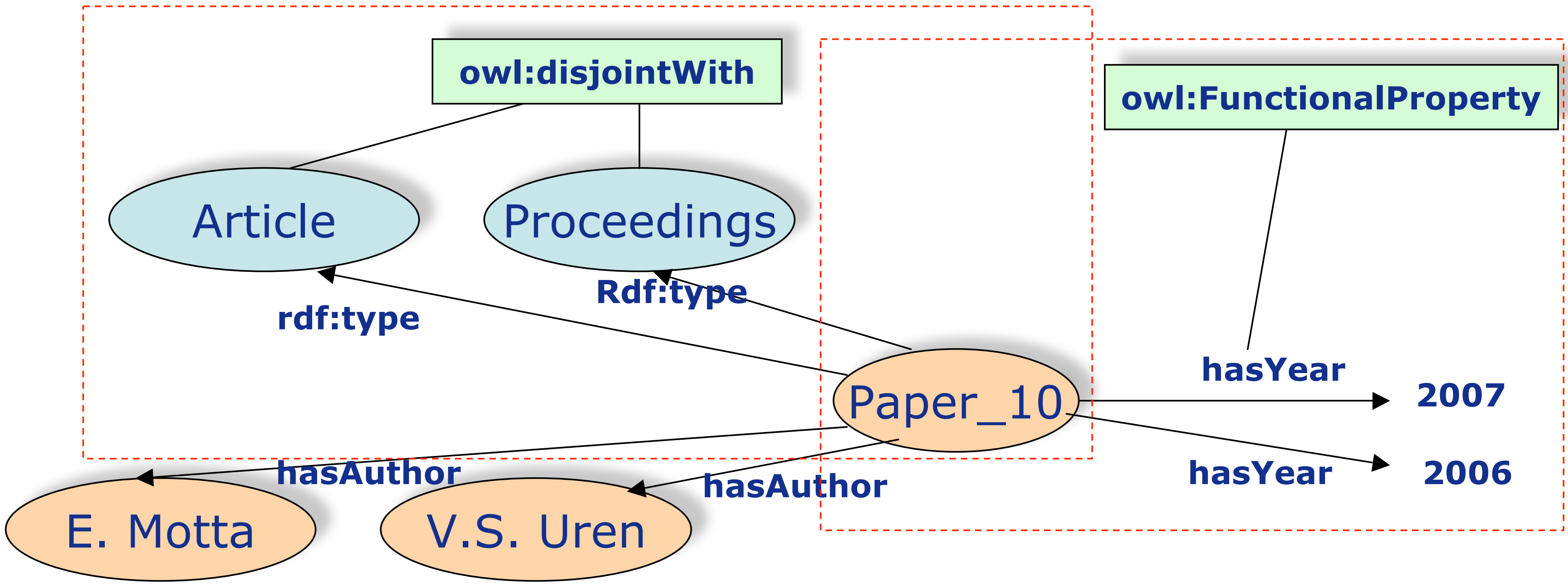


Outline

- Motivation
- Algorithm description
  - Conflict detection
  - Building belief networks
  - Belief propagation
- Limitations and future work

Conflict detection

- Goal: select minimal conflict sets (statements that together produce an inconsistency)

