Faster OWL Using Split Programs

Pascal Hitzler

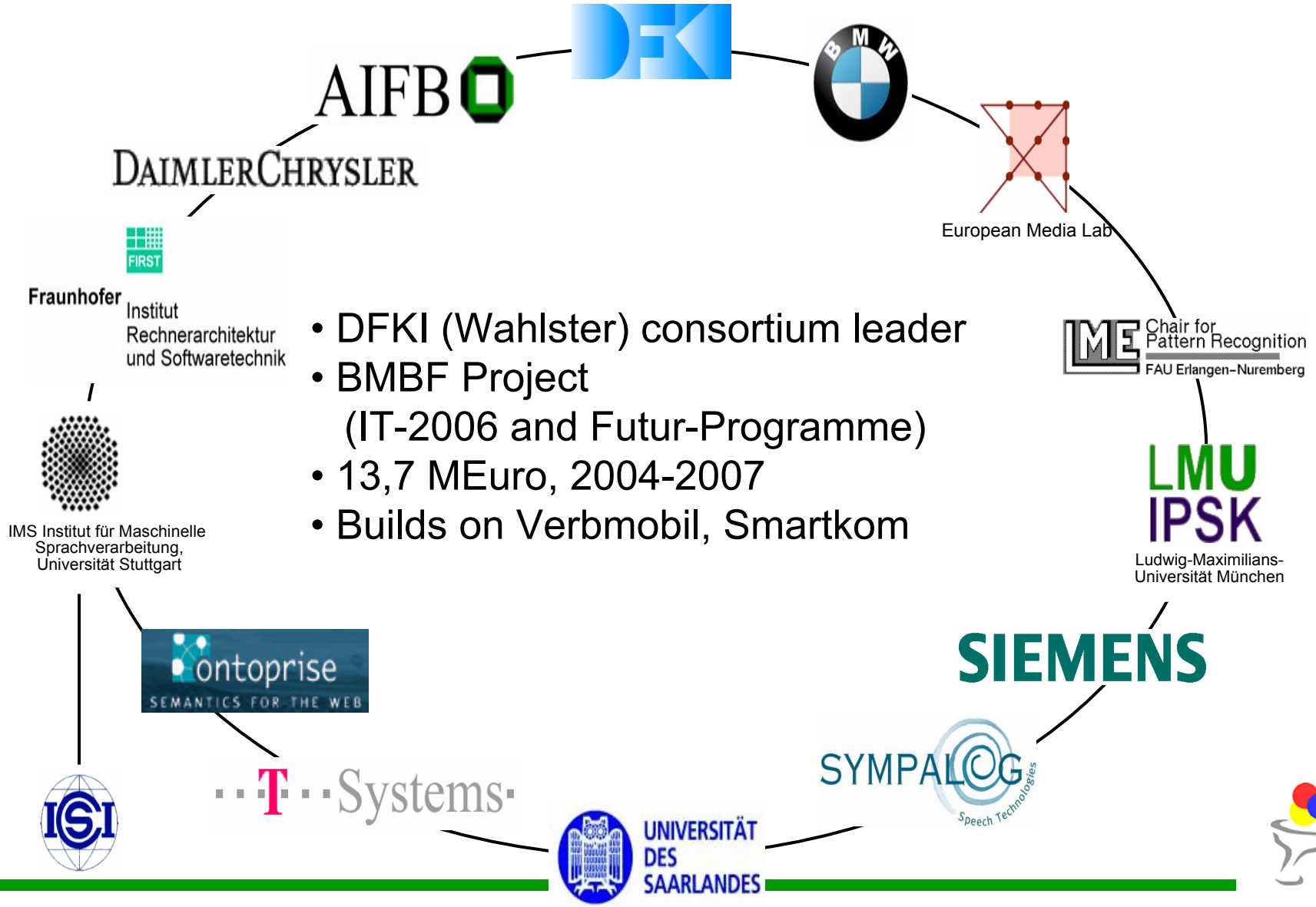
Denny Vrandecic

AIFB, University of Karlsruhe, Germany

Problem Description

- Reasoning with OWL DL is **hard**. (Expressivity vs. scalability)
- For certain Semantic Web applications quick responses are more important than absolute accuracy of answering.
  e.g. SmartWeb scenario.

SmartWeb: Mobile querying of the Semantic Web



- DFKI (Wahlster) consortium leader
- BMBF Project
  (IT-2006 and Futur-Programme)
- 13,7 MEuro, 2004-2007
- Builds on Verbmobil, Smartkom

AIFB

SmartWeb

KnowledgeWeb

Key Idea: Approximate Reasoning

- Reasoning with OWL DL is **hard**. (Expressivity vs. scalability)
- For certain Semantic Web applications quick responses are more important than absolute accuracy of answering.
  e.g. SmartWeb scenario.
- We **trade soundess for time**, using **approximate reasoning**.

Concrete Approach

To appear in: Hitzler & Vrandecic, Resolution-based approximate reasoning for OWL DL. In: Proceedings of ISWC2005, Galway, Ireland, November 2005.

We facilitate recent results due to

Hustadt, Motik, Sattler, Studer 2003/2004/2005

on **casting OWL-DL into disjunctive Datalog**.
(currently being implemented in **KAON2**
see http://kaon2.semanticweb.org)

Obtaining speed-up

1. Get rid of nominals. [linear]
2. Translate TBox into clausal form. [linear]
3. Saturate TBox by taking all consequences. ~~[exponential]~~
4. Eliminate function symbols. [linear]
5. Performing inferences. ~~[NP]~~ ⇒ **[P]**

Transforming disjunctive Datalog into non-disjunctive Datalog using split programs.

