Integrating Information with Ontologies

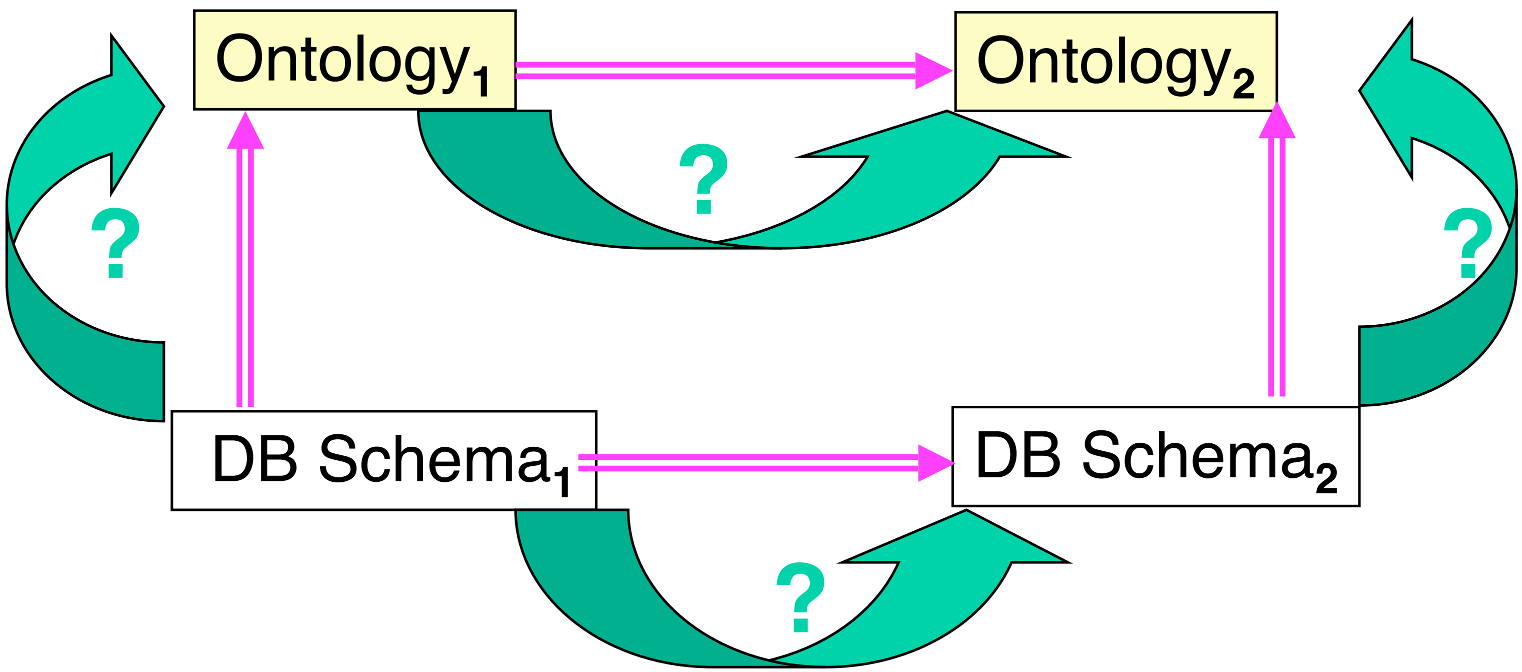
Alex Borgida
Rutgers University

How does one deal with *multiple graphs* since there is no "one and only true ontology"

