From Web Data to Visualization via Ontology Mapping

Owen Gilson

Swansea University, UK

December 2008

Outline:

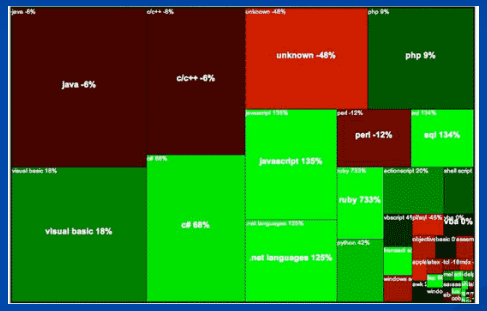
- What is Information Visualization?
- Motivation
- Related Work
- SemViz Pipeline
- Ontologies with Certainty Factors (OCF)
  - Domain Ontology (DO)
  - Visual Representation Ontology (VRO)
  - Semantic Bridge Ontology (SBO)
- Ontology Mapping Algorithm
- Results
- Summary

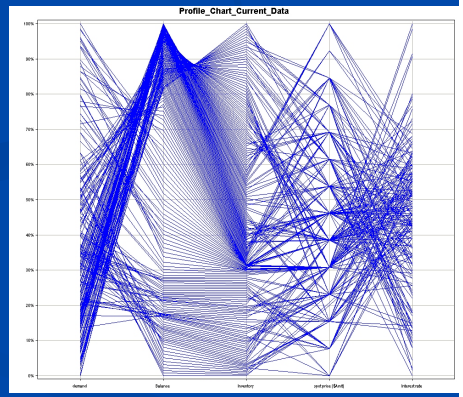
What is Information Visualization?

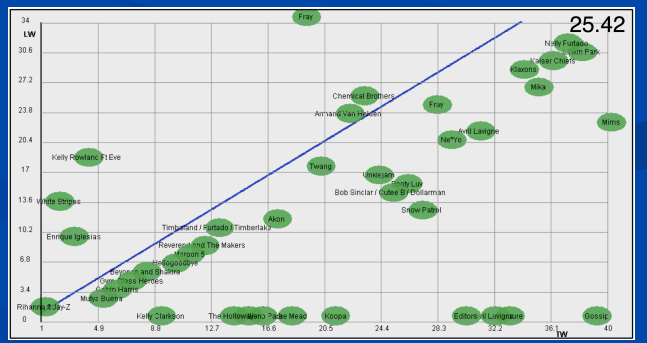
- "The use of computer-supported, interactive, visual representations of *abstract data* to amplify *cognition*"

[Card et al., 1999]

- Examples
  - Tree Map
  - Parallel Coordinates
  - 2D graph
  - Graph Network





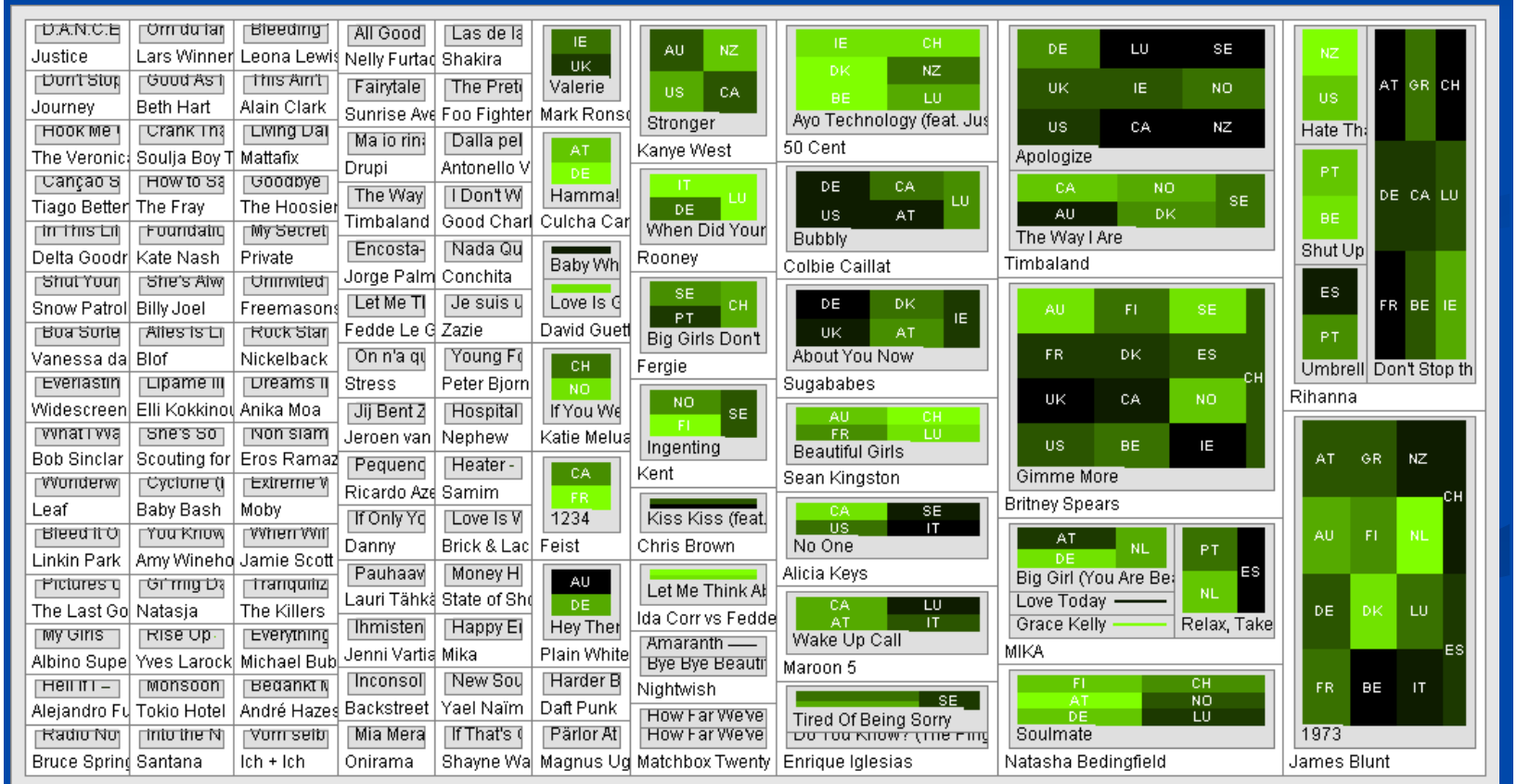


From Tabular Web Data to TreeMap

4 fields: 22 countries… ∴ 220 records…

- Country Name
- Artist Name
- Song Name
- Chart Position

ILOG Discovery [Baudel 2002]

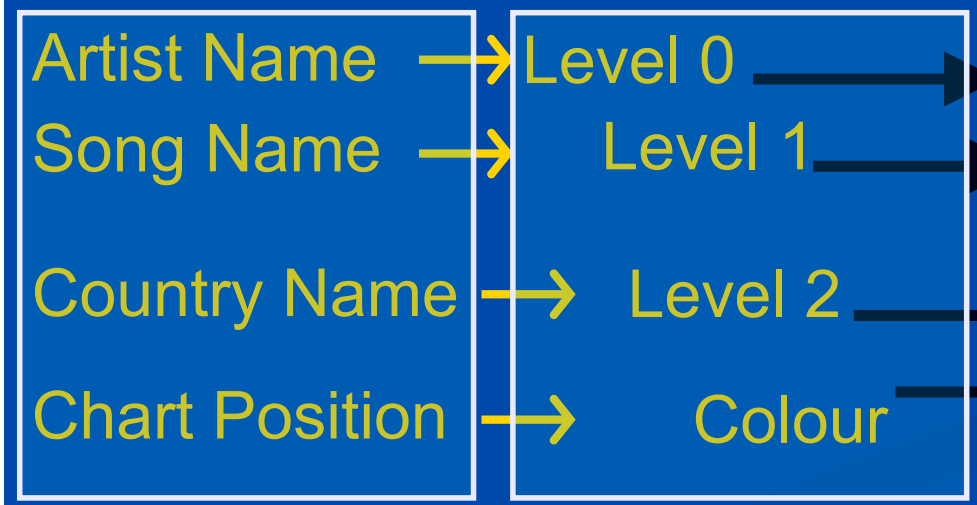


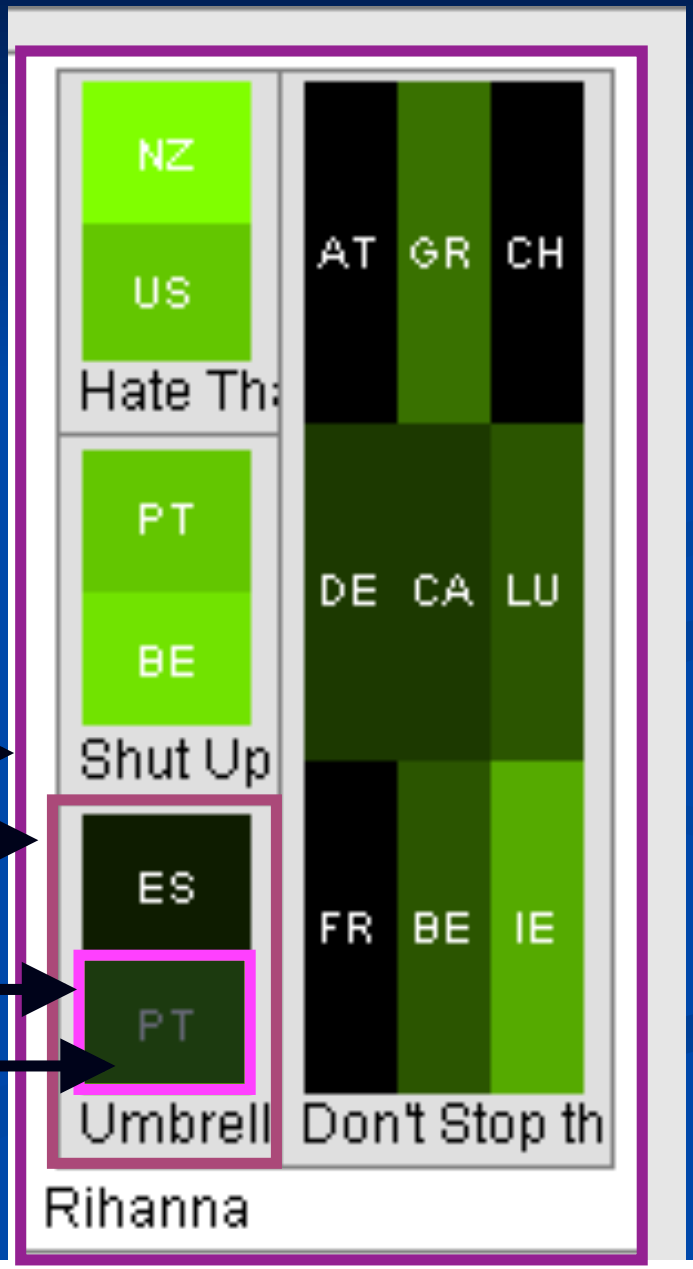
Visualization As Mapping

4 fields:

- Country Name
- Artist Name
- Song Name
- Chart Position

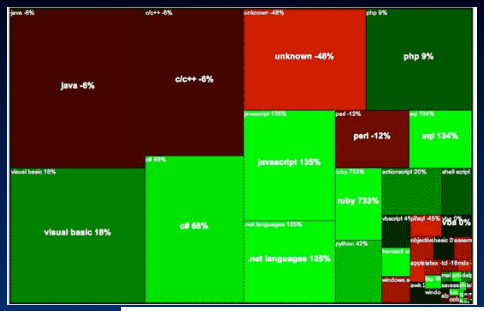
It's a Mapping Process…

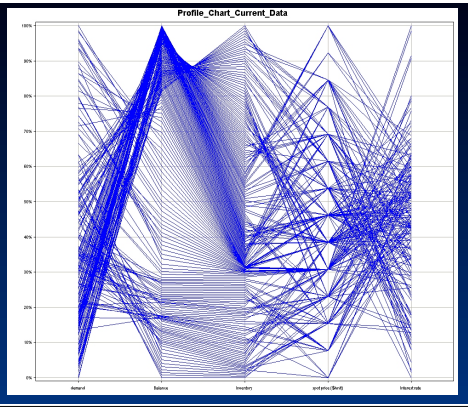


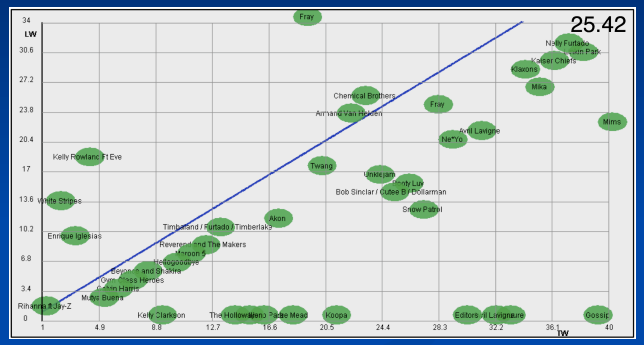


Motivation:

Tree Map
Parallel Coordinates
2D plot
Graph Network
…





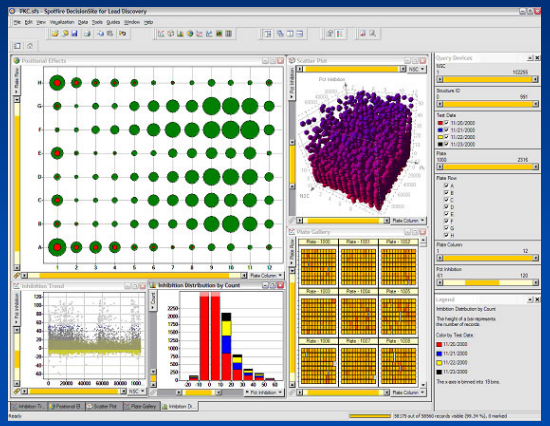


"Visualization for Everyone"

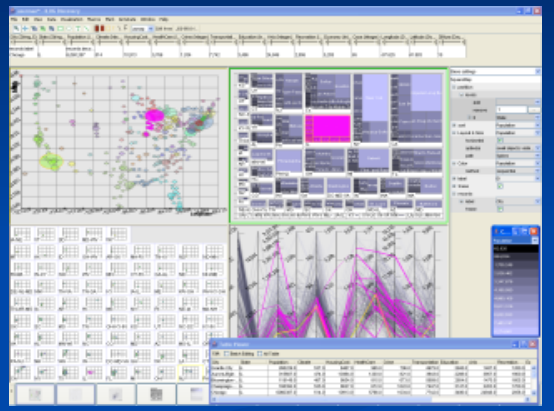
Related Work:

Trial and Error

Spot Fire
[Shneiderman 1999]

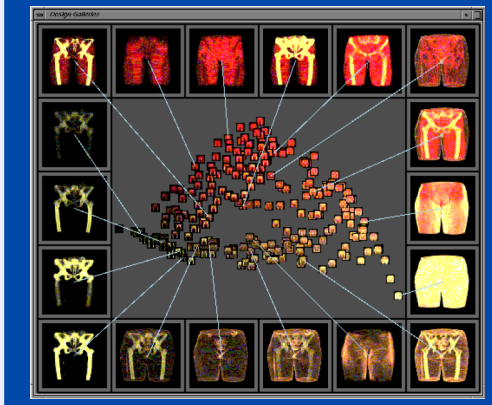


ILOG Discovery
[Baudel 2002]



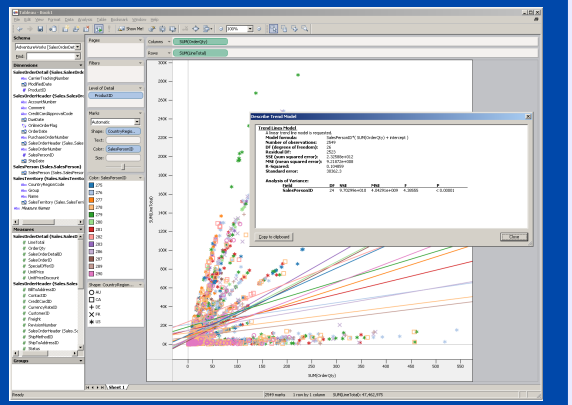
Design Galleries

[Marks 1997]



Automatic Visualization

Tableau (Show Me)
[Mackinlay 2007]



Problems:

1. How do we get the Web Data?

Extract data. Use a Screen Scraper.

```
<tr>
    <td>Artist</td>
    <td>Song</td>
    ...
</tr>
```

2. What is the source data meaning?

Capture *Domain Semantics*. Use an Ontology.

3. What can a visualization style convey?

Capture *Visual Representation Semantics*. Use an Ontology.

4. How do you present "source data meaning" using "visualization meaning"?

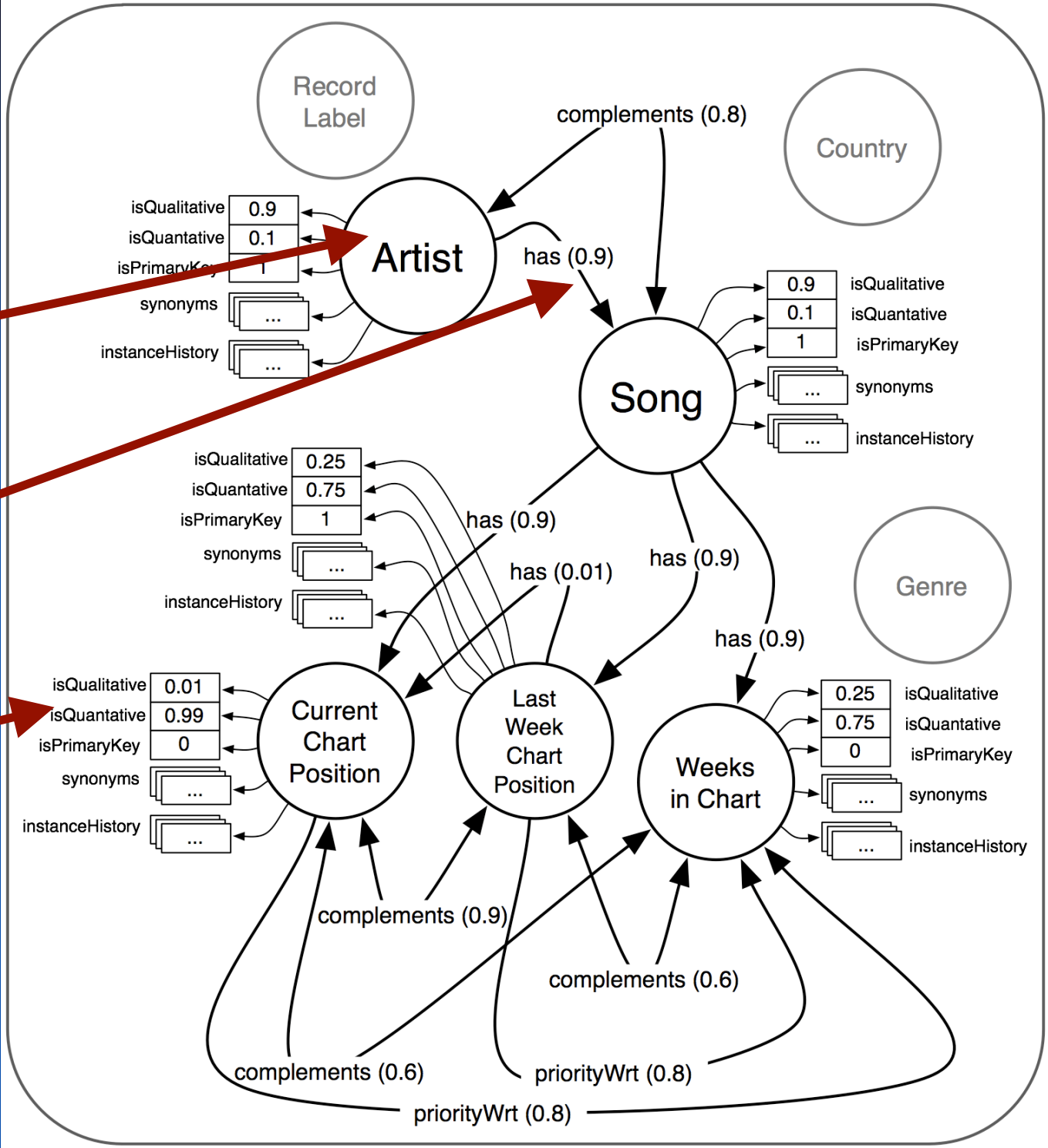
Map between *Source Data Entities* and *Target Artefacts*.
Use Ontology Mapping.

