Sketches, Views and Pattern-Based Reasoning

8th International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS 2013)

13–14 November 2013

George Mason University

C4I Center

Dr. Ralph L. Wojtowicz

Shepherd University
Shepherdstown, WV
rwojtowi@shepherd.edu

Baker Mountain Research Corporation
Yellow Spring, WV
ralphw@bakermountain.org

Shepherd University

Baker Mountain
Science Technology Service

Objectives and Themes

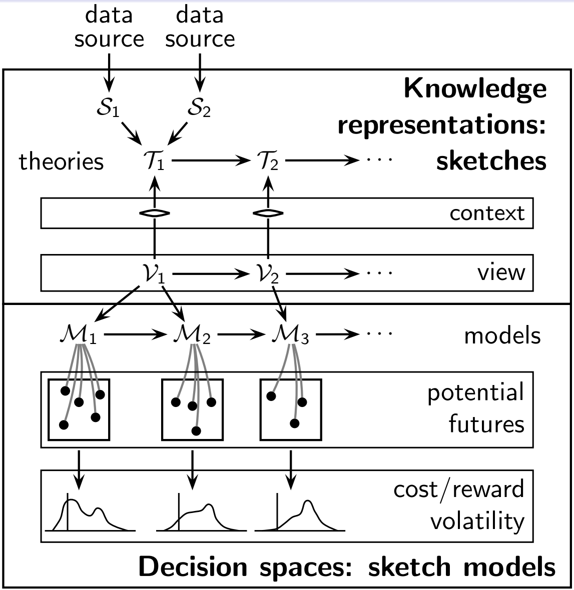
Objectives

- Introduce the sketch data model to researchers and practitioners of other semantic technologies
- Describe a program for its development and application

Themes

- Functional modeling paradigm: syntax; constraints; models; maps between sketches, models and logical theories
- Sketches (and ontologies) are presentations of knowledge
- A sketch generates a theory which is used to align presentations and extract context-sensitive views of knowledge bases
- Uncertainty and lack of information are accounted for in models
- Classes of sketches, like OWL species, correspond to fragments of the predicate calculus
- Programming framework: Easik + disparate research components

Sketch Conops

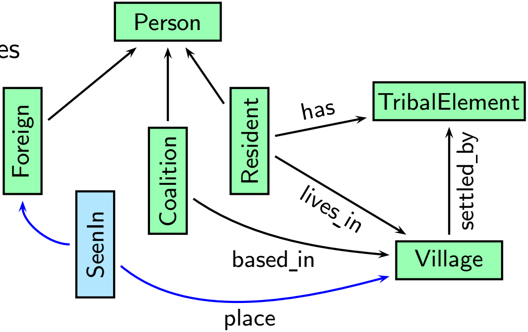


Historical Timeline

- 1943: Eilenberg and Mac Lane introduce category theory
- 1958: Kan introduces the concept of adjoints
- 1963: Lawvere characterizes quantifiers and other logical operations as adjoints
- 1968: C. Ehresman introduces sketch theory
- 1985: KL-ONE — First implementation of a description logic system
- 1985: Barr and Wells publish *Toposes, Triples and Theories*
- 1989: J. W. Gray publishes Category of Sketches as a Model for Algebraic Semantics
- 1990: Barr and Wells publish *Categories for Computing Science*
- 1995: Carmody and Walters publish algorithm for computing left Kan extensions
- 1999: RDF becomes a W3C recommendation
- 2000: Johnson and Rosebrugh apply sketch data model to database interoperability
- 2000: DARPA begins development of DAML
- 2001: Dampney, Johnson and Rosebrugh apply sketches to view update problem
- 2001: W3C forms the Web-Ontology Working Group
- 2004: RDFS and OWL become W3C recommendations
- 2008: Johnson and Rosebrugh release Easik software
- 2009: OWL2 becomes a W3C recommendation
- 2012: Johnson, Rosebrugh and Wood use sketches to formulate lens concept of view updates

Example Human Terrain Sketch

- Vertices represent both classes and relations (properties).
- Edges represent functions.
- Edges from relation vertices give type information.
- Instance data is represented in models.

