

Crowdsourcing: An Instance of Intensional Data Management

Pierre Senellart



10 September 2019, *Shonan meeting*

Intensional data is everywhere

Lots of data sources can be seen as **intensional**: accessing all the data in the source (**in extension**) is **impossible** or **very costly**, but it is possible to access the data through **views**, with some **access constraints**, associated with some **access cost**.

- **Indexes** over regular data sources
- **Deep Web** sources: Web forms, Web services
- Outcome of **complex automated processes**: information extraction, natural language analysis, machine learning, ontology matching
- **Logical consequences** of facts, costly to compute
- **Private data**: consumes *privacy budget* to acquire
- **Crowd data**: requires worker compensation, and time to acquire



Intensional Data Management

- Take into account the fact that access to data is **limited** and has a **cost** (computational, network, monetary, time, privacy...), to solve a user's **knowledge need**
- **Lazy evaluation** whenever possible
- Evolving probabilistic, structured view of the **current knowledge of the world**
- Solve at each step the problem: **What is the best access to do next** given my current knowledge of the world, the knowledge need, and the cost of the access
- **Knowledge acquisition plan** (recursive, dynamic, adaptive) that minimizes access cost, and provides probabilistic guarantees