1. Aligning/mapping ontologies/information sources -- a "hot" topic, with few real successes

2. Importing terms from one ontology to another -- a different approach

Finding complex mappings between information sources

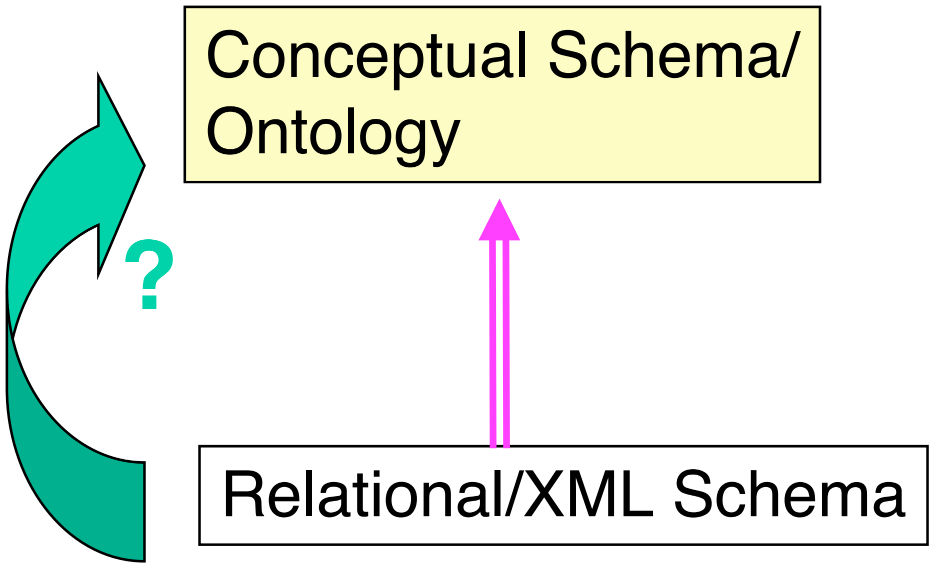


- ***Clio*** **[IBM]: To find** **complex mappings/queries** **from DB Scm$_1$ to DB Scm$_2$ use** **<u>additional information</u>:**

  simple "correspondences" between elements

- **Our work applies this *idea* (but entirely different techniques) to other situations**

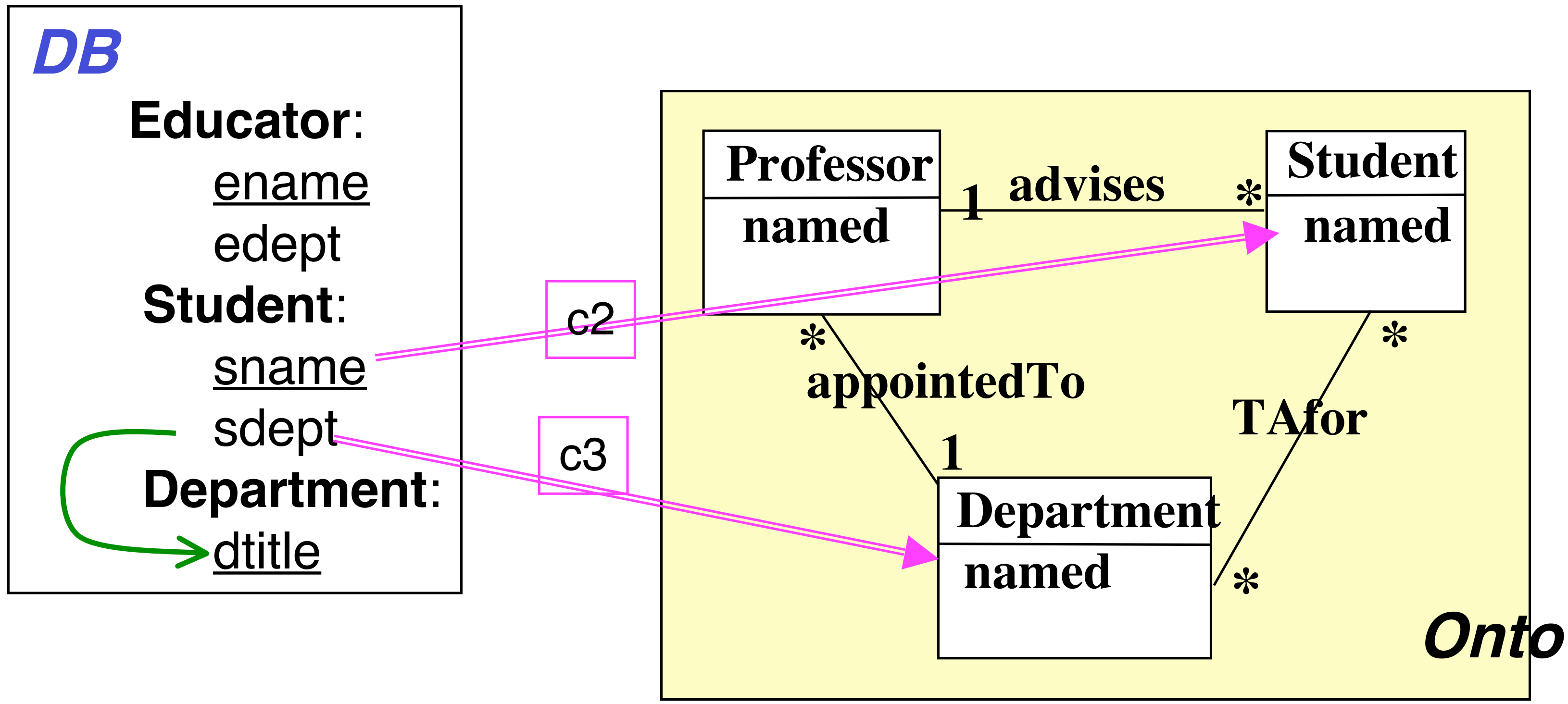
e.g., Finding mappings from DB schemas to Ontologies [with An,Mylopoulos, Miller]



