AGREEING TO DISAGREE: RECONCILING CONFLICTING TAXONOMIC VIEWS USING A LOGIC-BASED APPROACH

Yi-Yun Cheng[1], Nico Franz[2], Jodi Schneider[1], Shizhuo Yu[3], Thomas Rodenhausen[4], Bertram Ludäscher[1]

[1] School of Information Sciences, University of Illinois at Urbana-Champaign; [2] School of Life Sciences, Arizona State University;
[3] Department of Computer Science, University of California at Davis; [4] School of Information, University of Arizona

School of Information Sciences, The iSchool at Illinois

Biodiversity Knowledge Integration Center, Arizona State University

CIRSS

INTRODUCTION

Tina: *Hey Amy, can you recommend a signature dish from where you live?*

Amy: *Oh, definitely the half-smokes from the Northeast! They are these tasty half-pork and half-beef sausages.*

Tina: *What a coincidence! We have half-smokes in the South, too! Where do you live in the Northeast? New York? Boston?*

Amy: *Wrong guesses! Where do you live in the South?*

Tina and Amy together: *Washington, D.C.*

[The two of them look at each other, confused.]

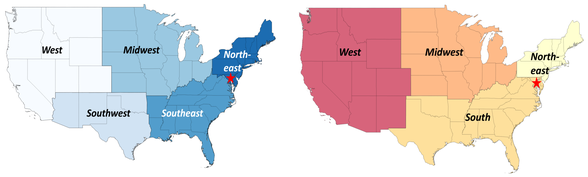


Figure 1. National Diversity Council map (NDC) vs. Census Bureau map (CEN)

"*In the face of incompatible information or data structures among users or among those specifying the system, attempts to create unitary knowledge categories are futile. Rather, parallel or multiple representational forms are required…*" (Bowker & Star, 2000).

RELATED WORK

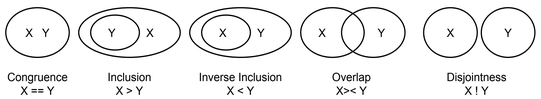
- **Taxonomy Alignment Problems (TAP)**

Taxonomies T_1, T_2 are inter-linked via a set of input *articulations A*, defined as RCC-5 relations, to yield a "merged" taxonomy T_3 .

- **Euler/X**

Articulations – a constraint or rule that defines a relationship (a set constraint) between two concepts from different taxonomies .

Region Connection Calculus (RCC-5)



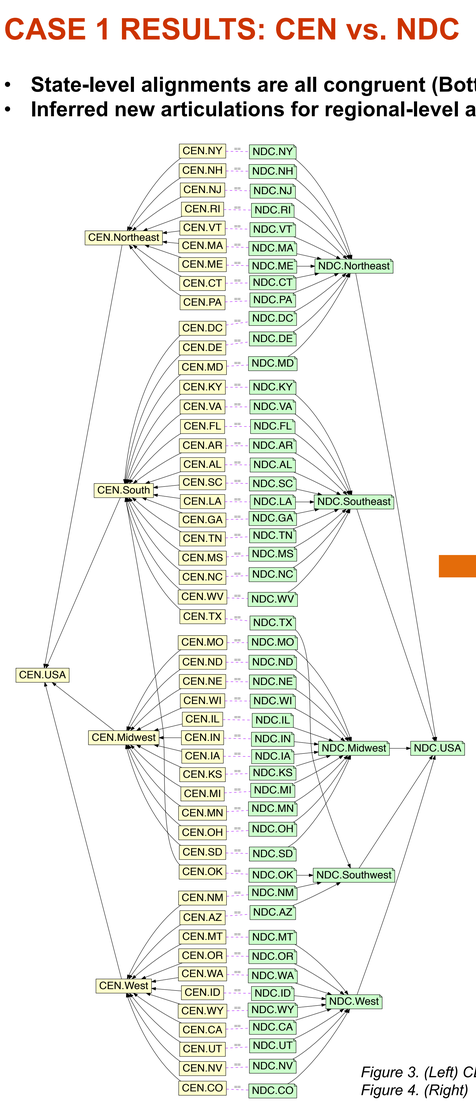
Possible Worlds – When encoding and solving TAPs via ASP, the different answer sets represent alternative taxonomy merge solutions or possible worlds (PWs).

