ParisMatch
Ontology Matching at Scale

Antoine Amarilli   Pierre Senellart   Fabian M. Suchanek

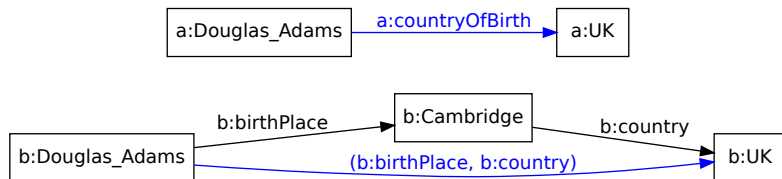Télécom ParisTech

Performance

- Millions of entities, tens of millions of facts.
- In-memory, multithreaded computation.

BerkeleyDB, SSD	RAM, 1 thread	RAM, 4 threads
20h30	1h15	26m

Table : Running times for the DBpedia–YAGO alignment task.
Intel Core i7-3820 CPU clocked at 3.60 Ghz, 48 GB of RAM.

Join Relations

- Simplest possible difference in structure between ontologies.
- Relations of one ontology correspond to join relations in the other ontology.
- Similar to the "join" operator of relational calculus.

Literal Similarity Functions

| Edgar R. Burroughs | Douglas Adams | and Constance Garnett |
| Edgar Rice Burroughs | Adams, Douglas | Constance Garnett |

- The original PARIS uses an exact literal equality function.
- Ad-hoc refinements: adjust for case, strip special characters...
- More general : index the literals using shingling :
    - hash the shingle sets through random hash functions
    - keep the minimum values (MinHashing)
- Use the index to compute pairs of similar literals efficiently!