NCSA SPIN FELLOW

Sahil Gupta
Rising Junior, Computer Science

Mentored by:
Prof. Bertram Ludäscher
Director, Center for Informatics
Research in Science and Scholarship

Project Available at:
https://github.com/idaks/**PW-explorer**

Reference for Taxonomy Alignment Example:
- Raunich, Salvatore, and Erhard Rahm. **Target-driven merging of taxonomies with ATOM**. *Information Systems* 42 (2014): 1-14.

An Extensible Possible Worlds Explorer for Answer Set Programming

Project Objectives

- To analyze the solutions (Possible Worlds) produced by Answer Set Programming (ASP) reasoners such as Clingo and DLV.
- To develop ways to query, visualize, and interpret this data using relational databases such as SQLite Databases and Panda DataFrames.
- To cluster similar and equivalent PWs to aid in finding fundamental solutions
- To identify and analyze the 'simplest' solutions to an underspecified problem, based on a user-defined distance function between PWs. (Occam's Razor: among competing explanations, the simplest one should be selected)
- To conduct a case study, demonstrating the feasibility of the approach.

Technical Approach

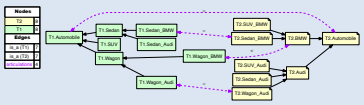
- Parse the output of Cllingo ASP using the Antlr parsing tool.
- Populate a relational database using the parsed data, with one table for each identified relation.
- Implement functionality to execute powerful SQL and Python/Panda queries on these databases/dataframes.
- Use these queries to discover interesting features about these Possible Worlds.
- Clustering of PWs with user-defined distance metric (e.g., size of symmetric distance).
- Analyze 'complexity' of the various solutions.
- Develop a stand-alone general tool that can be used by other tools like EulerX to then analyze more specific problems, using custom definitions of distance, complexity and visualization techniques.

Status

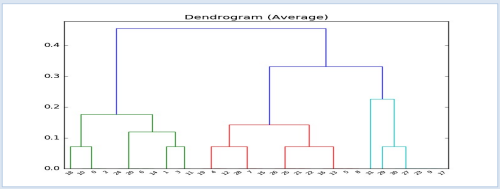
- Implemented Clingo Parser using Antlr which directly populates the parsed data into Panda DataFrames and SQLite Database.
- Implemented queries such as intersection, union, frequency of specific tuples, instances of a relation in a solution, difference, symmetric difference, identify redundant attributes in relations, identify unique tuples , etc.
- Implemented a general way of calculating dissimilarity/distance between two Possible Worlds using size of symmetric set difference as the distance metric.
- Developed an interactive command-line UI to query these Possible Worlds.
- Implemented a basic complexity analysis tool based on size of the solution.
- Implemented clustering of the Possible Worlds based on a distance-matrix generated using the aforementioned distance metric.
- Developed visualizations of these Possible Worlds such as Dendrograms and 2-D projection of these possible worlds.

Automobile Taxonomy Alignment Example:

Input:



Hierarchical Clustering (Dendrogram):

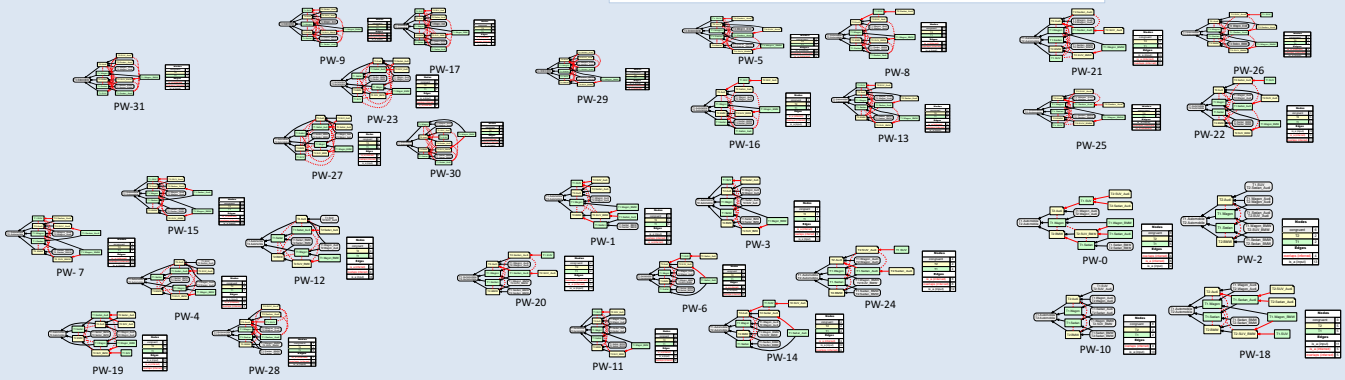


Complexity Definition: Number of ***overlap relations*** (><) in the PW:

```
select count(*) from rel_3
where x3 = "><"
  and pw = pw_id

pw_id ∈ [0,31]
```

Clusters Identified:



PWs: [31, 9, 17, 23, 27, 30, 29, 5, 8, 13, 16, 21, 22, 25, 26, 7, 15, 28, 4, 12, 19, 3, 11, 1, 6, 14, 20, 24, 18, 0, 2,10]
Complexities: [1.0, 0.79, 0.79, 0.79, 0.79, 0.79, 0.71, 0.57, 0.57, 0.57, 0.57, 0.5, 0.5, 0.5, 0.5, 0.43, 0.43, 0.43, 0.36, 0.36, 0.36, 0.29, 0.29, 0.21, 0.14, 0.14, 0.14, 0.14, 0.07, 0, 0, 0]

Decreasing Level of Complexity based on number of overlaps