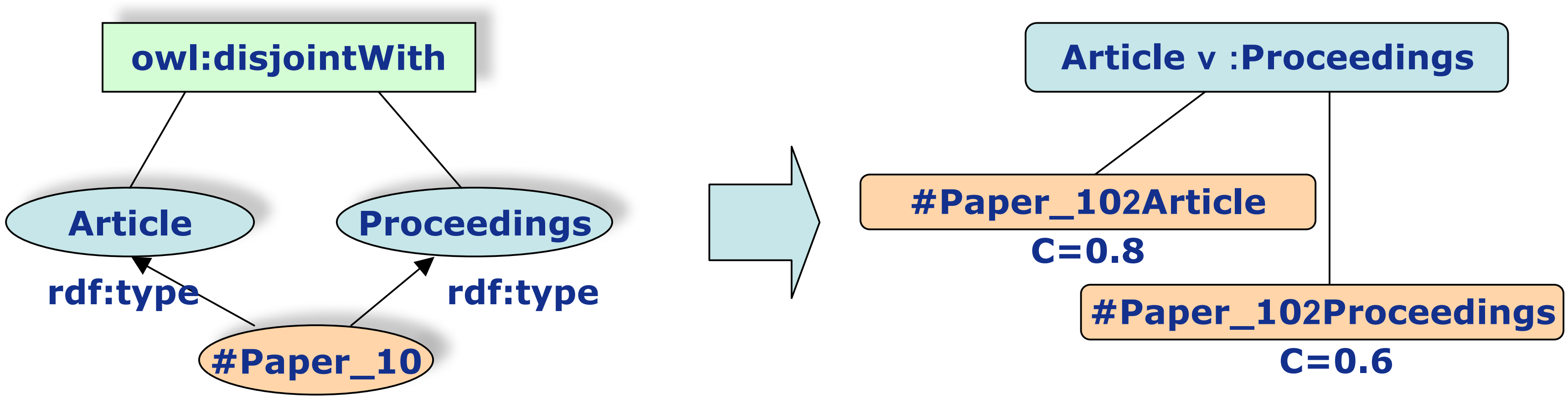


Conflict detection

- Identifying the cause of inconsistency
  - Using an OWL reasoner with diagnostic capabilities
  - initially Pellet
  - moving to RaDoN
- Find all minimal conflict sets
  - Use Reiter's hitting set algorithm [Reiter 1987]

Belief networks

- Valuation networks [Shenoy and Shafer 1990]
  - {ª, {Ω_{Ξ}}$_{X2ª}$, {¿$_1$,…,¿$_n$},+,-} – undirected graph
  - ª : variables – ABox statements
  - {Ω_{Ξ}}$_{X2ª}$: variable states – {true;false}
  - {¿$_1$,…,¿$_n$} : belief potentials
  - + : marginalization operator
  - - : combination operator – Dempster’s rule



Belief networks (cont)

- Network nodes – OWL axioms
  - Variable nodes
    - ABox statements (I2X, $R(I_1, I_2)$)
    - One variable – the statement itself
  - Valuation nodes
    - TBox axioms (XtY)
    - Mass distribution between several variables (I2X, I2Y, I2XtY)

Belief networks (cont)

- Belief network construction
  - Using translation rules
  - Rule antecedents:
    - Existence of specific OWL axioms (one rule per OWL construct)
    - Existence of network nodes
  - Example rule:
    - Explicit ABox statements:
      IF I2X THEN CREATE $N_1(I2X)$
    - TBox inferencing:
      IF *Trans*(R) AND EXIST $N_1(R(I_1, I_2))$ AND EXIST $N_2(R(I_2, I_3))$ THEN CREATE $N_3(\mathit{Trans}(R))$ AND CREATE $N_4(R(I_1, I_3))$