Domain Ontology (DO) : Music

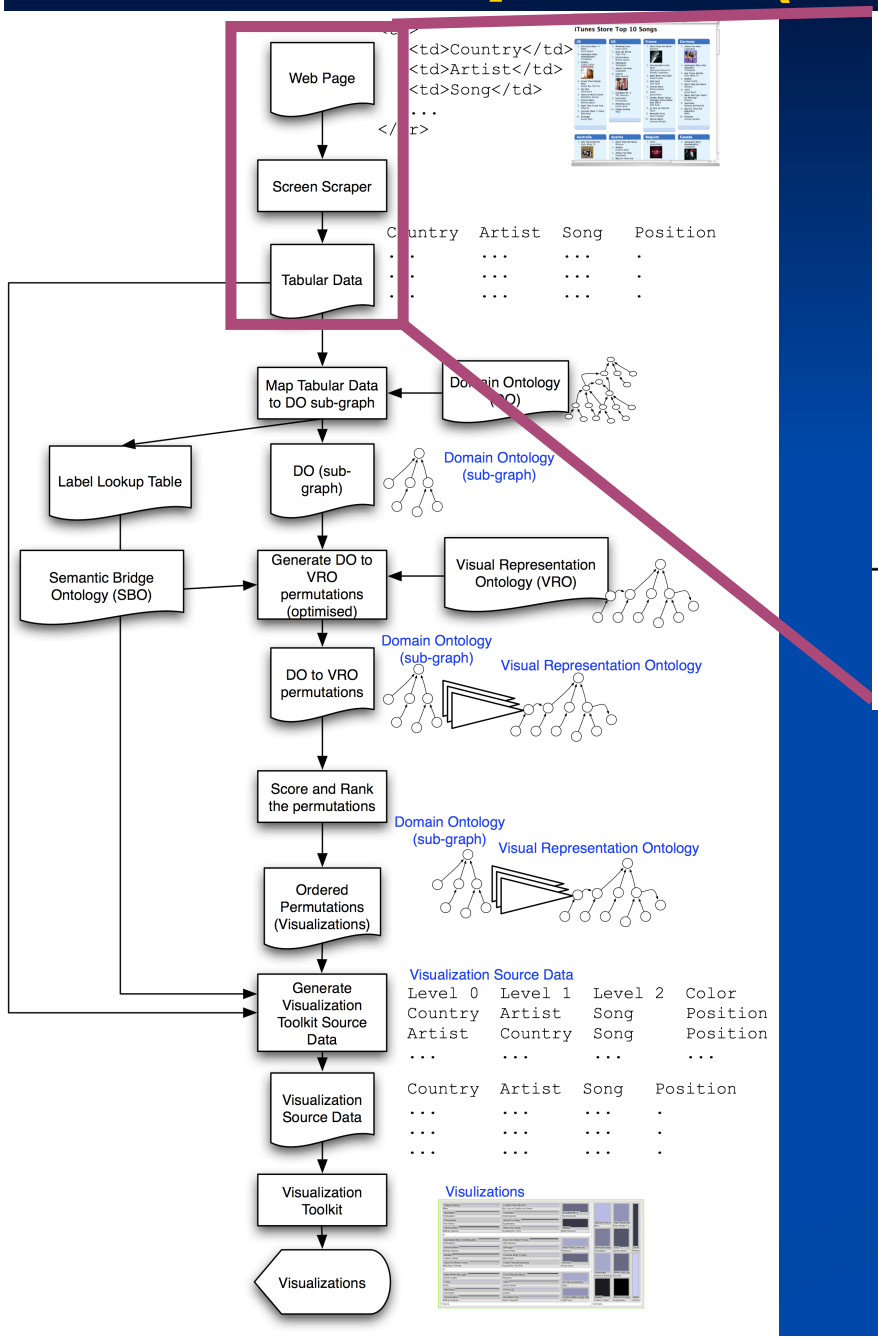
Concept

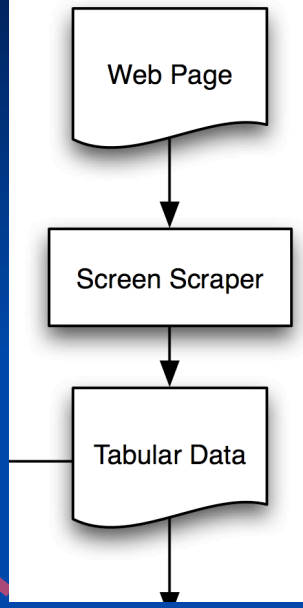
Weighted Relation (0 → 1)

Attribute



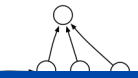
SemViz Pipeline (Web Page to Tabular Data)



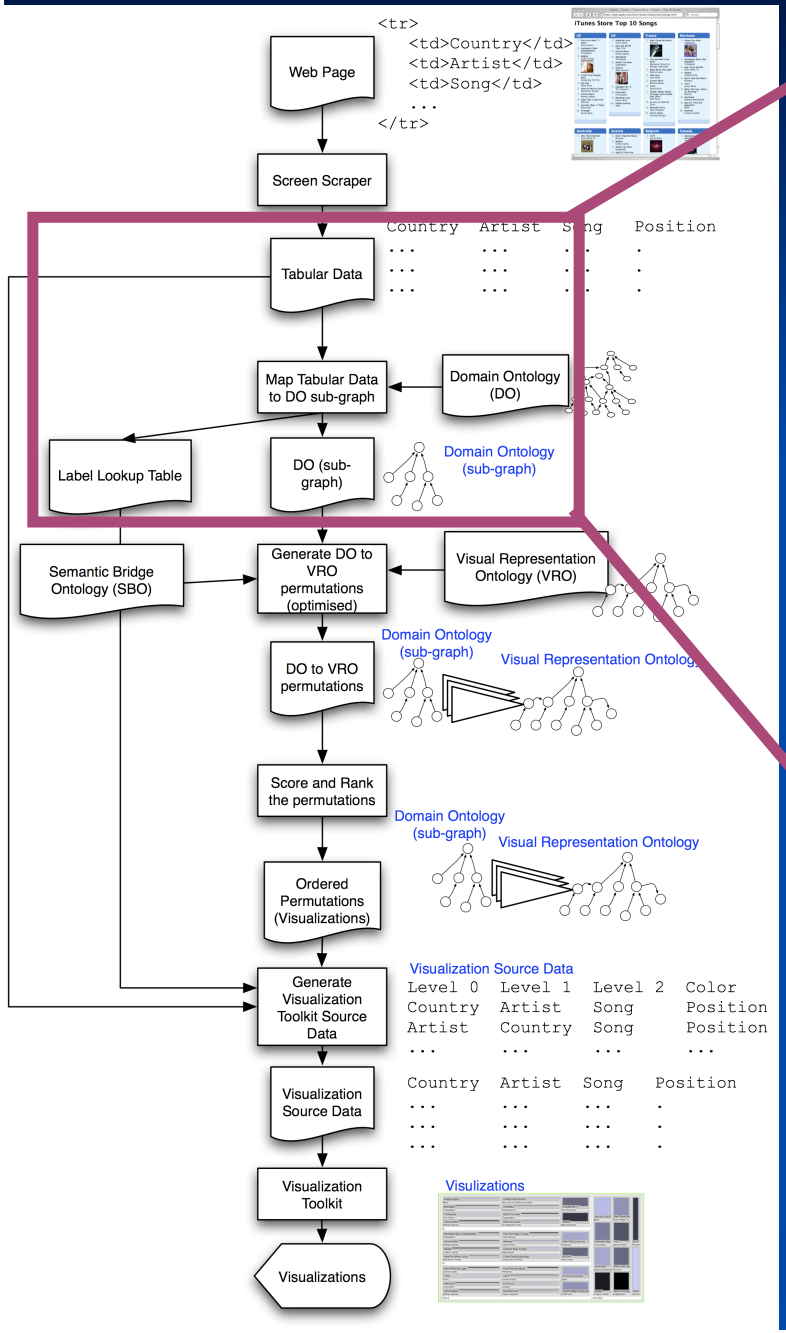


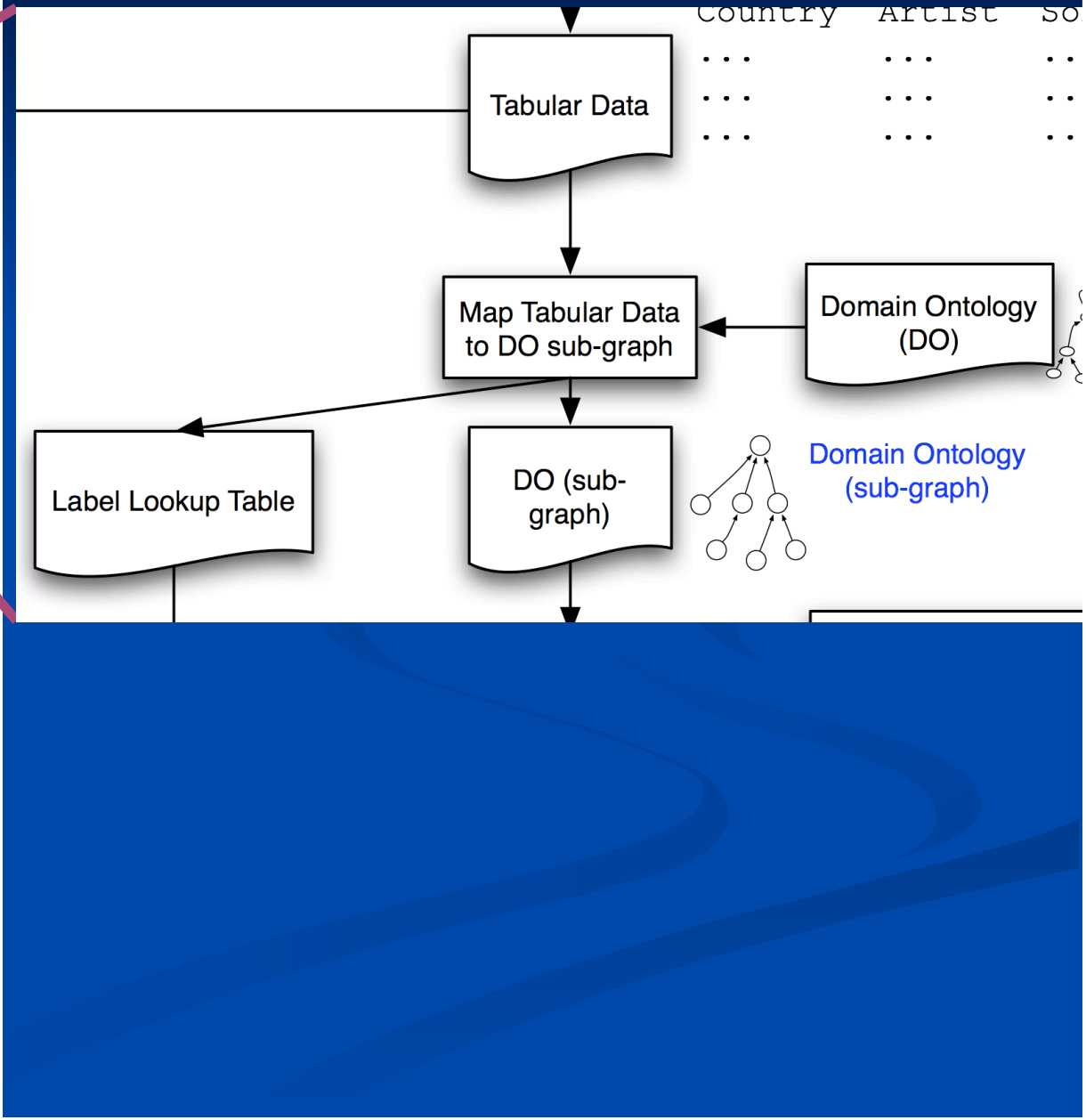

```
<tr>
   <td>Artist</td>
   <td>Song</td>
   ...
</tr>
```

Artist	Song	TW	LW
...	...	.	.
...	...	.	.
...	...	.	.