Useful for data integration and service composition

For Example…

INPUT:

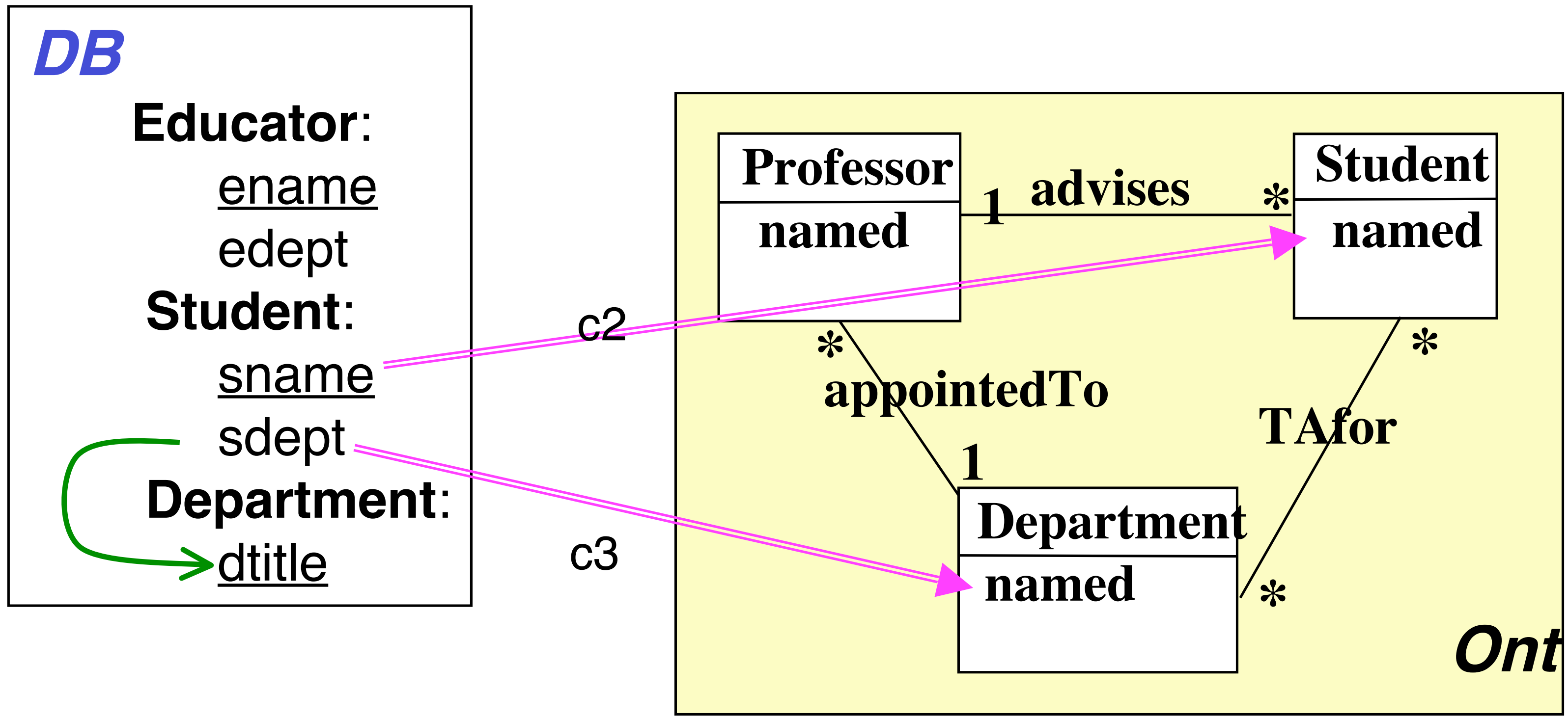


OUTPUT:

```
DB:Student(sname,sdept) →
      O:Student(x), O:Professor(y), O:Department(z),
      named(x,sname), named(z,sdept),
      advises(y,x), appointedTo(y,z).
```

For Example…

INPUT:



OUTPUT:

```
DB:Student(sname,sdept) →
      O:Student(x), O:Professor(y), O:Department(z),
      named(x,sname), named(y,ename), named(z,dtitle),
      advisedBy(x,y), appointedTo(y,z).
```

***Not output:** shortest path!*

```
DB:Student(sname,sdept) → Student(x), Department(z),
named(x,sname),named(z,dtitle),TAfor(x,z).
```