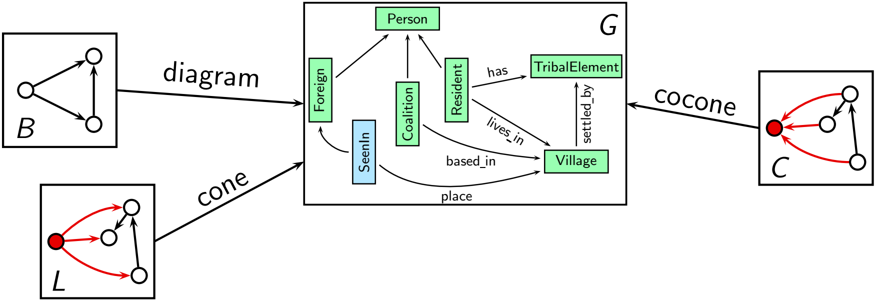


Intended semantic constraints:

- Commutativity of the Resident-Village-TribalElement triangle
- Foreign, Coalition and Resident as mutually exclusive and exhaustive subtypes of Person

Sketch $(G, \mathcal{D}, \mathcal{L}, \mathcal{C})$

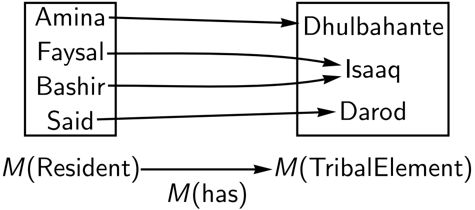
- All semantic constraints in a sketch are expressed using graph maps.
- A sketch $(G, \mathcal{D}, \mathcal{L}, \mathcal{C})$ consists of:
  - An underlying graph G
  - A set $\mathcal{D}$ of diagrams $B \to G$
  - A set $\mathcal{L}$ of cones $L \to G$
  - A set $\mathcal{C}$ of cocones $C \to G$



Set-Based Sketch Models

Set-based model of a graph

- Each vertex V is mapped to a set $M(V)$.
- Each edge $V \xrightarrow{e} W$ is mapped to a function $M(V) \xrightarrow{M(e)} M(W)$.

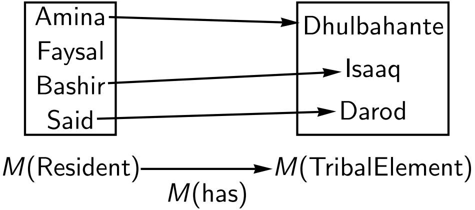


Set-based model of a sketch $(G, \mathcal{D}, \mathcal{L}, \mathcal{C})$

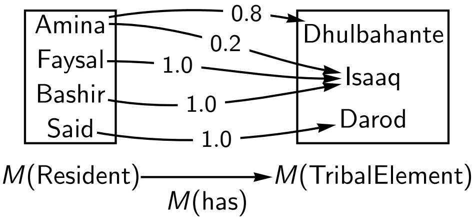
- A sketch model is, first, a model of the underlying graph G.
- Sketch constraints impose additional requirements on models.
- Expressiveness of the sketch imposes requirements on suitable categories of semantic models.

Sketch Models in Other Semantic Categories

- Partial function model of a graph edge Resident $\xrightarrow{\text{has}}$ TribalElement



- Stochastic matrix model of a graph edge Resident $\xrightarrow{\text{has}}$ TribalElement



Categorical Semantics of Sketches

- Classes of constraints (cones and cocones) are distinguished by the shapes of their base graphs.
- Classes of sketches are distinguished by their classes of constraints.
- Like OWL species (DL, EL, RL), these have different expressive powers.
- Unlike OLW, sketch classes were formulated for their expressive power relative to logical theories, not for computational complexity properties.