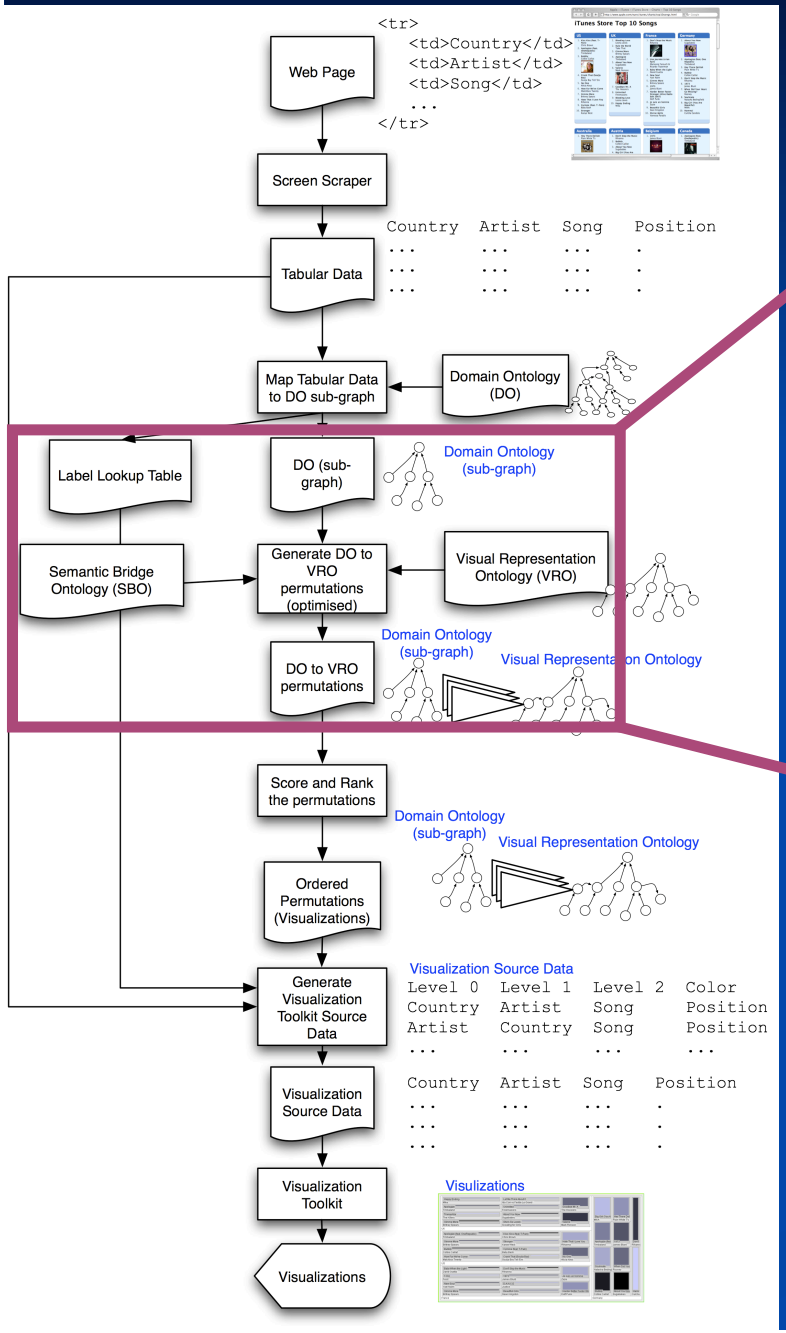


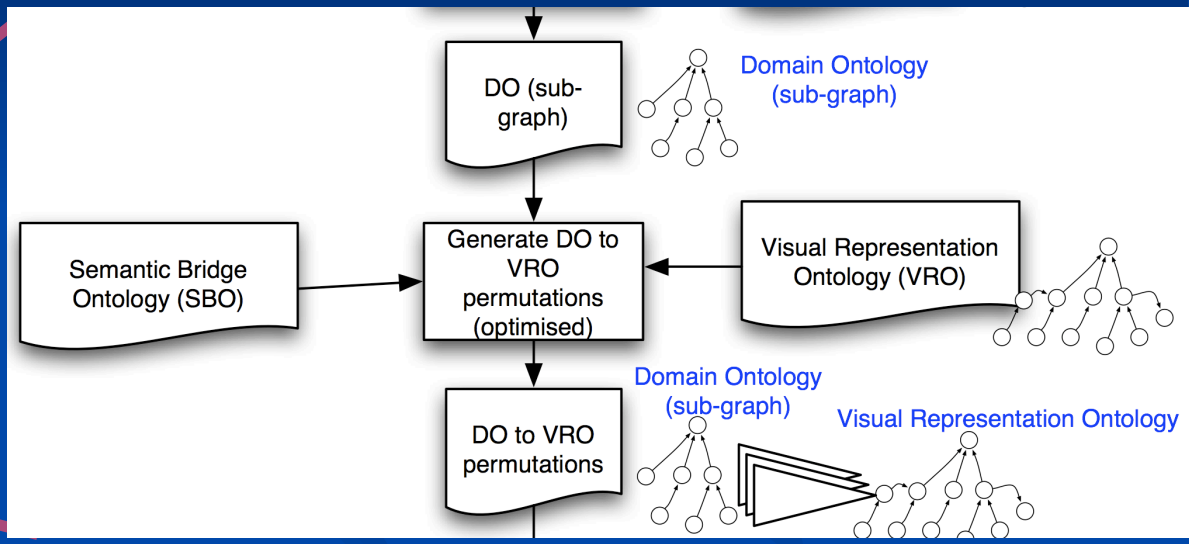
SemViz Pipeline (Tabular Data to DO sub-graph)



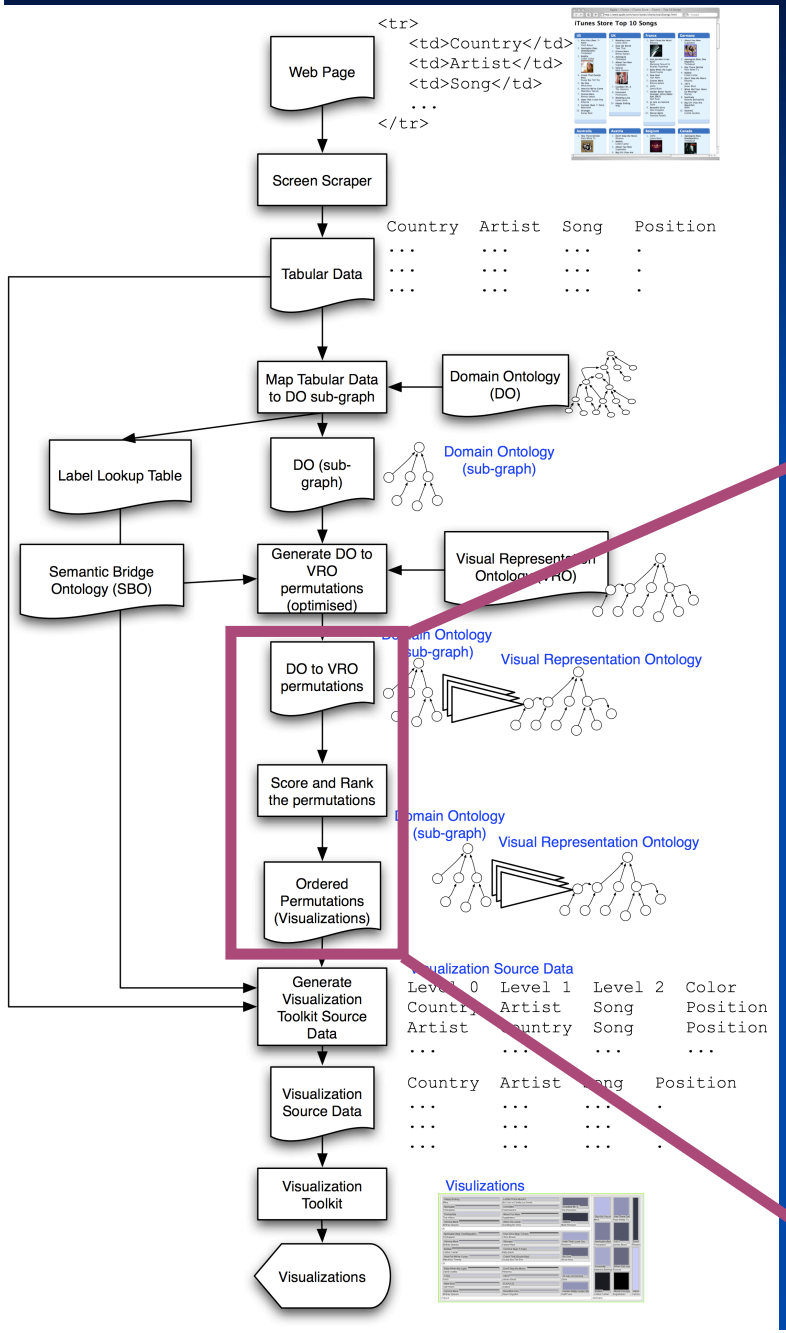


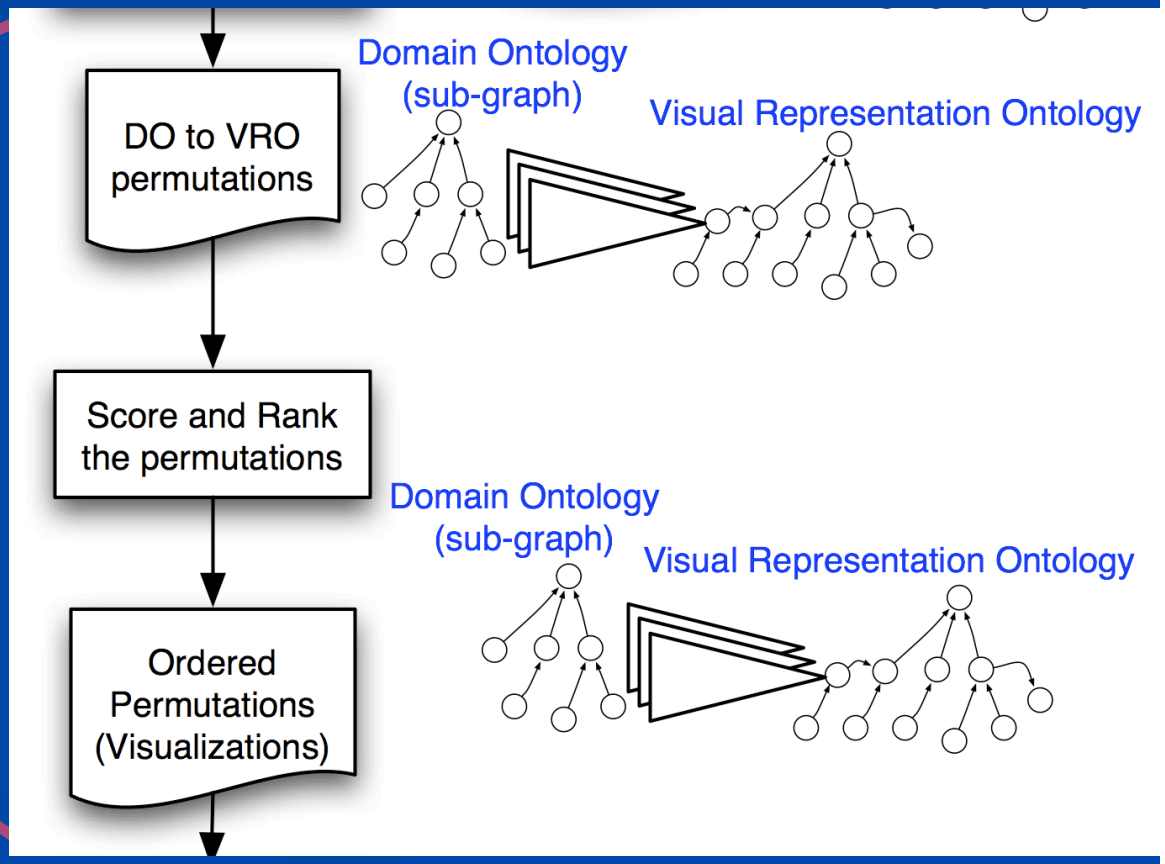
SemViz Pipeline (DO sub-graph to VRO)