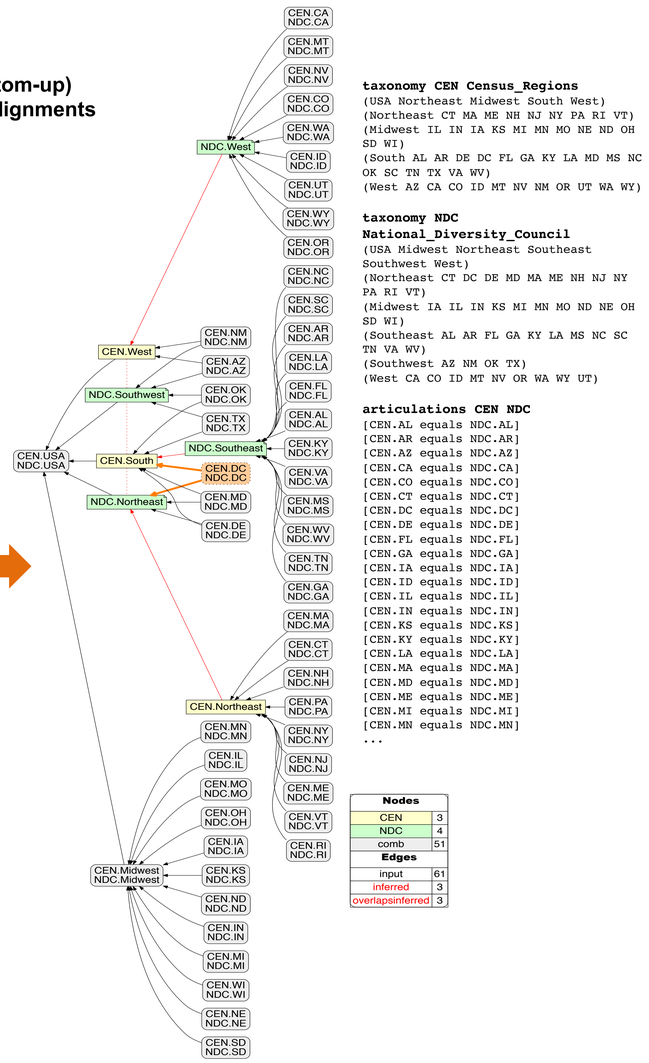
- **Github link**: https://github.com/EulerProject/ASIST17
- **Email**: yiyunyc2@illinois.edu

Quick Scan!

CASE 1 RESULTS: CEN vs. NDC

- **State-level alignments are all congruent (Bottom-up)**
- **Inferred new articulations for regional-level alignments**





taxonomy CEN Census_Regions
(USA Northeast Midwest South West)
(Northeast CT MA ME NH NJ NY PA RI VT)
(Midwest IL IN IA KS MI MN MO NE ND OH SD WI)
(South AL AR DE DC FL GA KY LA MD MS NC OK SC TN TX VA WV)
(West AZ CA CO ID MT NV NM OR UT WA WY)

taxonomy NDC National_Diversity_Council
(USA Midwest Northeast Southeast Southwest West)
(Northeast CT DC DE MD MA ME NH NJ NY PA RI VT)
(Midwest IA IL IN KS MI MN MO ND NE OH SD WI)
(Southeast AL AR FL GA KY LA MS NC SC TN VA WV)
(Southwest AZ NM OK TX)
(West CA CO ID MT NV OR WA WY UT)

articulations CEN NDC
[CEN.AL equals NDC.AL]
[CEN.AR equals NDC.AR]
[CEN.AZ equals NDC.AZ]
[CEN.CA equals NDC.CA]
[CEN.CO equals NDC.CO]
[CEN.CT equals NDC.CT]
[CEN.DC equals NDC.DC]
[CEN.DE equals NDC.DE]
[CEN.FL equals NDC.FL]
[CEN.GA equals NDC.GA]
[CEN.IA equals NDC.IA]
[CEN.ID equals NDC.ID]
[CEN.IL equals NDC.IL]
[CEN.IN equals NDC.IN]
[CEN.KS equals NDC.KS]
[CEN.KY equals NDC.KY]
[CEN.LA equals NDC.LA]
[CEN.MA equals NDC.MA]
[CEN.MD equals NDC.MD]
[CEN.ME equals NDC.ME]
[CEN.MI equals NDC.MI]
[CEN.MN equals NDC.MN]
...