Small sample of the sketch semantics landscape

Sketch Class	Set	Partial Func.	Stoch. Matrices	Čencov Cat.	Prob. 0 Refl.	Dempster Shafer	Fuzzy Sets	Convex Sets
Regular	•	•	•	•	•	•	•	•
Finite Limit	•	•	×	×	×	×	•	•
Finite Coproduct	•	•	•	•	•	•	•	•
Entity-Attribute	•	•	×	×	×	×	•	•
Mixed	•	•	×	×	×	×	•	•

Sketch Maps and Model Maps

- A sketch map $\mathcal{S}_1 \to \mathcal{S}_2$ is a graph map

$$G_1 \longrightarrow G_2$$

  that preserves all the constraints of $\mathcal{S}_1$.

$$B \longrightarrow G_1 \longrightarrow G_2$$

- We use sketch maps to formulate the Alignment Problem.

- Given models M_1 and M_2 of a sketch $\mathcal{S}$, a model map $M_1 \to M_2$ is a collection of functions (one for each vertex V of G)

$$M_1(V) \xrightarrow{\tau_V} M_2(V)$$

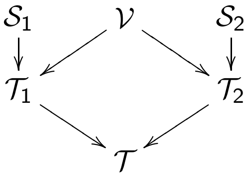
  that are consistent with the edges of G.

- Example:

$$\begin{array}{c} \text{Resident} \\ \text{live\_in} \downarrow \\ \text{Village} \end{array} \qquad \begin{array}{ccc} M_1(\text{Resident}) & \xrightarrow{\tau} & M_2(\text{Resident}) \\ {\scriptstyle M_1(\text{lives\_in})} \downarrow & & \downarrow {\scriptstyle M_2(\text{lives\_in})} \\ M_1(\text{Village}) & \xrightarrow[\tau]{} & M_2(\text{Village}) \end{array}$$

Presentations

- A sketch or ontology is a presentation of knowledge.
- A sketch generates a mathematical structure called its theory.
- Different presentations may generate identical theories.
- Sketches $\mathcal{S}_1$ and $\mathcal{S}_2$ representing common concepts are aligned by finding a sketch $\mathcal{V}$ and sketch maps as shown.

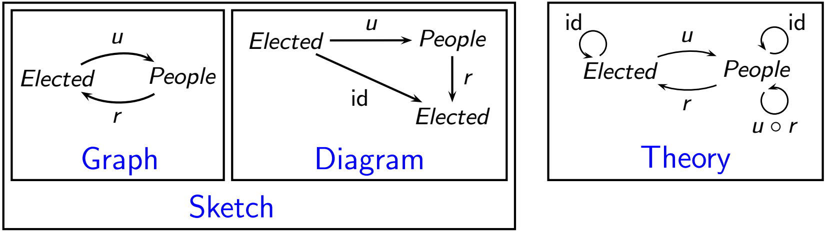


- $\mathcal{V}$ is the overlap or intersection.
- We can demonstrate solution of the alignment problem using two sketches as distinct presentations of the same concepts.

Civics Sketch $\mathcal{S}_1$

First formulation of civics concepts:

- Two classes: People and Elected officials
- People have Elected representatives via r.
- Elected officials are instances of people via u.
- Elected officials represent themselves via a diagram.



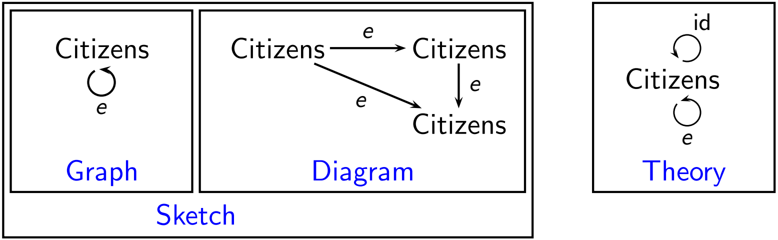
- The diagram truncates the infinite list of composites (property chains).