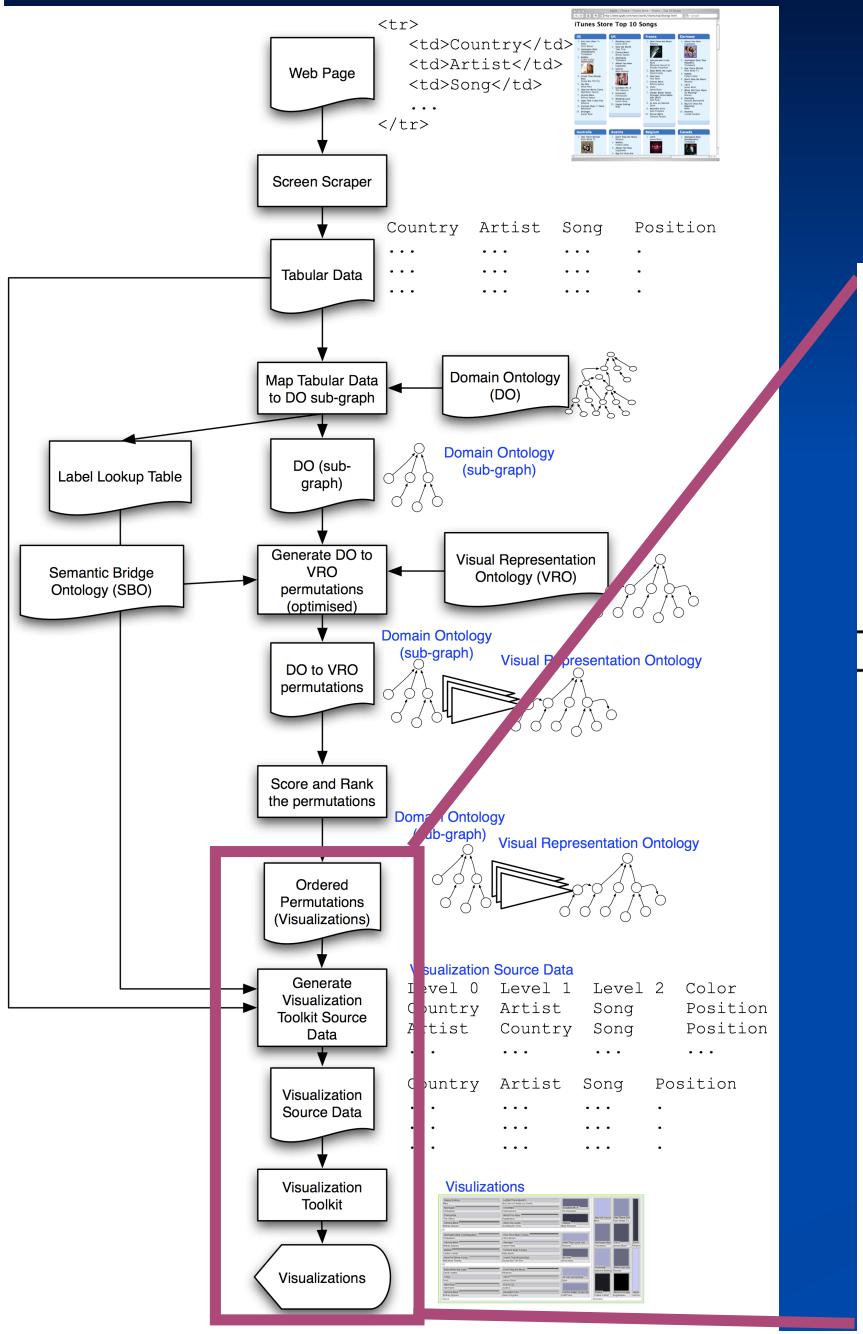


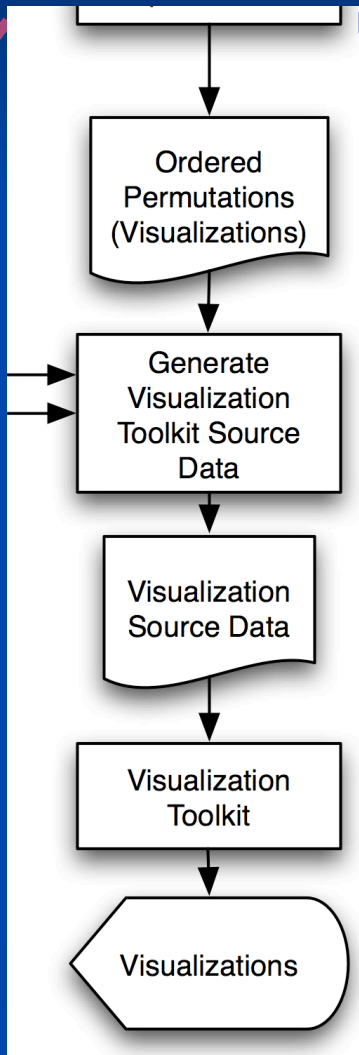
SemViz Pipeline (Score DO to VRO)





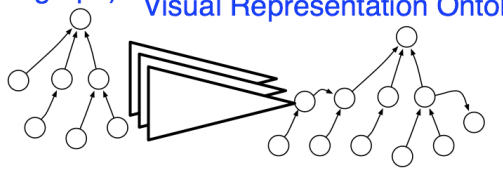
SemViz Pipeline (Generate Visualizations)





Domain Ontology (sub-graph)

Visual Representation Ontology

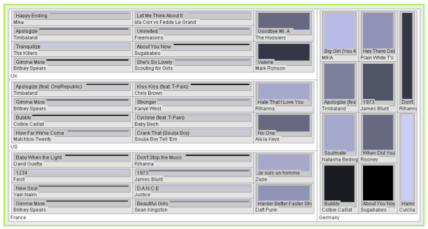


Visualization Source Data

Level 0	Level 1	Level 2	Color
Country	Artist	Song	Position
Artist	Country	Song	Position
...	...	...	...

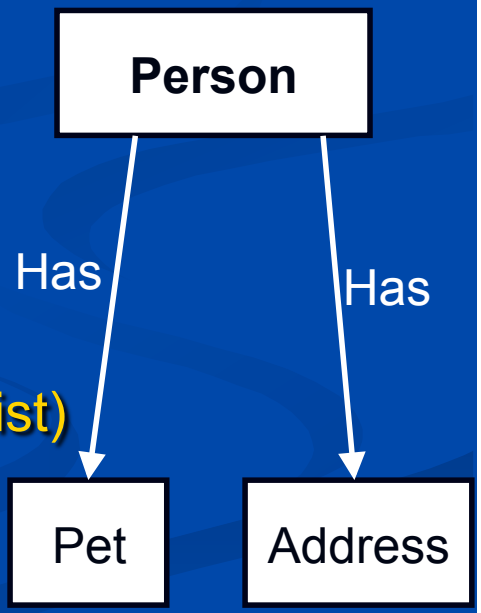
Country	Artist	Song	Position
...	...	...	.
...	...	...	.
...	...	...	.

Visulizations



“Traditional” Ontologies

- Graph Representation of Domain Knowledge
- Syntactically and Semantically rich
- Hierarchy of concepts and properties
- (e.g. Music Ontology Specification (MOS))

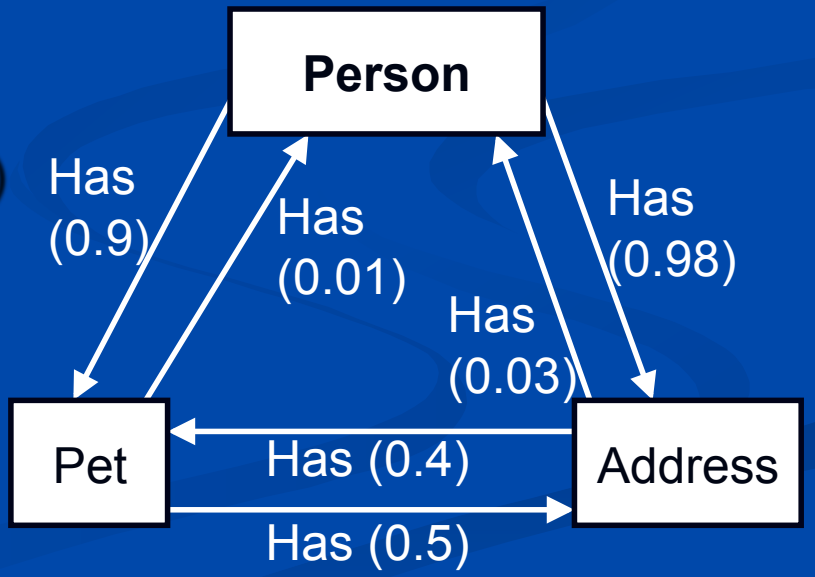


- Absolute relations (it exists, or does not exist)
- RDF / OWL

Ontologies with Certainty Factors (OCF)

- (Still a) Graph Representation of Domain Knowledge
- (Can be) Syntactically and Semantically rich
- (Can have a) Hierarchy of concepts and properties

- Weighted relations (0.0 → 1.0)
- Fully connected graph



- RDF (with weighted properties through Reification)

Traditional vs Ontologies with Certainty Factors

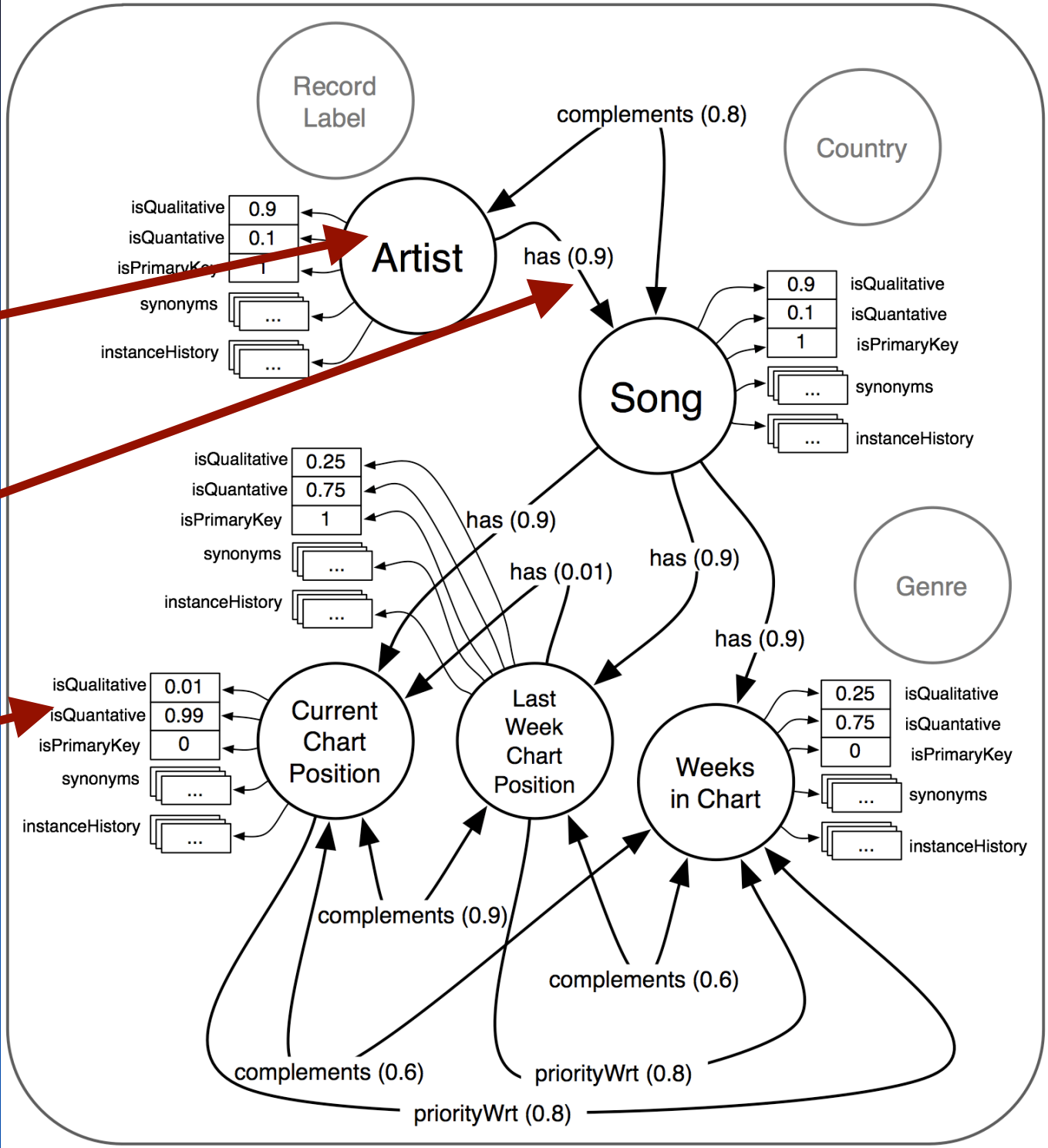
	Traditional (RDF/OWL)	OCF (RDF)
Explicit Conceptualisation	Yes	Can be
Syntactically and Semantically rich	Yes	Can be
Hierarchy of concepts	Yes	Can be
Hierarchy of properties	Yes	Can be
Relations	Absolute	Weighted (0.0 → 1.0)
Technology	RDF / OWL	RDF
Connectedness	Partial	Fully

