



# A formal approach to design a large scale domain ontology

**BISWANATH DUTTA**

INDIAN STATISTICAL INSTITUTE  
DOCUMENTATION RESEARCH AND TRAINING CENTRE  
BANGALORE, INDIA

EMAIL: [BISU@DRTC.ISIBANG.AC.IN](mailto:BISU@DRTC.ISIBANG.AC.IN)

**Dutta, B., Chatterjee, U. and Madalli, D. P. (2015), "YAMO: Yet Another Methodology for large-scale faceted Ontology construction", Journal of Knowledge Management, Vol. 19 Iss 1 pp. 6 – 24.**

# Introduction

---

- A brand new step-by-step approach
- Provides a set of guiding principles
- Approach is domain independent
- Approach is motivated by the **facet analysis** and **analytico-synthetic** classification (Ranganathan, 1967)
  - This ensures the design of an ontology consisted of clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of concepts of a domain of interest

# Past Approaches

---

- **DILIGENT** focuses on ontology evolution rather than initial ontology designing (Vrandecic *et al.*, 2005)
- **Toronto Virtual Enterprise (TOVE)** mainly highlights ontology evaluation and maintenance (Gruninger and Fox, 1995)
- **ENTERPRISE** discusses the informal and formal phases of ontology construction, but is unable to clearly state how an ontological concept can be identified (Uschold *et al.*, 1995)
- **IDEF5** (KBSI, 1994) and **METHONTOLOGY** (Fernandez *et al.*, 1997) provide more emphasis on ontology maintenance
  
- **Problem**: there exists no such methodology that gives a detailed description of the steps along with a set of principles that are to be undertaken to build an ontology.

# Talk Overview

---

Two-way approach

Ten steps

Guiding principles

Result

Conclusion

# Ontology

---

- “a formal, explicit specification of a shared conceptualization”
- A formal explicit description of concepts or classes in a domain of discourse, with properties (roles or slots) of each concept describing various features and attributes of the concepts (*Noy and McGuinness, 2001*)
- An ontology potentially brings out the conceptual knowledge by establishing richer semantic relationships.