$$u \circ r \qquad r \circ u \qquad u \circ r \circ u \qquad r \circ u \circ r \qquad \cdots$$

Civics Sketch $\mathcal{S}_2$

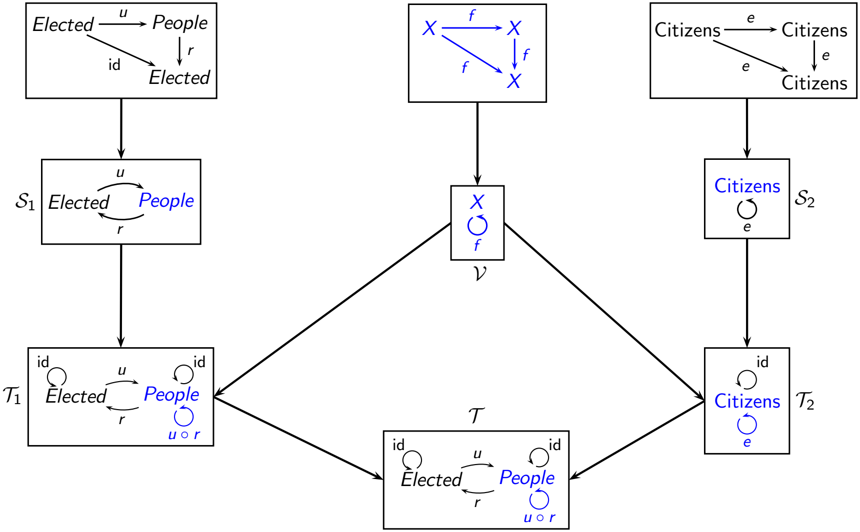
Alternative formulation of the concepts:

- One class: Citizens
- Citizens have elected representatives via e.
- Elected officials represent themselves via a diagram.



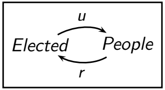
- Number and names of vertices in $\mathcal{S}_1$ and $\mathcal{S}_2$ differ.
- The edges u and r of $\mathcal{S}_1$ have no corresponding edges in $\mathcal{S}_2$.
- The edge e of $\mathcal{S}_2$ has no corresponding edge in $\mathcal{S}_1$.

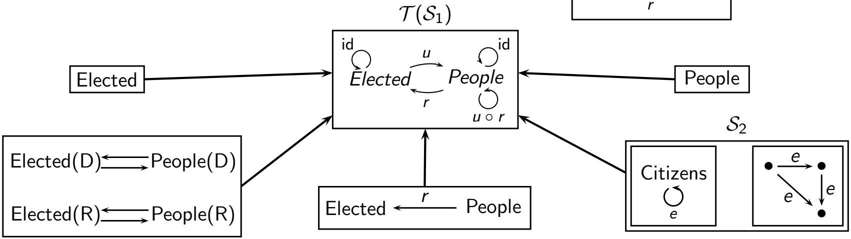
Alignment of the Civics Sketches



Views

- A view $\mathcal{V} \Longrightarrow \mathcal{S}$ of a sketch $\mathcal{S}$ is a sketch $\mathcal{V}$ and a sketch map
$$\mathcal{V} \longrightarrow \mathcal{T}(\mathcal{S})$$
from $\mathcal{S}$ to the theory $\mathcal{T}(\mathcal{S})$ of $\mathcal{S}$.
- Below are several views of the civics sketch $\mathcal{S}_1 =$





- A model of $\mathcal{S}$ induces a model of $\mathcal{T}(\mathcal{S})$ and of its views $\mathcal{V} \to \mathcal{T}(\mathcal{S}) \stackrel{M}{\to}$ Set.
- Views may be composed $\mathcal{V}_2 \Longrightarrow \mathcal{V}_1 \Longrightarrow \mathcal{S}$.
- View Update Problem: Under what conditions can updates to a model of $\mathcal{V}$ be propagated to a model of $\mathcal{S}$?