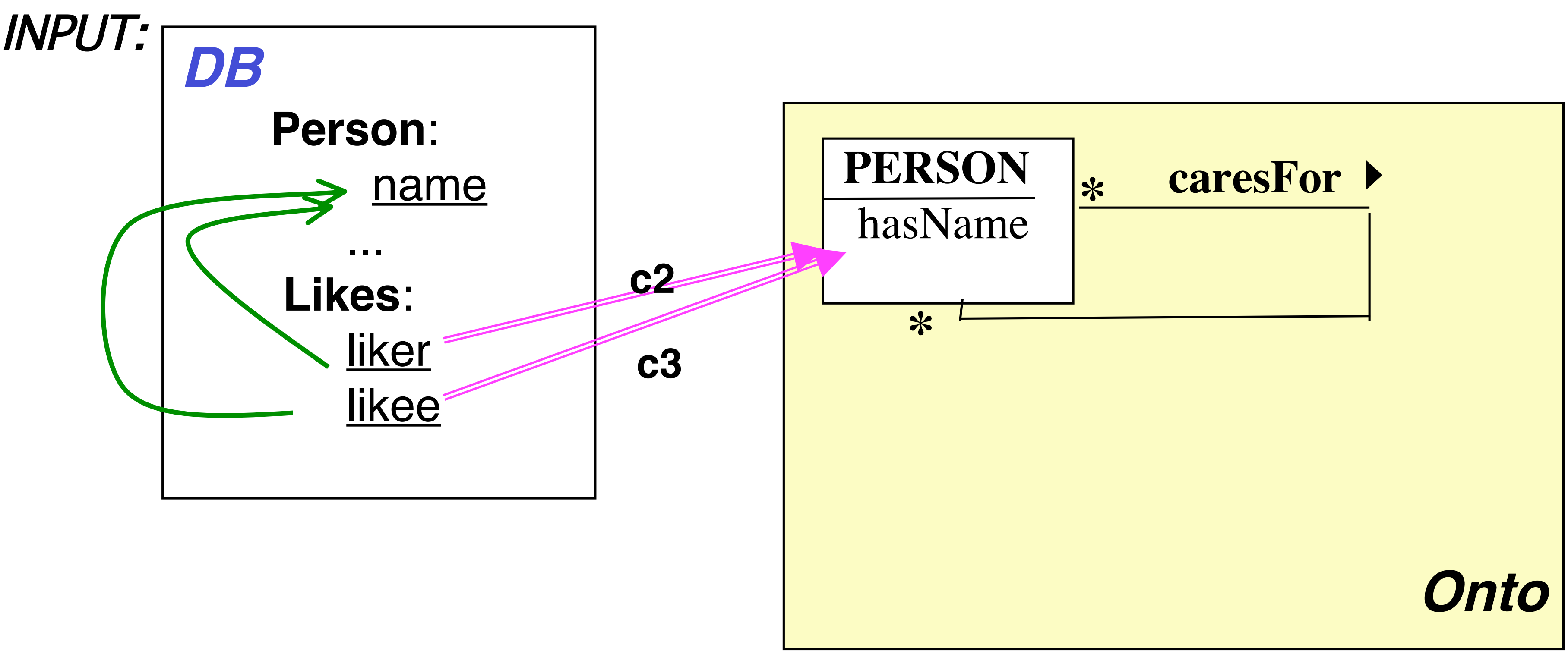
General Strategy

- **Our prototype tool exploits richer semantics in the conceptual schema ("object", association cardinality, IsA hierarchy, disjointness,...)**
- **Also, uses theory of relational schema design from ER diagrams as basis of heuristics**