# Two-way approach

---

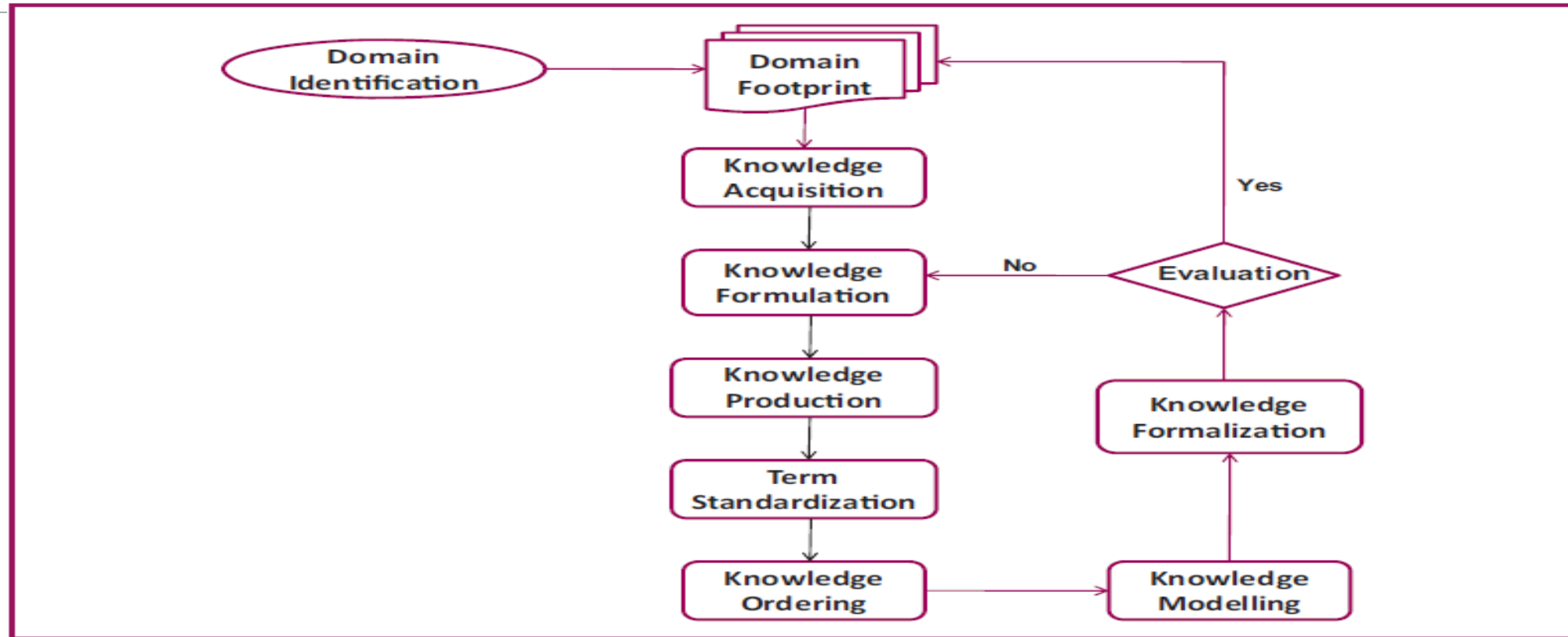
- **Top-down approach**

- Involves in drawing the big-picture of an ontology at an abstract level
- Proceeds from an abstract level and reaches to a concrete level

- **Bottom-up approach**

- Involves in identifying and studying the characteristics of base concepts and assembling them depending upon their similar features
- Proceeds from a concrete ground and reaches to an abstract level

# Yet Another Methodology for Ontology development (YAMO) Steps



\*Documentation at each step

# Principles

---

- Principle of relevance
- Principle of ascertainability
- Principle of permanence
- Principle of exclusiveness
- Principle of exhaustivity
- Principle of consistency
- Principle of context
- Principle of Helpful Sequence

(Ranganathan, 1967)



# Step0: Domain identification

---

- Identify the domain based on the project goal and application needs.
- E.g., food, disaster, music, movie

# Step1: Domain footprint

---

- Create a set of use scenarios and based on that create a set of questions.
- E.g., Scenario: visiting a restaurant

1. What is the special item available for the day?
2. How many pieces of chicken will be served in the plate?
3. How much time will it take to serve the dish?
4. Will the sauce be spicy/hot/mild/sweet?
5. Which is the most popular vegetarian item of the restaurant?

6. How will the dish be prepared (fried/roasted/sautéed)?
7. Does the restaurant serve halal meat?
8. What is available for starters?
9. What are the main ingredients present in the dish?
10. What are the desserts available for diabetic patient?

# Step2: Knowledge acquisition

---

- Involves in identifying a set of terms relevant to the domain.
- **E.g.**, Salad, chicken, eggplant, chicken kebab, ice cream, bacon, bean, avocado, whisky, tomato, butter, almond, spinach, protein shake, white wine, humus, oatmeal, coffee, wine, milk, lettuce, ...

# Step3: Knowledge formulation

---

- Involves in **analyzing** the terms collected in the previous step.
- Analysis is done based on the definition, characteristic and appropriateness of the identified terms.
- **E.g.**,
  - *red wine*: wine having a red color derived from the skins of dark-colored grapes;
  - *white wine*: pale yellowish wine made from white grapes with skins removed before fermentation;
  - *pink wine*: pinkish table wine from red grapes whose skins are removed after fermentation began.

# Step4: Knowledge production

---

- This phase results in **facet** discovery and arrangement.