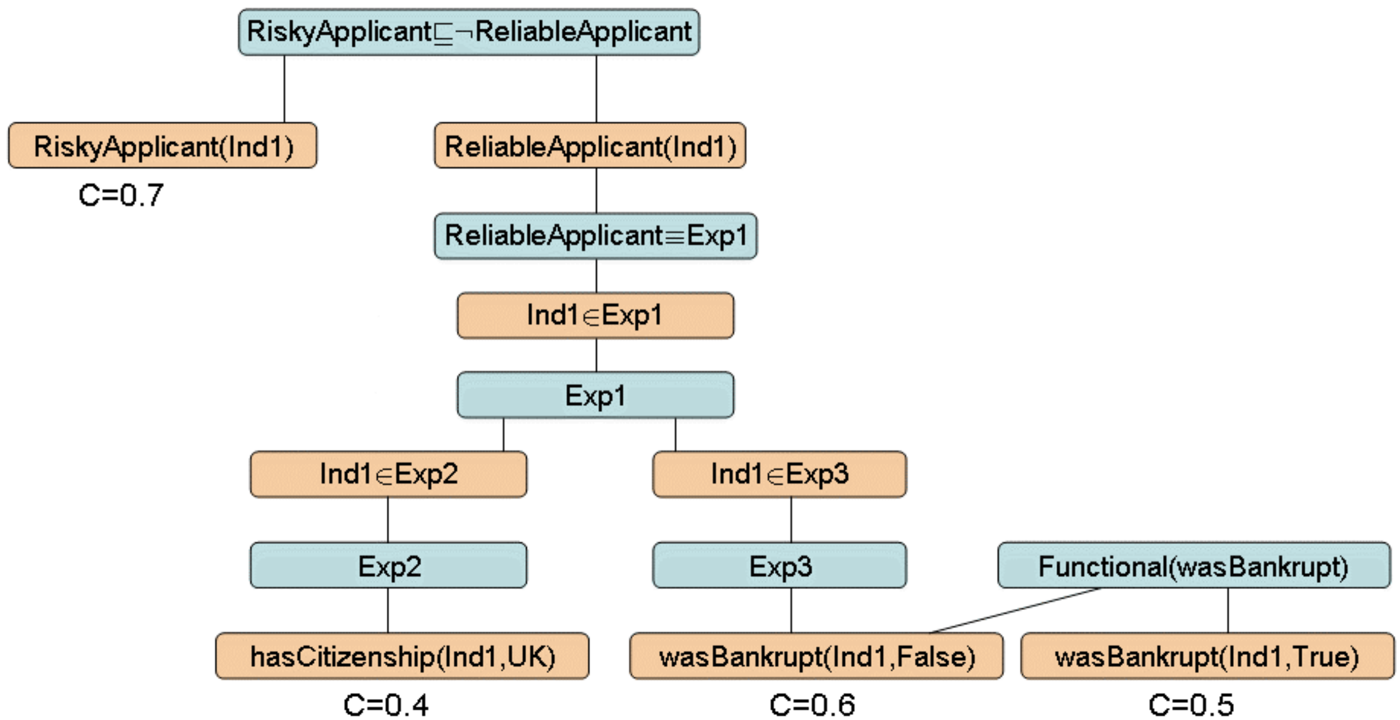
Example ontology

- "Somebody is a reliable applicant if (s)he has a UK citizenship and has never been bankrupt before"
- TBox
  - RiskyApplicantvLoanApplicantu:ReliableApplicant
  - ReliableApplicant´LoanApplicantu
    9wasBankrupt.Falseu
    9hasCitizenship.UK
  - >v·1wasBankrupt
- ABox
  - Ind12RiskyApplicant Sup = 0.7
  - hasCitizenship(Ind1, UK) Sup = 0.4
  - wasBankrupt(Ind1, False) Sup = 0.6
  - wasBankrupt(Ind1, True) Sup = 0.5

Example ontology

- Conflict set 1
  - RiskyApplicantvLoanApplicantu:ReliableApplicant
  - ReliableApplicant´LoanApplicantu
    9wasBankrupt.Falseu
    9hasCitizenship.UK
  - Ind12RiskyApplicant Sup = 0.7
  - hasCitizenship(Ind1, UK) Sup = 0.4
  - wasBankrupt(Ind1, False) Sup = 0.6
- Conflict set 2
  - >v·1wasBankrupt
  - wasBankrupt(Ind1, False) Sup = 0.6
  - wasBankrupt(Ind1, True) Sup = 0.5

Example network



- Exp1=9wasBankrupt.Falseu9hasCitizenship.UK
- Exp2=9hasCitizenship.UK
- Exp3=9wasBankrupt.False

Assigning mass distributions

- Variable nodes
  - Direct assignment of masses to explicit statements
- Valuation nodes
  - crisp logic TBox:
    m({possible states}) = 1
- Example
  - Node: XtY
  - Variables: I2X, I2Y, I2XtY
  - Distribution: m({0;0;0},{0;1;1},{1;0;1},{1;1;1})=1

