



Querying Intensional Data

Pierre Senellart



October 16, 2012



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
- The Web or social networks as partial graphs that can be expanded by **crawling**
- Outcome of **complex automated processes**: information extraction, natural language analysis, machine learning, ontology matching
- **Crowd data**: (very) partial views of the world
- etc.



What we need for querying intensional data

- Answering queries using **views**
- Query feasibility under **access restrictions**
- **Recursive languages**: iterating over answers, feeding back an output value as input to a service, repeated accesses to similar sources
- Plans and cost models for **good sequences of accesses**
- **Uncertainty management**: the views may be uncertain; the outcome of an access can be modeled as an uncertain process
- **Provenance** for correlation management and explanation
- Gathering **statistics** about the behavior of the sources