e.g algorithm

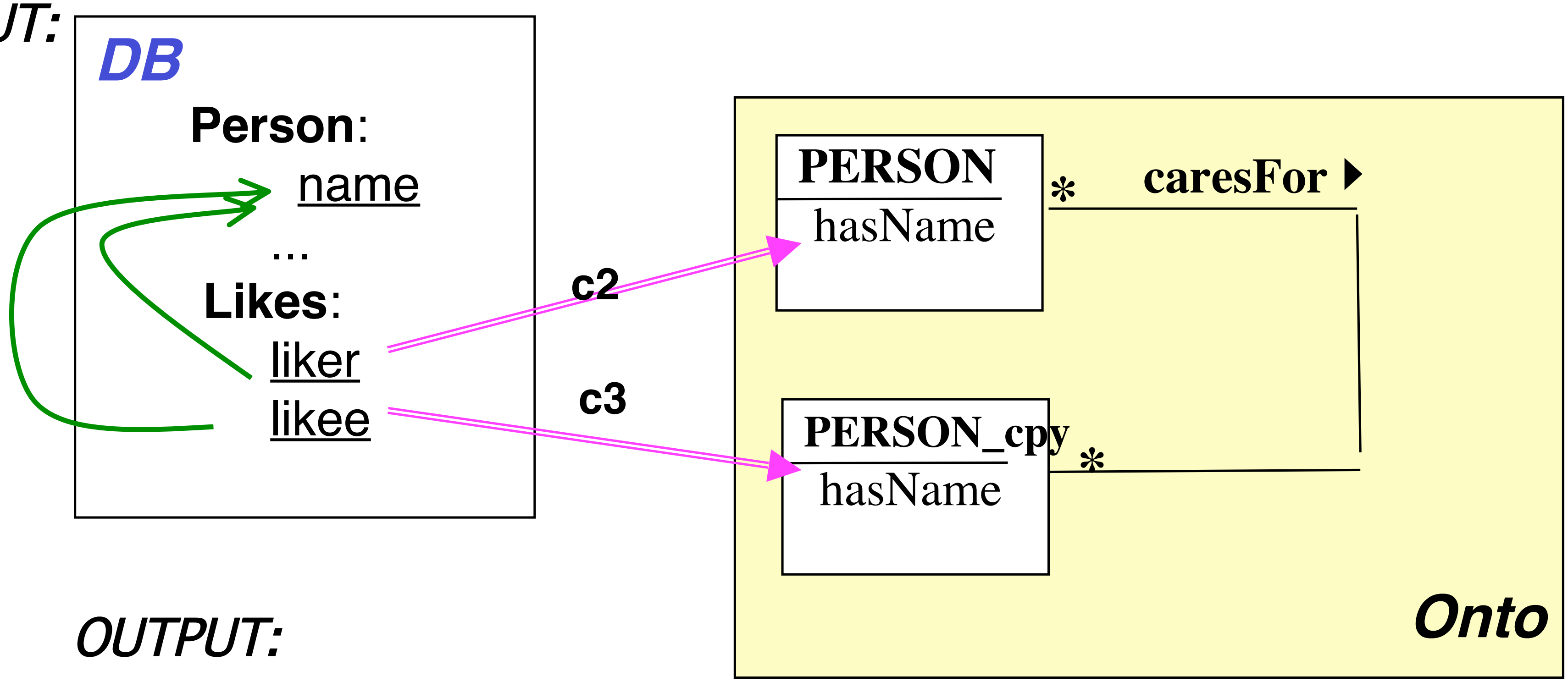
1. **Mark concept nodes having attribute(s) corresponding to table T's column(s).**
2. **Create a "skeleton tree" of marked nodes which have correspondences to the *key columns* of table T, by finding cheapest spanning tree of the "appropriate kind"**