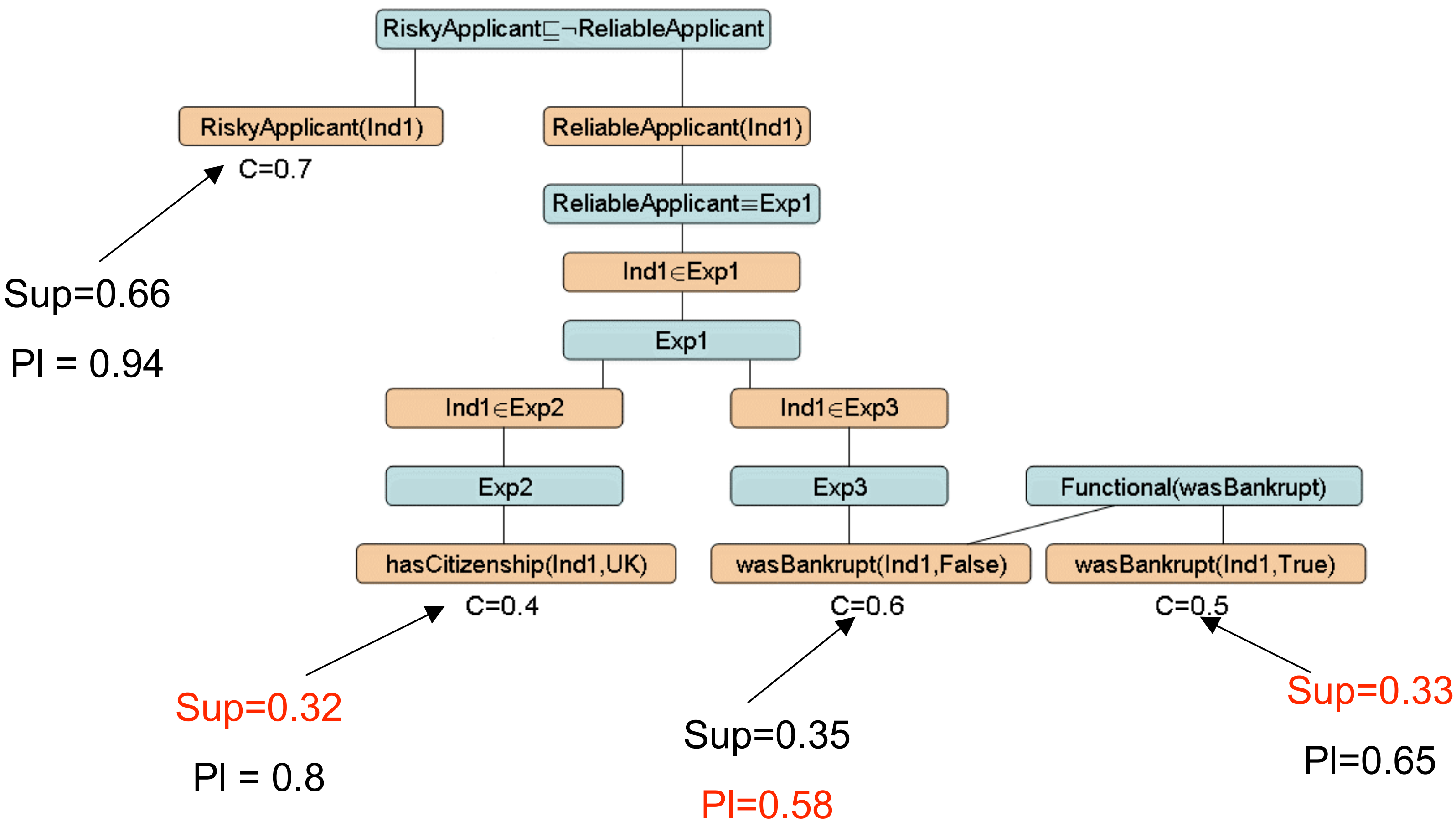
Belief propagation

- Axioms for valuation networks
  [Shenoy and Shafer 1990]
- Marginalization

  $m^{+C}(X)=\sum_{Y^{+C}=X} m(Y)$
- Combination (Dempster's rule)

  $m_1\text{-}m_2(X)=(\sum_{X_1 u X_2=X} m_1(X_1)m_2(X_2))/$
  $(1-\sum_{X_1 u X_2=\circledR} m_1(X_1)m_2(X_2))$