Domain Ontology (DO) : Music

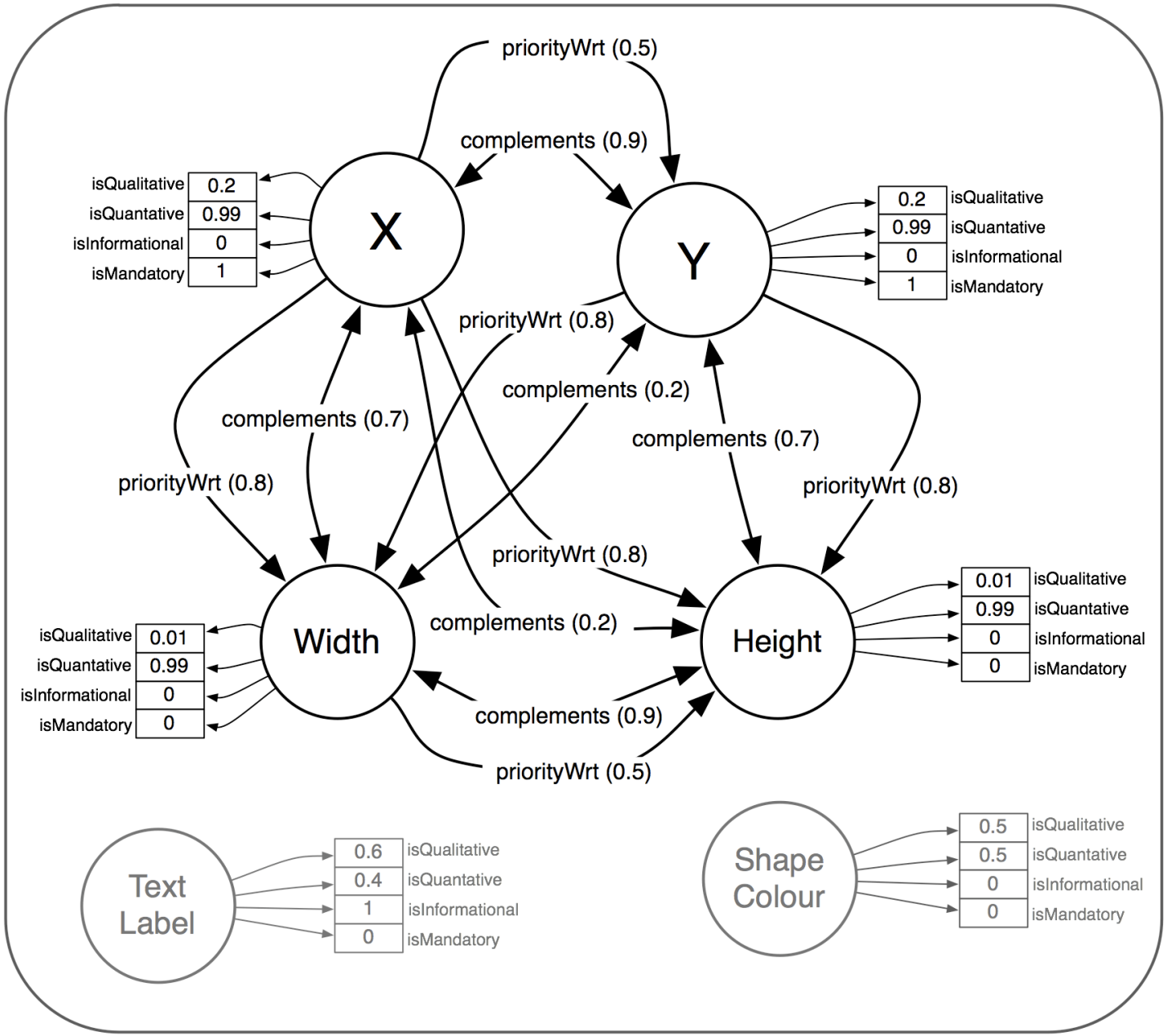
Concept

Weighted Relation (0 → 1)

Attribute

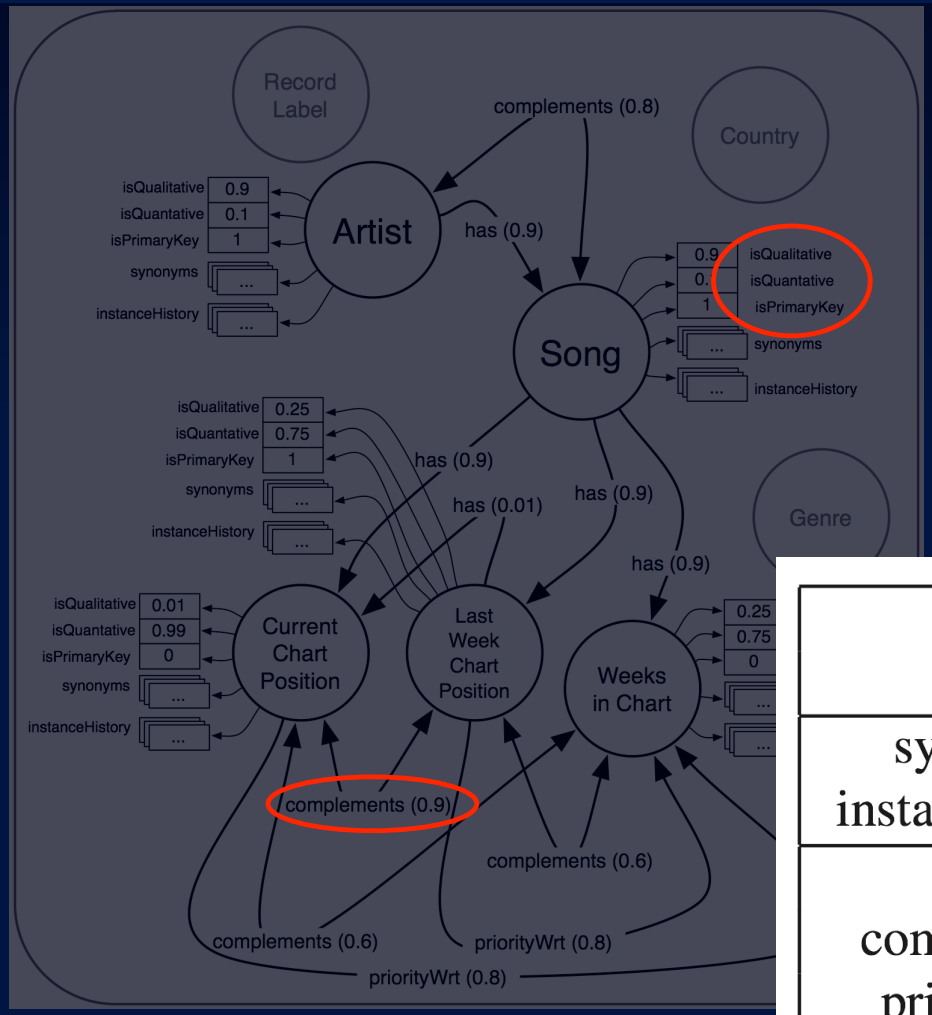


Visual Representation Ontology (VRO) : 2D plot

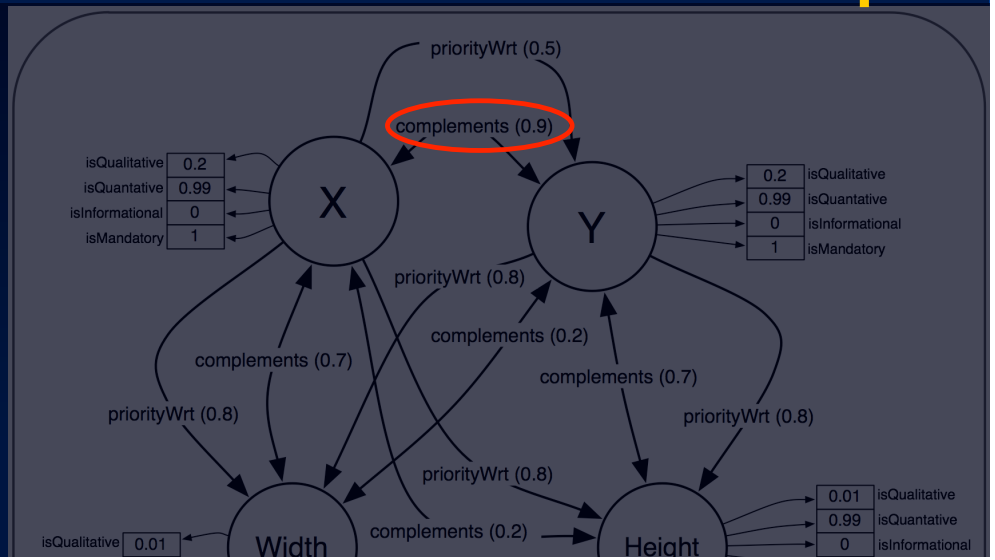


Semantic Equivalence

DO : Music



VRO : 2D plot



DO	VRO	Relationship / Attribute
synonyms	-	A
instanceHistory	-	A
has	contains	R
complements	complements	R
priorityWrt	priorityWrt	R
isQualitative	isQualitative	A
isQuantitative	isQuantitative	A
isPrimaryKey	isInformational	A
-	isMandatory	A

Mapping Permutations

iTunes:
- Country Name
- Artist Name
- Song Name
- Chart Position

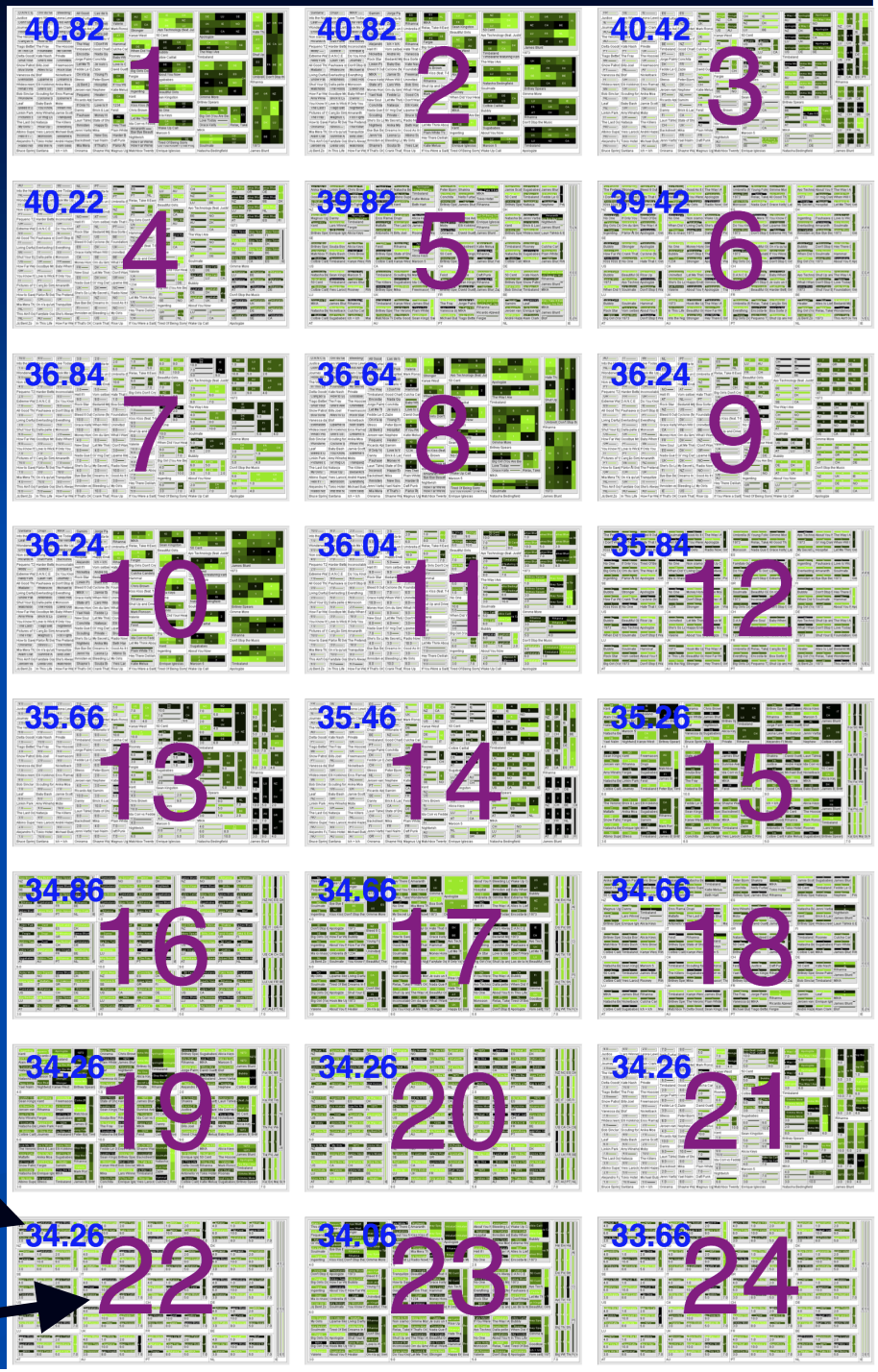
TreeMap:
- Level 0
- Level 1
- Level 2
- Colour