3. **Connect other marked nodes to the skeleton by shortest paths that are "function-like" (cardinality upper bound 1) -** (Because key uniquely determines these values in the table)
4. **Translate resulting tree into Datalog formula by recursive descent. (Can also be translated to OWL!)**

e.g., An interesting complication



Solution: duplicate node

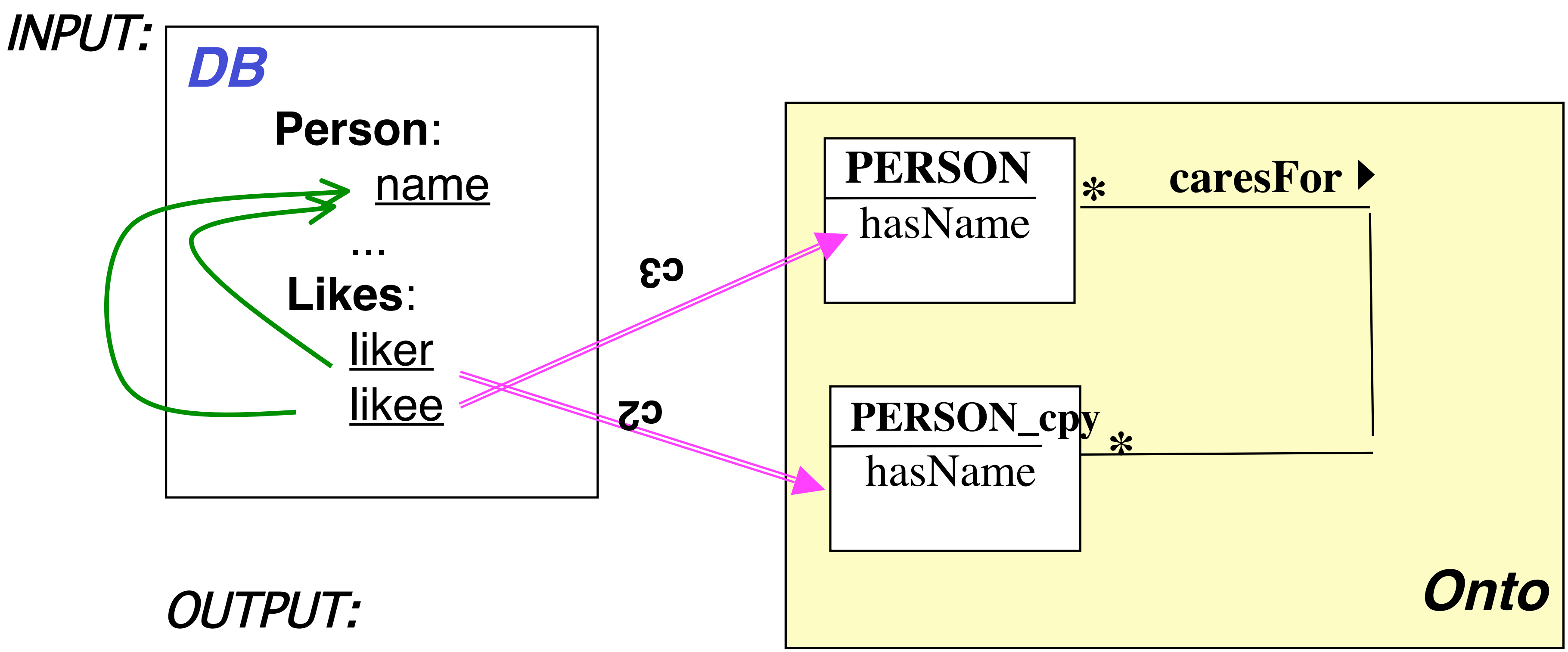
INPUT:



OUTPUT:

```
DB:Likes(liker,likee) →
       O:Person(x), hasName(x,liker),
       O:Person(x), hasName(x,likee),
       caresFor(x,y)
```

Yet another solution!




```
DB:Likes(liker,likee) →
      O:Person(x), hasName(x,likee),
      O:Person(x), hasName(x,liker),
      caresFor(x,y)
```

Somewhat reassuring

If we treat EER diagrams as notation for logic then the standard relational schema design rules found in textbooks assign a specific formula to each table.

- ***Theorem* 1 ("completeness") For any specific table T1 derived from E1, one of the formulas returned by our algorithm is the one assigned by the above design rules.**

- ***Theorem 2 ("soundness")* If the algorithm returns a semantic formula, there was a way of deriving that table from E1 with that semantics.**

The algorithm works in other cases too, but of course not in all ‘denormalized’ schemas.

ontology concept **book** with properties **hasTitle, hasAuthors**

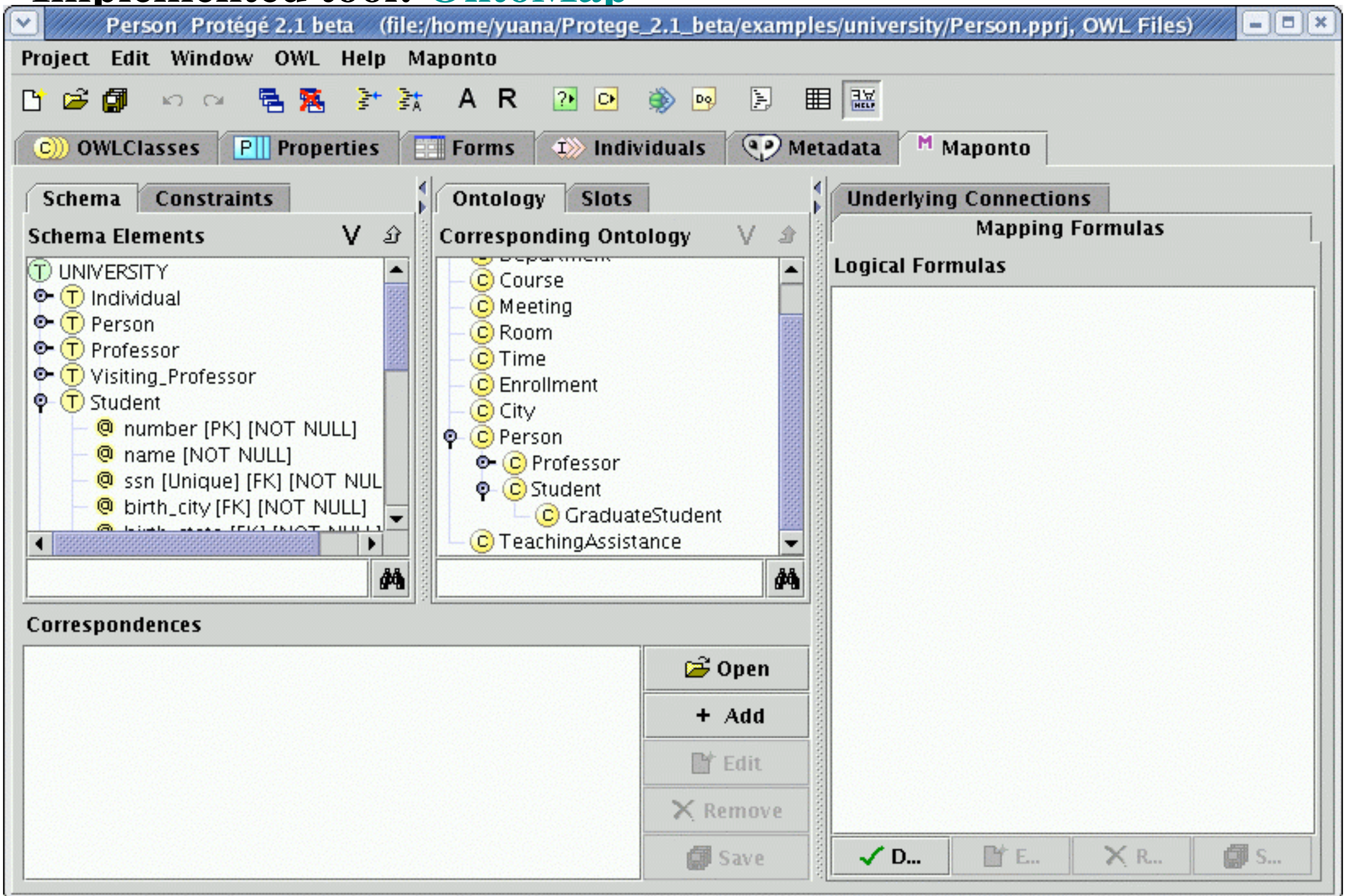
vs

database schema

```
table book(title, first_author, other_authors,...)
```

Empirical Evaluation

- **Implemented tool: OntoMap**



Experiments

- **To measure the usefulness of the tool, we want to know how much can a user benefit from its use.**
- **Suppose that MapOnto returns formula instead of the desired correct formula Ψ. If there are**
  - n **atoms in the returned formula Φ;**
  - m **atoms in the correct formula Ψ;**
  - c **atoms in common to Ψ and Φ**
- **Then tool user needs to delete n-c atoms from Φ and add m-c atoms to Φ.**

Labor savings=1- ((n-c)+(m-c))/m

References

- Yuan An, Alex Borgida, Renee Miller, John Mylopoulos. ``A Semantic Approach to Discovering Schema Mapping Expressions'', *Int. Conf. on Data Engineering ICDE* 2007
- Yuan An, Alex Borgida, John Mylopoulos,``Discovering the Semantics of Relational Tables Through Mappings '', *Journal of Data Semantics VII*, pp.1--32, 2006
- Yuan An, Alex Borgida, John Mylopoulos: ``Constructing Complex Semantic Mappings Between XML Data and Ontologies.'' *International Semantic Web Conference ISWC 2005*: 6-20

Experimental Results: “labor savings” for XML Schema semantics