Propagation example

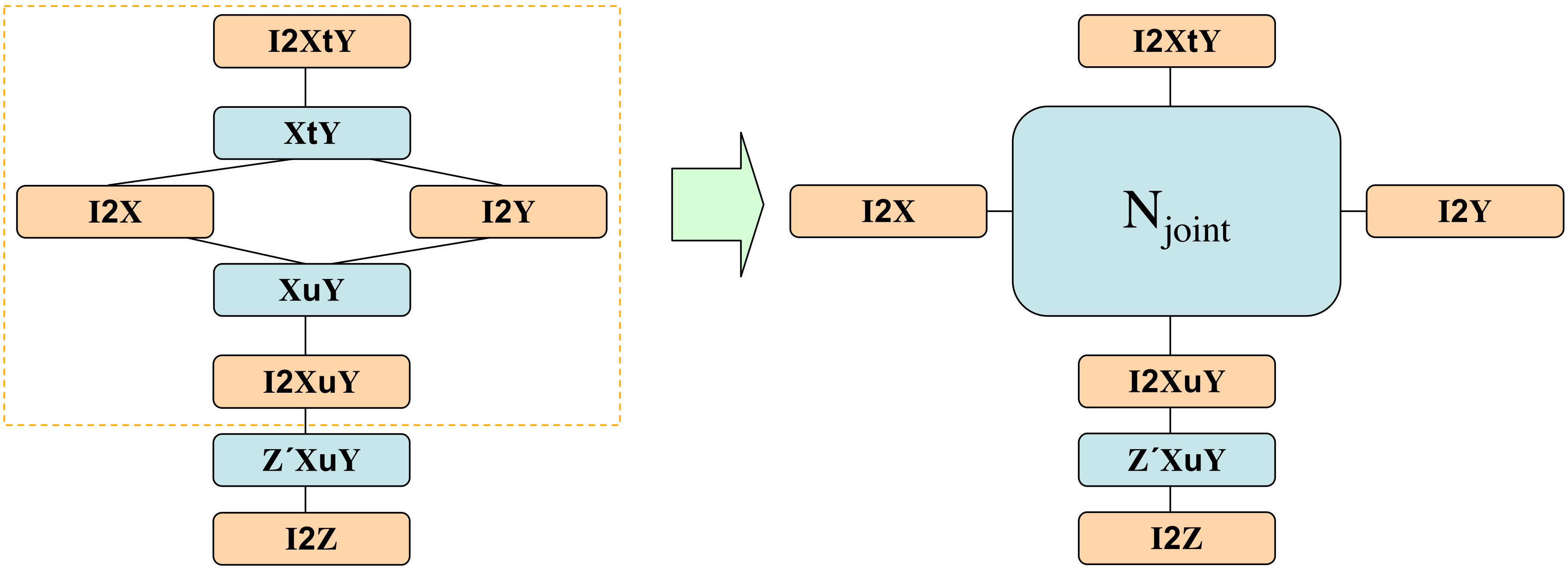


Summary

- Translation
  - From: minimum inconsistent OWL-DL subontology
  - To: valuation network
- Dempster-Shafer theory
  - Representing ignorance
  - Switching to Bayesian probabilities (support only) when needed

Assumptions

- Valuation network has to be a tree
- Elimination of loops
  - Replace a loop with a single node containing joint distribution