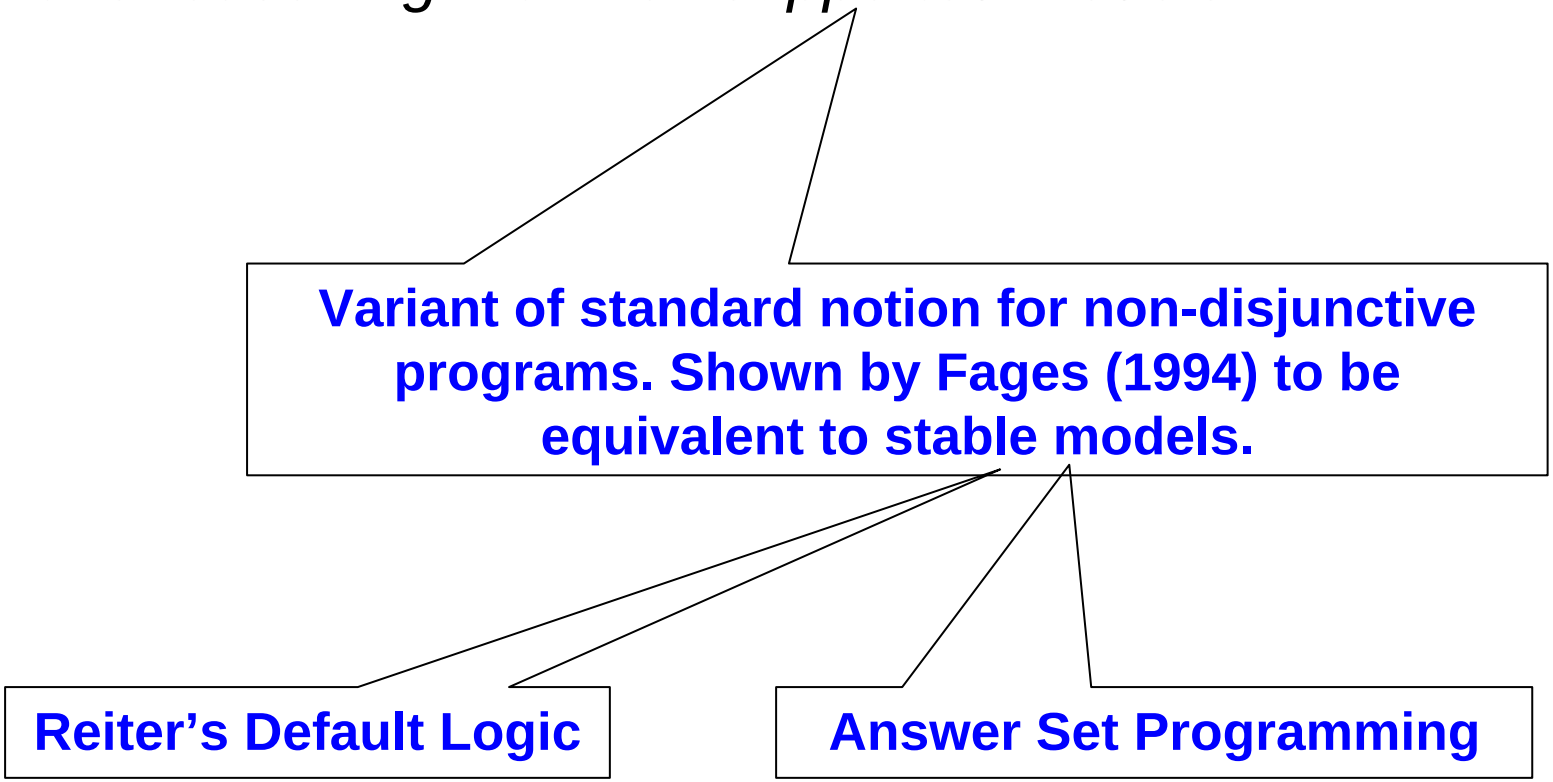
Presented here!

Can be done offline.

Semantic description

Inference boils down to

brave reasoning with well-supported models.



Screech Performance

Galen ontology
673 axioms, 175 classes
randomly populated with 500 individuals

267 disjunctions in 133 rules eliminated

Time (DD)	Time (SPLIT)	Instances	Class Name
11036 ms	6489 ms	154/154	Biological_object
11026 ms	5959 ms	9/9	Specified_set
11006 ms	6219 ms	9/13	Multiple
11015 ms	5898 ms	16/16	Probe_structural_part_of_heart
11036 ms	7711 ms	4/4	Human_red_blood_cell_mature
11055 ms	5949 ms	24/58	Biological_object_that...

Table 2. Performance comparison for instance retrieval using disjunctive datalog (DD) vs. the corresponding split program (SPLIT), on the KAON2 datalog engine. *Instances* indicates the number of instances retrieved using DD versus SPLIT, e.g. class *Multiple* contained 9 individuals, while the split program allowed to retrieve 13 (i.e. the 9 correct individuals plus 4 incorrect ones). The full name of the class in the last row is Biological_object_that_has_left_right_symmetry.

Thanks!

http://logic.aifb.uni-karlsruhe.de/screech