Challenge: Construct Views Tailored to Mission Contexts

- Research area with narrower scope: context-sensitive Internet search
  - Google patent for "methods, systems and apparatus including computer program products, in which context can be used to rank search results" (USPTO 8,209,331 — 2012)
  - Yandex personalized web search challenge: www.kaggle.com
- Techniques to infer context from activities and rank data elements
  - Variable-length hidden Markov model
  - Parametric models of users
  - RankNet, LambdaRank, RankSVM
- Performance metrics used for context-sensitive rankings
  - Normalized discounted cumulative gain (scoring in Kaggle competition)
  - Kendall's τ comparison of rankings
  - Jaccard distance between top N rankings and target

Transforming Sketches into Logical Theories

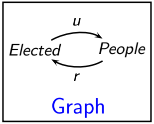
- Sketches are related to first-order logical theories by theorems of the form: Given any sketch $\mathcal{S}$ of class X, there is a logical theory $\mathbb{T}$ of class Y for which $\mathcal{S}$ and $\mathbb{T}$ have equivalent classes of models.
- D2.2 of Johnstone's *Sketches of an Elephant: A Topos Theory Compendium* gives explicit constructions of $\mathbb{T}$ from $\mathcal{S}$ and conversely.

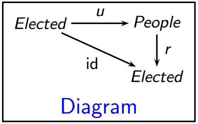
Class of Sketches	Fragment of Predicate Calculus	Logical Connectives
finite limit	cartesian	$=, \top, \wedge, \exists^*$
regular	regular	$=, \top, \wedge, \exists$
coherent	coherent	$=, \top, \wedge, \exists, \bot, \vee$
geometric	geometric	$=, \top, \wedge, \exists, \bot, \bigvee$
σ-coherent	σ-coherent	$=, \top, \wedge, \exists, \bot, \bigvee_{i=1}^{\infty}$
finitary	σ-coherent	

* In cartesian logic, only certain existentially quantified formulae are allowed.

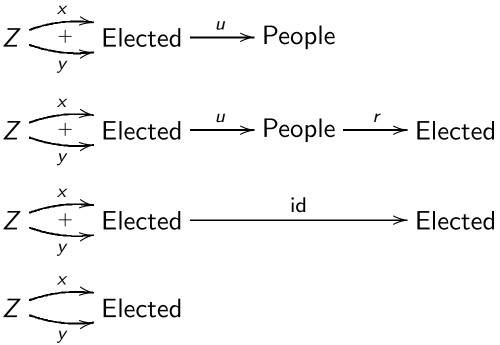
Graphical Inference

In civics sketch $\mathcal{S}_1$, we may conclude that Elected is a subclass of People.





Given any x and y as shown:



It follows that u is a monomorphism (one-to-one) in any model.

Software Infrastructure

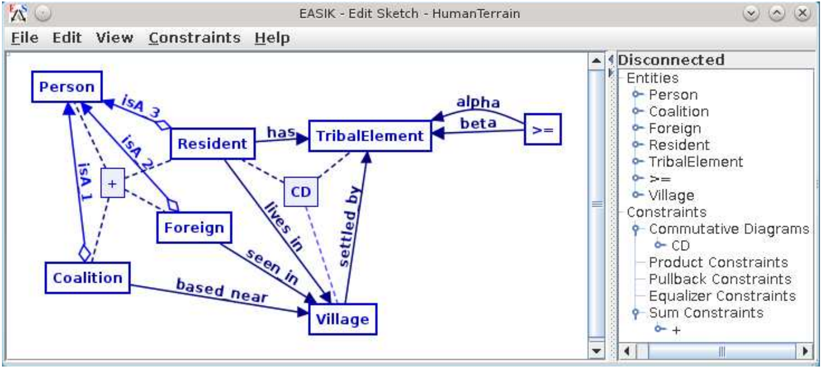
- Relative to Jena platform, infrastructure for working with sketches is meager.
- Set-based models can be implemented using database features.