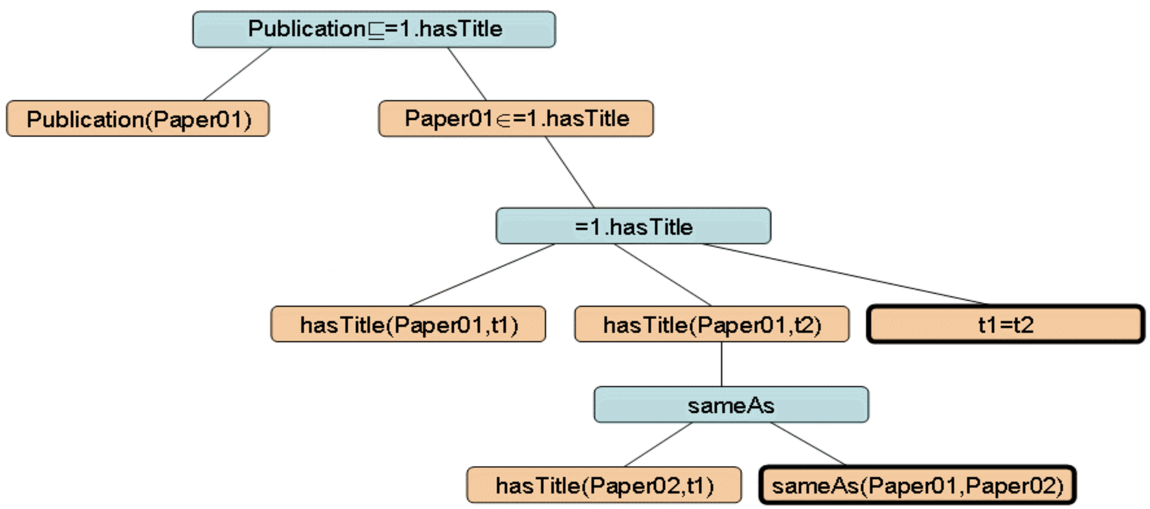
Limitations

- Only applicable to ABox conflicts
  [Pearl 1990]
- Does not consider identity uncertainty

paper1	**hasTitle**	**AquaLog: An Ontology portable Question Answering Interface for the Semantic Web**
paper1	**hasTitle**	**AquaLog: An ontology-driven Question Answering System to interface the Semantic Web**
paper1	**owl:sameAs**	**paper2**

Model extension

- The belief propagation model has to be extended
  - Extracted statements – uncertainty from the extraction methods
  - Auxiliary statements – uncertainty from object identification methods



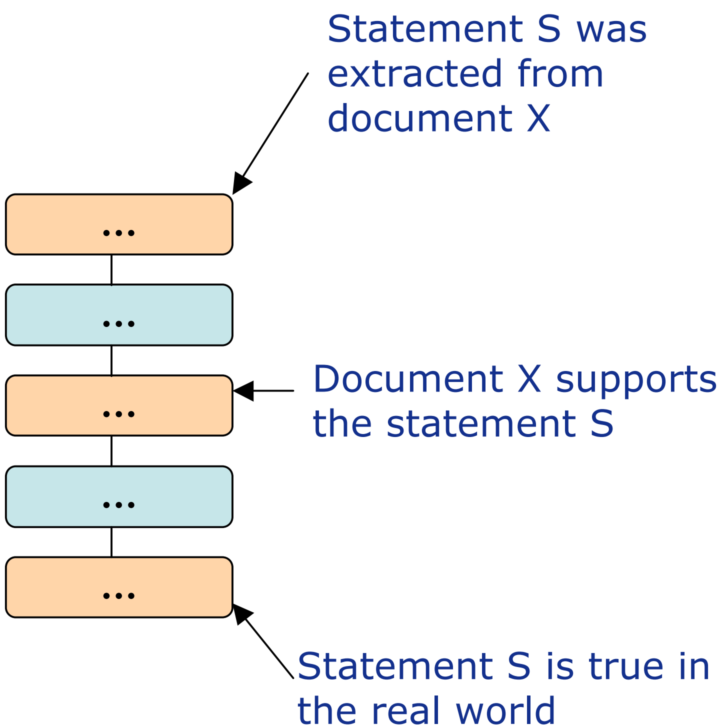
Future Work

- Conducting experiments
  - No project data available (yet…)
  - Scientific publications dataset
  - Geographic data extracted using GATE
- Comparing with Bayesian approach
  - Additional expressivity vs computational complexity

Questions?

Thanks for your attention

Representing ignorance (example)



- Complementary representation
  - Extraction uncertainty
  - Reliability of sources
- Assigning conditional probabilities
  - P(document X does not support the statement S| statement S is correct) = ?