Assuming all fields are mapped only once, there are:

24 permutations

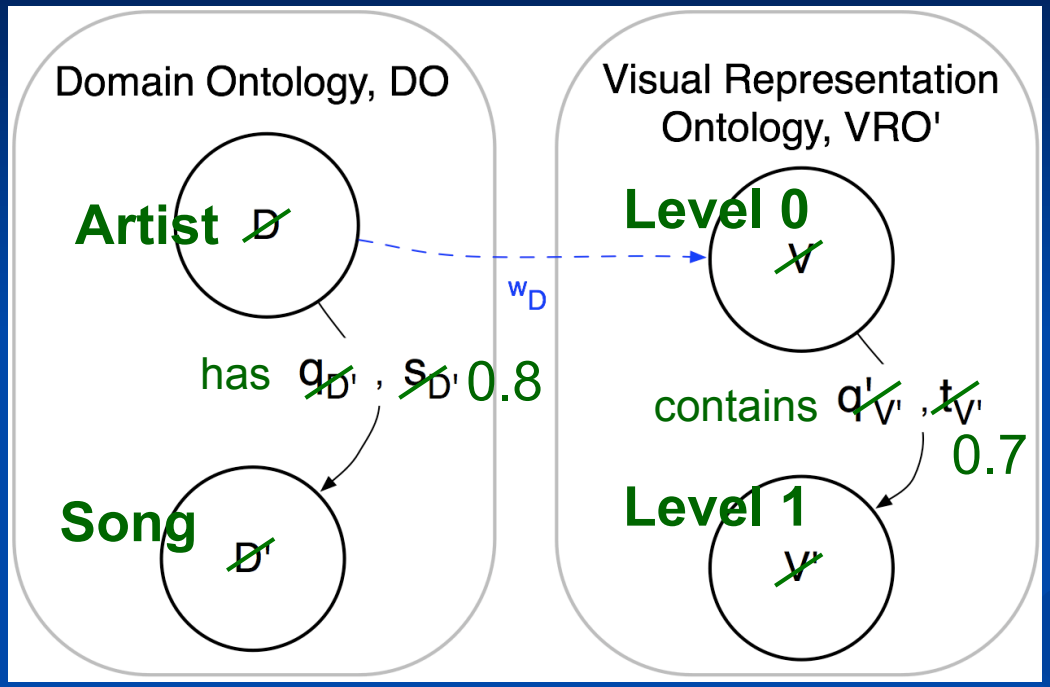
Which visualization is best?

Calculate "Cognitive Value"



Scoring Algorithm

- Based on "Ontology Mapping With Uncertainty" [Mitra et al, 2005]



Let Θ be the mapping from DO to VRO

So: $V = \Theta(D)$
$V' = \Theta(D')$

w_D is the weighting of the concept pair DV.

If: $q_{D'} \sim q'_{V'}$
$0 \le s_{D'} \le 1$
$0 \le q'_{V'} \le 1$

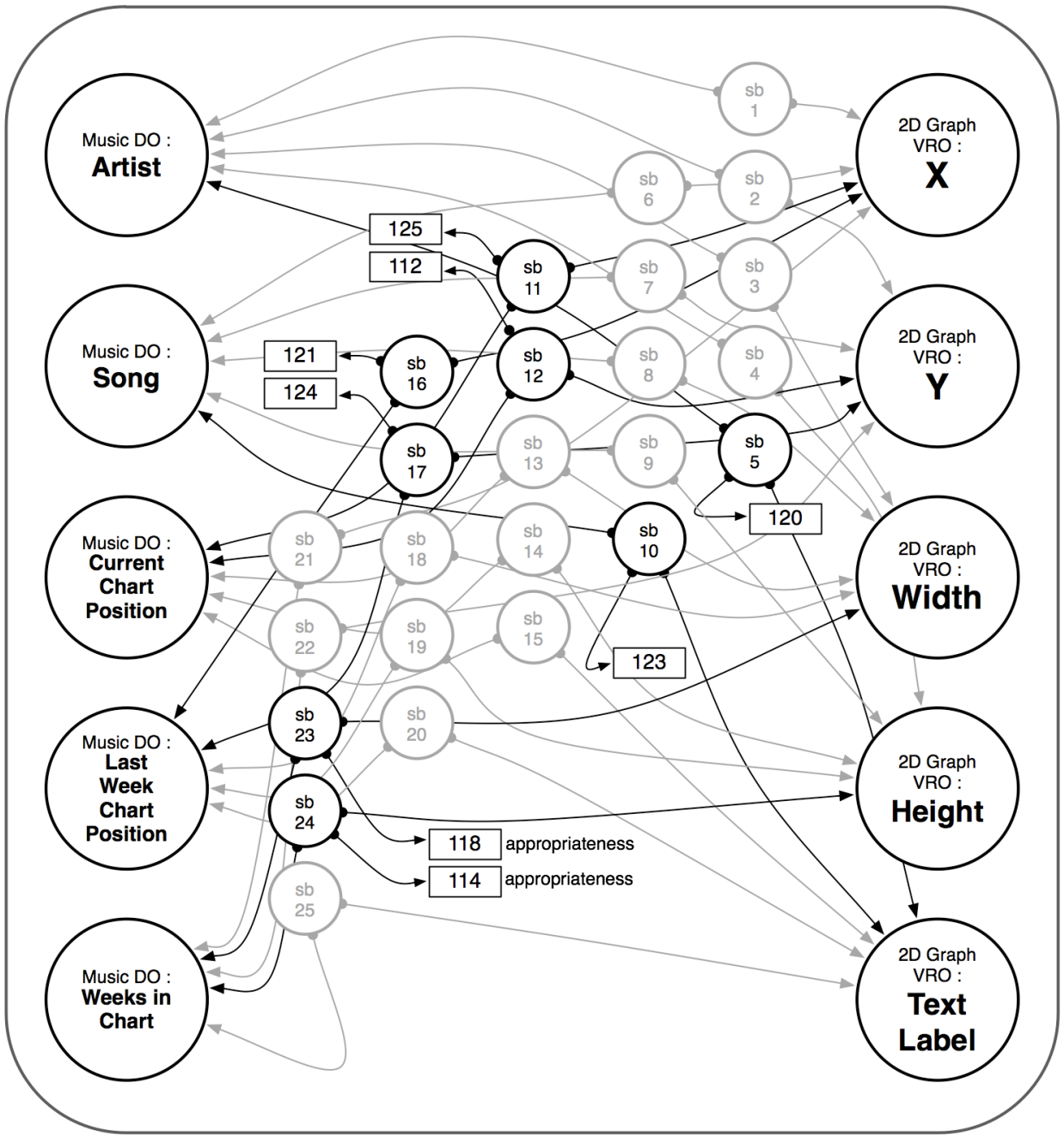
$$w_D = \sum_{D' \neq D} f_1(s_{D'}, t_{V'})$$

$$totalw_\theta = \sum_{D \in DO} w_D$$

$$f_1(s,t) := 1 - |s - t|$$

$w_D = 1 - | s_{D'} - t_{V'} |$
$= 1 - | 0.8 - 0.7 |$
$= 0.9$

Semantic Bridge Ontology (SBO)

