Ontologies & Management of Partially Complete Data

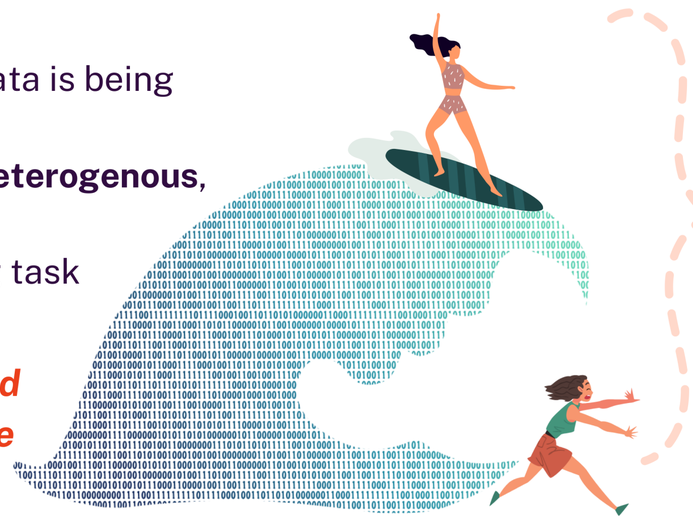
Sanja Lukumbuzya
TU Wien

Ontology-based data accesss should support data with both complete and incomplete parts, but how hard is this task?

BIG DATA FLOOD

- A staggering amount of data is being generated every day
- This data is **incomplete**, **heterogenous**, and often **lacks structure**
- Processing it is a daunting task

How do we avoid drowning in the data and instead learn to surf the wave?



ONTOLOGIES

- are **logical theories**
- are **high-level descriptions** of some domain
- offer an **intuitive vocabulary** for posing queries
- facilitate **data integration** and **knowledge sharing**
- improve **answer quality** through **background knowledge**

ONTOLOGY-BASED DATA ACCESS (OBDA)

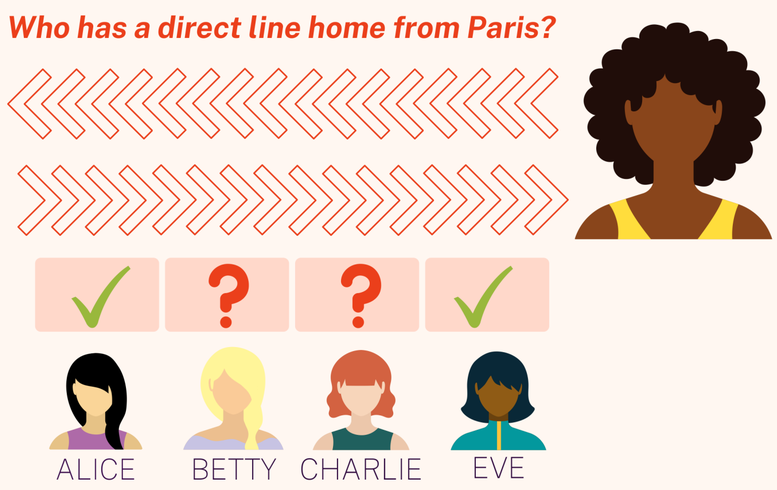
NAME	HOME
ALICE	VIENNA
BETTY	DELFT
CHARLIE	NULL
EVE	BRUSSELS

FLIGHT NR	FROM	TO
JU 1357	PARIS CDG	BELGRADE
FR 2987	PARIS CDG	VIENNA
EJU8323	PARIS ORY	LONDON LHR
W6 2829	PARIS BVA	BRATISLAVA

LINE	FROM	TO	NOTE
258	PARIS GDN	BRUSSELS	VIA LILLE
351	PARIS GDN	LONDON	EXPRESS

$$\text{Person} \sqsubseteq =1\ \text{livesIn.City}$$
$$\text{Location} \sqsubseteq =1\ \text{in.City}$$
$$\text{TrainLine} \sqsubseteq \text{PTL}$$
$$\text{Flight} \sqsubseteq \text{PTL}$$
$$\text{PTL} \sqsubseteq =1\ \text{from.Location}$$
$$\text{PTL} \sqsubseteq =1\ \text{to.Location}$$
$$\text{ExpressTL} \equiv \text{TrainLine} \sqcap \neg \exists\, \text{stopsAt.T}$$

Who has a direct line home from Paris?



query answering = logical inference

OPEN OR CLOSED?

- Standard OBDA assumes **information incompleteness** and makes the **open-world assumption** (OWA)

  anything not explicitely stated is considered unknown

- Traditional DB systems assume completeness and make the **closed-world assumption** (CWA)

  anything not explicitely stated is considered false

Which one do we choose?

WE NEED BOTH!

- We want **languages** that allow us to view **some parts of the data under the CWA**, and **others under the OWA**
- Prominent approaches mostly based on **Description Logics (DLs) = decidable FOL in funny syntax:**
  - Circumscription
  - Epistemic queries
  - Combinations of **DLs** and **logic rules**
  - **Closed predicates**

OUR RESEARCH

- **Closed predicates:** specify which ontology predicates are interpreted exactly as given in the data
- Known to sometimes **increase the complexity** of reasoning
- But, for a **large class of queries** and **ontologies** written in **expressive DLs** that are close relatives of the Web Ontology Language, closed predicates **do not make reasoning harder** [1]
- In fact, answering such ontology-mediated queries can be **polynomially reduced** to reasoning in a **variant of Datalog** [1]
- We can also decide which predicates are **implicitly bounded** by closed predicates, further pushing the **decidability** of query answering [2]

- **DLs + rules:** specify which predicates have their extensions computed by logic rules
- Our framework **Resilient Logic Programs (RLPs)** [3] is based on a combination of DLs and rules but has a novel semantics
- RLPs support systems that should be configured to always **react successfully to the environment whose exact state is unknown to us a priori**
- They use **ontologies** to capture the **unknown state of the environment** and rules to compute the **possible reactions** under given settings

[1] Lukumbuzya, Ortiz, Simkus. Resilient Logic Programs: Answer Set Programs Challenged by Ontologies. AAAI 2020
[2] Gogacz et al. Datalog Rewritability and Data Complexity of ALCHOIF with Closed Predicates. KR 2020
[3] Lukumbuzya, Simkus. Bounded Predicates in Description Logics with Counting. IJCAI 2021

logics LOGICAL METHODS IN COMPUTER SCIENCE
TU WIEN FWF Der Wissenschaftsfonds.