Sketch	database schema
Vertex	table with automatically-generated (`Serial`) key
Edge $A \xrightarrow{e} B$	foreign key in A-table referencing B-table key
Constraints	triggers

- Challenge: management of distributed models of sketches, views and constraints — Google Megastore, Tenzing and Spanner; Apache Cassandra
- Reasoning
  - Utilize computation category theory tools: various research activities including Rydeheard and Burstall (ML implementations) 1988
  - Transform to first-order theory and employ theorem prover
- Theory of a (linear) sketch
  - Carmody-Walters algorithm for computing left Kan extensions: generalizes Todd-Coxeter procedure used in computational group theory
  - Complexity difficult to characterize: can depend on order of constraints

Easik Tool for Modeling with Sketches

- Entity Attribute Sketch Implementation Kit (Easik)
- http://mathcs.mta.ca/research/rosebrugh/Easik
- Build sketches, views and constraints
- Interface with MySQL or Postgres for (set-based models)
- No reasoning engine



Conclusions

- The sketch data model demonstrates valuable features
  - Functional paradigm: syntax, models and maps
  - Separation of knowledge representation from its models
  - Uncertainty and lack of information accounted for in models
  - Context-sensitive views which can be composed and combined
  - Formulation of the alignment problem using a well-defined mathematical construction (theory of a sketch)
  - Reasoning via graphical arguments or transformation to predicate calculus fragment
- Challenges
  - Implement sketch constraints on large, distributed models
  - Leverage insights, datasets and performance metrics from the narrower problem of context-sensitive Internet search
  - Develop and implement semi-automated alignment tool
  - Integrate reasoning and modeling algorithms with instance data into a common software platform
  - Characterize sketch classes corresponding to OWL species

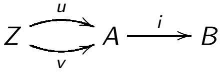
References

- M. Barr and C. Wells. *Toposes Triples and Theories*. Springer-Verlag. 1985
- M. Barr and C. Wells. *Category Theory for Computing Sciences*. Prentice-Hall. 1990
- S. Carmody, M. Leeming and R. F. C. Walters. The Todd-Coxeter Procedure and Left Kan Extensions. *J. Symbolic Computation*. **19**:459–488. 1995
- C. Ehresmann. Esquisses et types de structures algébriques. *Bul. Inst. Politehn. Iași.* **14**:1–14. 1968
- J. W. Gray. The Category of Sketches as a Model for Algebraic Semantics. In *Categories in Computer Science and Logic*. V. 92 of Contemporary Mathematics. AMS. 1989
- M. Johnson and R. Rosebrugh. Sketch Data Models, Relational Schema and Data Specifications. *Electronic Notes in Theoretical Computer Science*. **61**(6):1–13. 2002
- M. Johnson, R. Rosebrugh and R. J. Wood. Lenses, Fibrations and Universal Translations. *Mathematical Structures in Computer Science*. **22**:25–42. 2012
- P. E. Johnstone. *Sketches of an Elephant: A Topos Theory Compendium*. Oxford University Press. 2002
- J. J. M. M. Rutten, M. Kwiatkowska, G. Norman, and D. Parker. *Mathematical Techniques for Analyzing Concurrent and Probabilistic Systems*. AMS. 2004
- D. E. Rydeheard and R. M. Burstall. *Computational Category Theory*. Prentice-Hall. 1988

Appendix

Subtype Constraints via Cones

- A is a subclass of B if and only if each A instance is a B instance.
- Replace this membership-based formulation by a functional one: A is a subclass of B if and only if there is a one-to-one function $i : A \to B$.
- i is one-to-one if: for any x, y in A, if $i(x) = i(y)$ then $x = y$.
- Equivalently, for any u and v as shown below, if $i \circ u = i \circ v$ then $u = v$.



- Equivalently, there is a unique φ making the diagram below left commute.
- That is, the diagram below right is a cone.