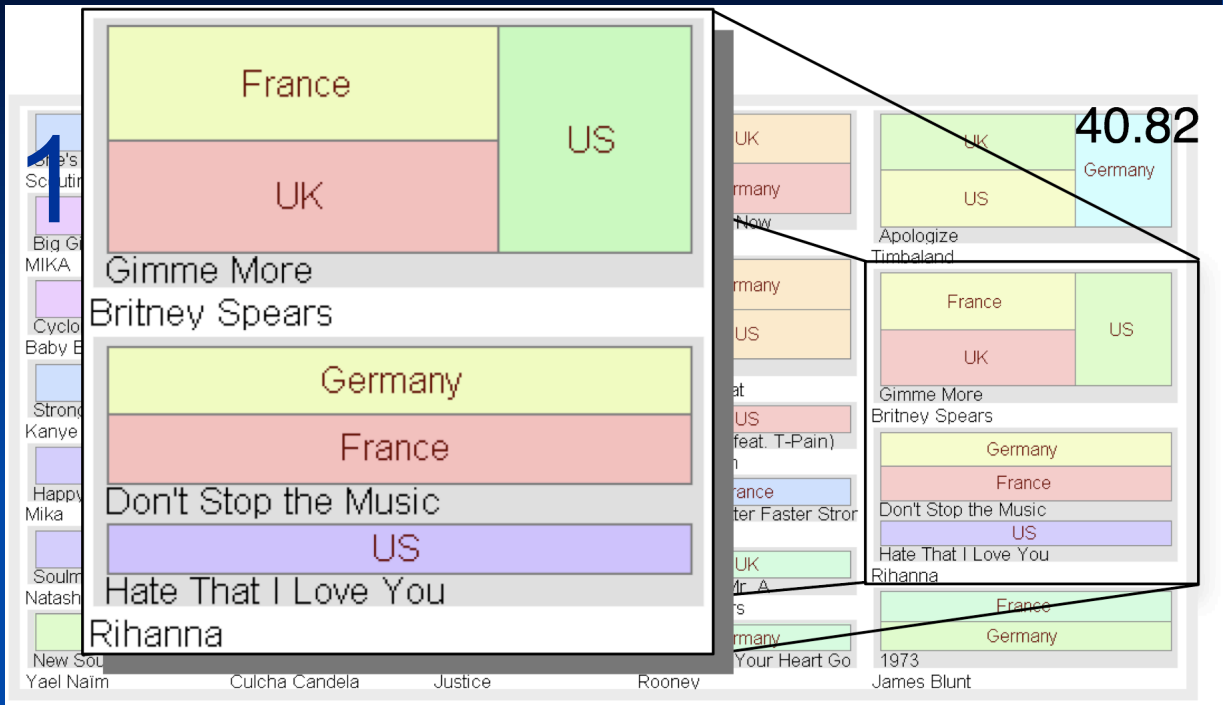


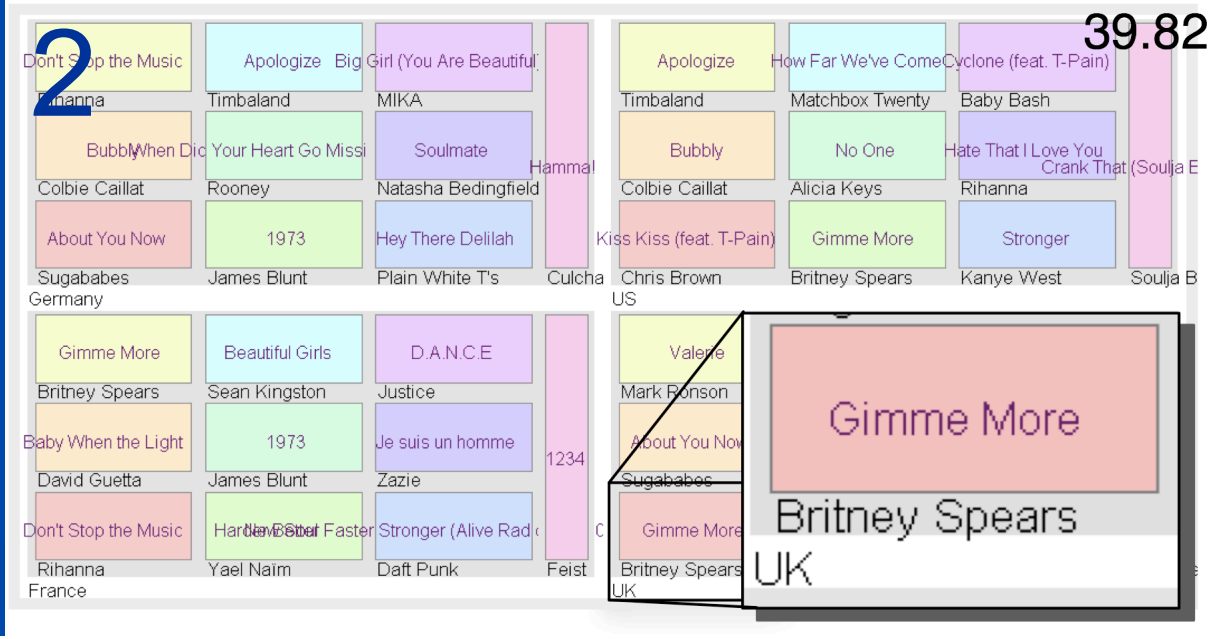
- Stores *available* expert knowledge
- Records all possible mappings between all DO's and all VRO's
- Fully-connected graph
- Each mapping has an *appropriateness* value
- Default is 100, increased or decreased as necessary.
- 2 purposes:
  - Reduces complexity.
  - Increases accuracy.

iTunes Country Chart to TreeMap

Source (iTunes)	→	Target (TreeMap)
Country		Colour
Artist Name		Level 0
Song Name		Level 1
Chart Position		Level 2

# of mapping permutations: 24
of *visible* map'g permutations: 24



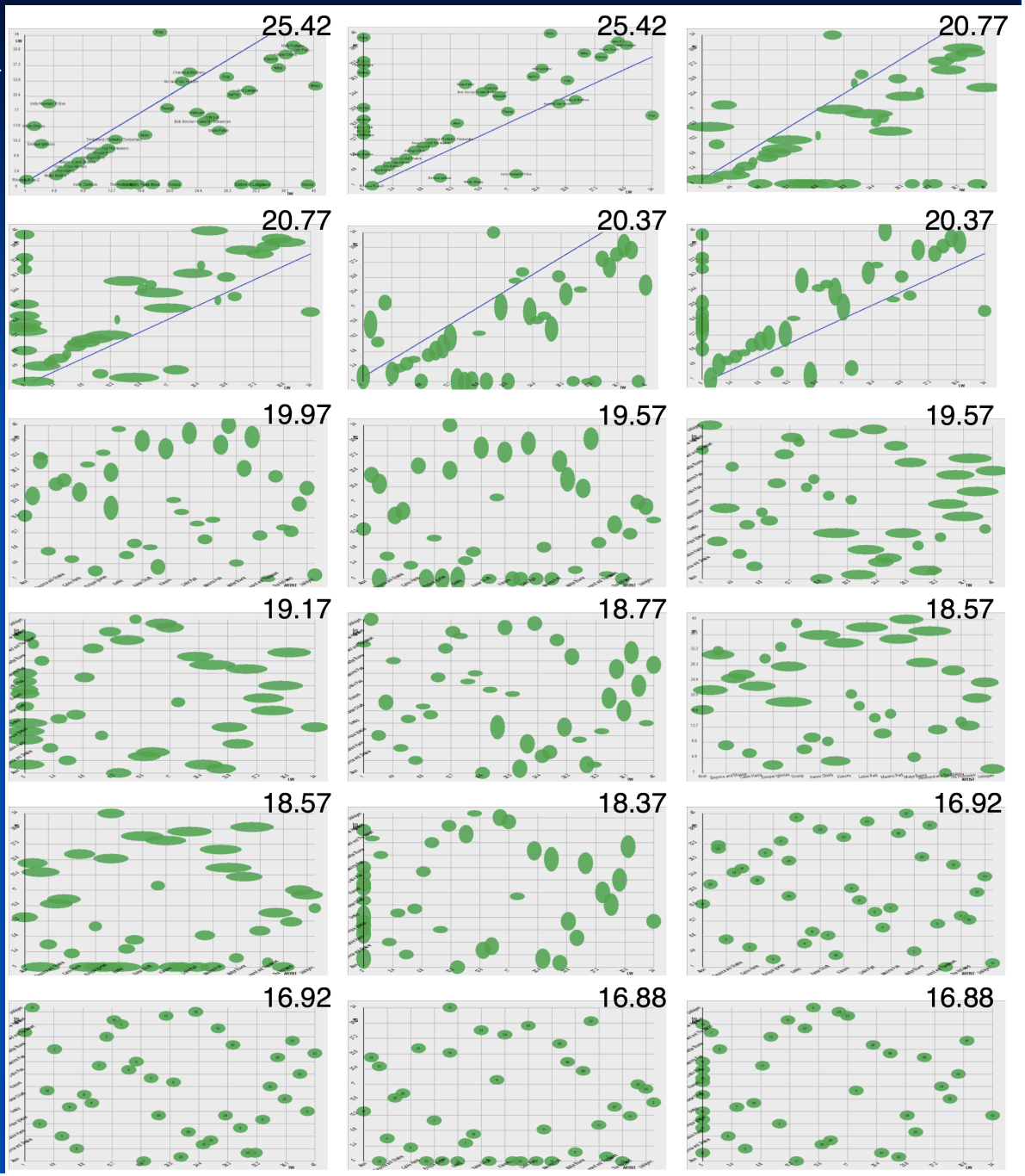


BBC Top 40 to 2D plot

Source (BBC)	Target (2D plot)
This Week	X position
Last Week	Y position
Artist Name	Width
	Height
	Text Label

# of mapping permutations: 60

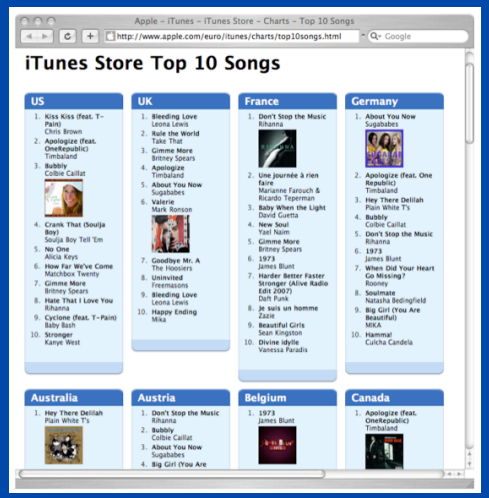
of *visible* map'g permutations: 18

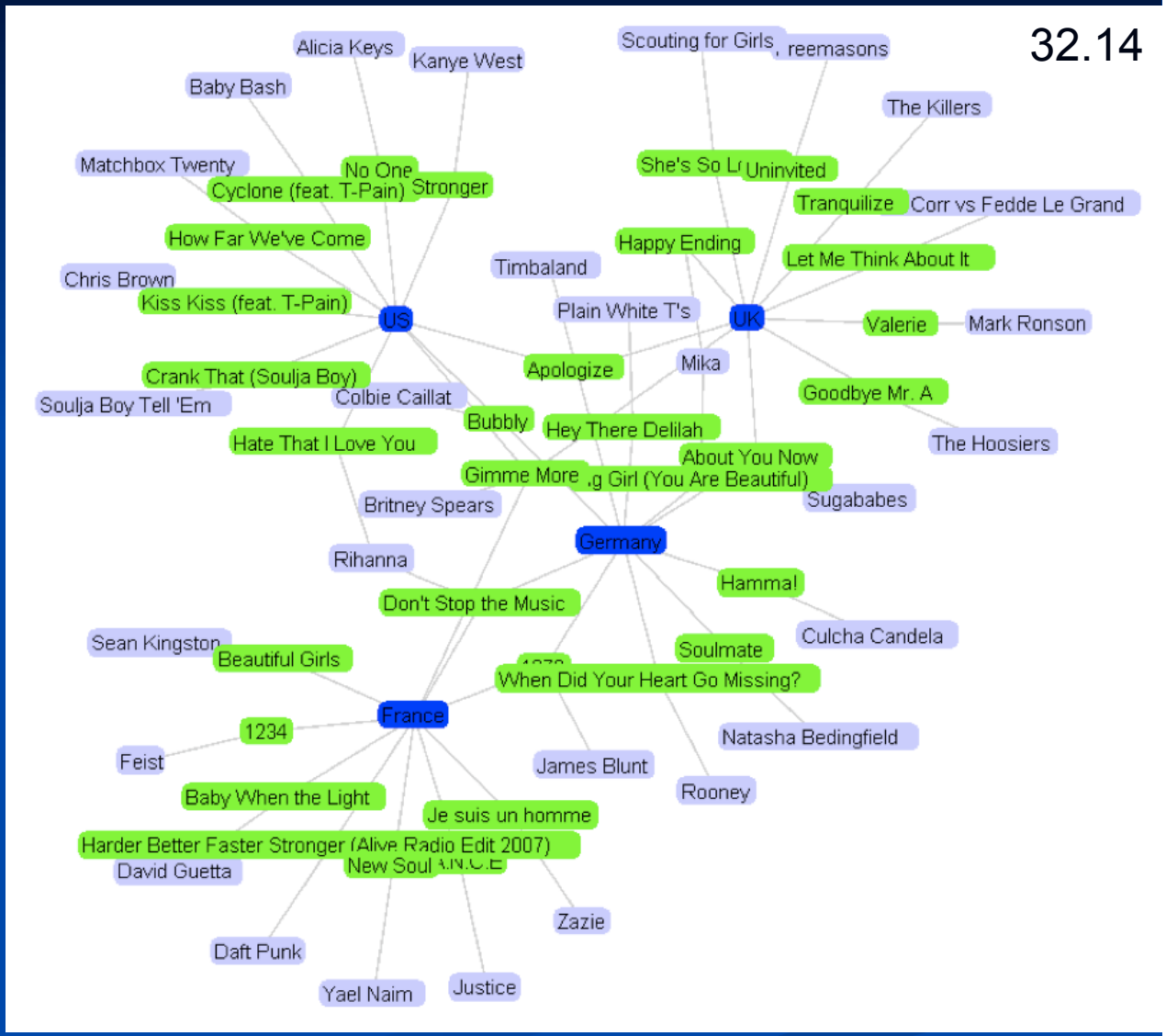


Prefuse (Graph Network)

Source (iTunes)		Target (Graph)
Country	→	Node 0
Song Name		Node 1
Artist Name		Level 2

# of mapping permutations: 6
of *visible* map'g permutations: 6





Summary:

- Ontologies with Certainty Factors
- Ontology Mapping

- **Pragmatic Visualization Pipeline**
  - Domain Ontology (DO)
  - Visual Representation Ontology (VRO)
  - Semantic Bridge Ontology (SBO)

- **Scored (and ranked) visualizations**
  - Automatic Information Visualization for Non-Experts

Thank you

Owen Gilson
Swansea University, UK
cs.swan.ac.uk/~csowen