Nodes	
CEN	3
NDC	4
comb	51
Edges	
input	61
inferred	3
overlapsinferred	3

Figure 3. (Left) CEN-NDC taxonomy alignment problem with 49 input articulations between T_{CEN} and T_{NDC}
Figure 4. (Right) The unique possible world (PW) T_3 reconciling T_{CEN} and T_{NDC} via inferred relationships

CASE 2 RESULTS: CEN vs. TZ

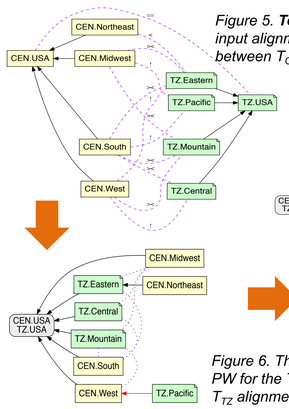


*Figure 5. **Top-down** input alignments between T_{CEN} and T_{TZ}*

Figure 6. The unique PW for the T_{CEN} with T_{TZ} alignment

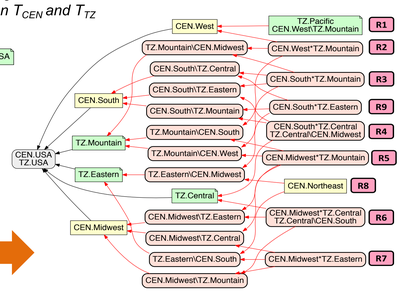
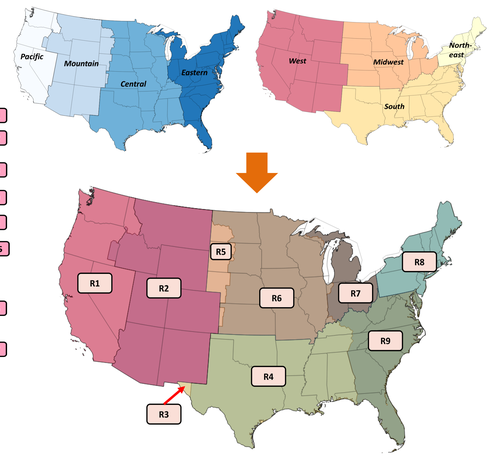
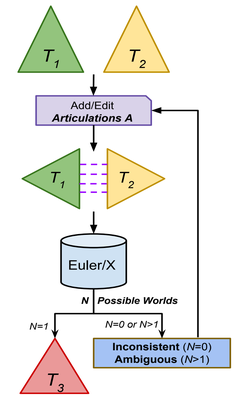


Figure 7. Combined concepts solution for T_{CEN} and T_{TZ}



RESEARCH DESIGN

Step 1. Supply input taxonomies T_1 and T_2
Step 2. Formulate RCC-5 articulations between T_1 and T_2
Step 3. Iteratively edit articulations in Euler/X



taxonomy CEN Census_Regions
(USA Midwest South West Northeast)

taxonomy TZ Time_Zone
(USA Pacific Mountain Central Eastern)

articulations CEN TZ
[CEN.Midwest disjoint TZ.Pacific]
[CEN.Midwest overlaps TZ.Eastern]
[CEN.Midwest overlaps TZ.Mountain]
[CEN.Northeast is_included_in TZ.Eastern]
[CEN.South disjoint TZ.Pacific]
[CEN.South overlaps TZ.Central]
[CEN.South overlaps TZ.Eastern]
[CEN.South overlaps TZ.Mountain]
[CEN.USA equals TZ.USA]
[CEN.West disjoint TZ.Central]
[CEN.West disjoint TZ.Eastern]
[CEN.West overlaps TZ.Mountain]

Figure 2. The process of aligning taxonomies T1 and T2 with Euler/X

CONCLUSION

- **Our logic-based taxonomy alignment approach can be used to solve crosswalking issues**
  We will be able to mitigate the membership condition problems that occur in equivalent crosswalking.

- **RCC-5 approach preserves the original taxonomies while providing an alignment view**
  We can solve data integration problems that happen in the more coarse-grained relative crosswalking, which otherwise is subjected to information loss.

- **Our study also underscores the benefits of designing different alignment workflows (Bottom-up vs. Top-down) to match the needs of specific taxonomy alignment problems**
  Bottom-up approach: seems to work well whenever we have non-overlapping relationships at the leaf-level (lowest-level) articulations, and we are not sure how the higher-level concepts should be aligned.

  Top-down approach: seems favorable when there is an expectation of certain higher-level articulations in conjunction with under-specified, complex, and often overlapping leaf-level relations.

ACKNOWLEDGEMENTS

Support of the authors' research through the National Science Foundation is kindly acknowledged (DEB-1155984, DBI-1342595, and DBI-1643002). The authors thank Professor Kathryn La Barre for her comments and suggestions. We would also like to thank Dr. Laetitia Navarro and Jeff Terstriep for help with creating map overlays in QGIS.