

Panel on uncertainty modeling

Information quality and uncertainty

Marie-Jeanne Lesot

Laboratoire d'Informatique de Paris 6 (LIP6)

Université Pierre et Marie Curie (UPMC)

IPMU'12

July 9-13, 2012



Motivations

- From an application point of view
 - **uncertainty = key concept of information quality**
 - information quality \sim information usefulness
 - potential impact: relevance, importance, seriousness, precision
 - attached confidence: degree of certainty
 - information quality = important issue
 - essential for information management and decision making
- From a theoretical point of view
 - raises challenging issues
 - covers a **wide range of uncertainty types**
- Fictitious running example

on July 8th, the International Olympic Committee declared
“In 2024, the Summer Olympic Games will probably take place in November”

Content-related uncertainty: what is said?

2024 Summer Olympic Games dates = November

- Comparison with background knowledge: **plausibility**
 - a priori confidence
 - equivalent to a surprise effect
- Comparison with other pieces of information: **credibility**
 - looking for corroboration
 - confirmation or invalidation by other pieces of information

Formulation-related uncertainty: how is it said?

“In 2024, the Summer Olympic Games **will probably** take place in November”

- Information presentation: **linguistic uncertainty**
 - as expressed by the chosen words, e.g. adverbs, modal verbs
 - possibly, probably, certainly, might, may, could, should, must, ...
- Linguistic **imprecision**
 - approximate compatibility to the query: partial answer
 - e.g. November → around the end of the year

Source-related uncertainty: who says it?

The [International Olympic Committee](#) declared

“In 2024, the Summer Olympic Games will probably take place in November”

- Expertise level: **competence**
 - is the source qualified on the subject?
 - e.g. the IOC \neq a taxi driver
- Source trust level: **reliability**
 - reputation, trustworthiness
 - sincerity, intention, commitment degree
- Issue of successive sources: hearsay
 - A says that B says that ... that F

Context-related uncertainty: when is it said?

On **July 8th**, the IOC declared

“In 2024, the Summer Olympic Games will probably take place in November”

- **Temporal context**

- distance between information and reported content
- dynamic evolution of confirmation/invalidation
- evolution of the source characteristics

- Context of other assertions: **credibility**

- corroboration degree wrt confirmation/invalidation
- possibly refined by affinity/hostility/independence relation between the sources

Automatic processing-related uncertainty: how is it extracted?

- **Natural language processing:** e.g.
 - linguistic uncertainty modeling
 - e.g. probably → moderate ; undeniably → low
 - negation handling
- **Automatic extraction tools:** e.g.
 - date recognition
 - named entity recognition: acronym resolution, aliases
 - relationship extraction

Several uncertainty type classifications

- **Objective vs subjective** uncertainty
 - objective: recognition rate, confirmation/invalidation
 - subjective: content plausibility, source reliability, competence
- **General vs contextual** uncertainty
 - general: source reliability, recognition rates, linguistic modeling
 - contextual: source competence, content credibility, temporal dimensions
- **Different levels** of uncertainty
 - structural doubts about the phenomena: content plausibility, recognition rate
 - imprecision level: linguistic imprecision
 - subjective level: source reliability and competence

Formal modeling frameworks for information scoring

- Probabilities
 - in particular for automatic tool uncertainty: error recognition rates
- Evidence theory
 - “Information evaluation in fusion: a case study”. L. Cholvy. IPMU 2004
- Extended multivalued logic
 - “Trust-moderated information-likelihood. A multi-valued logics approach”. A. Revault d’Allonnes, H. Akdag, and O. Poirel. CiE 2007.
- Possibility theory
 - “Proposition of a semi-automatic possibilistic information scoring process”. M.-J. Lesot, T. Delavallade, F. Pichon, H. Akdag, B. Bouchon-Meunier, P. Capet. EUSFLAT 2011

Some other references for information quality and uncertainty

- “Annex to STANAG2022”. NATO, Information handling systems. 1997.
- “An extension of STANAG2022 for information scoring”.
J. Besombes and A. Revault d’Allonnes. Fusion 2008
- “Reasoning about trust: a formal logical framework”.
R. Demolombe. iTrust 2004.
- “Relevance and truthfulness in information correction and fusion”.
F. Pichon, D. Dubois, T. Dencœux. IJAR 2011.
- “Uncertainty detection for information extraction”.
B. Goujon. Int. Conf. on recent advances in natural language processing, 2009.
- “Assessment and analysis of information quality: a multidimensional model and case studies”.
L. Berti-Équille, I. Comyn-Wattiau, M. Cosquer, Z. Kedad, S. Nugier, V. Peralta,
S. Si-Saïd Cherfi, V. Thion-Goasdoué. Int. J. of Information Quality 2011.