Edible Food	Drinkable Food
Animal Origin Food	Alcoholic Drink
Meat Product	Fermented Beverage
Bird Product	Wine
Chicken Kebab	Red Wine
Fish Product	Distilled Beverage
Smoked Salmon	Whisky

# Step5+6: Term standardization and ordering

---

- Standardizes the terms.
  - E.g., term beverage (any liquid suitable for drinking) has synonymous terms like *drink*, *drinkable*, and *potable*.
- Knowledge Ordering involves in ordering the terms within the array as per the system goals.
  - E.g., increasing and decreasing complexity of knowledge, increasing and decreasing quantity, literary warrant, centre to periphery, periphery to centre, chronological order, canonical order, alphabetical order, later in evolution, etc.).

Edible Food	Drinkable Food
Animal Origin Food	Alcoholic Drink
Meat Product	Distilled Beverage
Fish Product	Whisky
Smoked Salmon	Fermented Beverage
Bird Product	Wine
Chicken Kebab	Red Wine

# Step7: Knowledge modelling

- Representation of the derived knowledge based on **DERA framework** (a faceted knowledge organization framework) (Giunchiglia and Dutta, 2011).

ENTITY TYPE	ATTRIBUTES	RELATION
<ul style="list-style-type: none"> <li>✚ Food           <ul style="list-style-type: none"> <li>▪ Edible Food               <ul style="list-style-type: none"> <li>○ Plant Origin Food                   <ul style="list-style-type: none"> <li>• Processed Vegetable                       <ul style="list-style-type: none"> <li>Cereal Product</li> <li>Fruit Product</li> <li>Legume Product</li> <li>Vegetable Product</li> <li>Mixed Vegetable Product</li> </ul> </li> <li>• Raw Vegetable                       <ul style="list-style-type: none"> <li>Root Vegetable</li> <li>Stem Vegetable</li> <li>Edible Flower</li> <li>Edible Fruit</li> <li>Edible Seed</li> </ul> </li> </ul> </li> <li>○ Animal Origin Food                   <ul style="list-style-type: none"> <li>• Egg Product</li> <li>• Milk Product</li> <li>• Meat Product</li> </ul> </li> <li>○ Mixed Origin Food                   <ul style="list-style-type: none"> <li>• Soup</li> <li>• Snack</li> <li>• Salad</li> <li>• Dessert</li> </ul> </li> </ul> </li> <li>▪ Drinkable Food               <ul style="list-style-type: none"> <li>○ Alcoholic Drink                   <ul style="list-style-type: none"> <li>• Fermented Beverage</li> <li>• Distilled Beverage</li> <li>• Fortified Beverage</li> <li>• Cocktail</li> </ul> </li> <li>○ Non Alcoholic Drink                   <ul style="list-style-type: none"> <li>• Drinking Water</li> <li>• Nourishing Drink</li> <li>• Refreshing Drink</li> <li>• Stimulating Drink</li> </ul> </li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✚ Price</li> <li>✚ Rating</li> <li>✚ Regional Name</li> <li>✚ Critics Review</li> <li>✚ Serving Size</li> <li>✚ Image</li> <li>✚ Calorie Content</li> <li>✚ Cooking Level</li> <li>✚ Expiry Date</li> <li>✚ Manufactured Date</li> <li>✚ Serving Temperature</li> <li>✚ Special Occasion</li> <li>✚ Available Quantity           <ul style="list-style-type: none"> <li>▪ Alcohol Volume</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✚ Type Of Food           <ul style="list-style-type: none"> <li>▪ Belief Based Food</li> <li>▪ Weight Control Food</li> <li>▪ Vegetarian Food</li> </ul> </li> <li>✚ Accompaniment           <ul style="list-style-type: none"> <li>▪ Raita</li> <li>▪ Dips</li> <li>▪ Pickle</li> <li>▪ Sauce</li> <li>▪ Chatney</li> </ul> </li> <li>✚ Ingredient           <ul style="list-style-type: none"> <li>▪ Animal Origin Ingredient</li> <li>▪ Plant Origin Ingredient</li> </ul> </li> <li>✚ Recipe           <ul style="list-style-type: none"> <li>▪ Recipe Category</li> <li>▪ Recipe Type</li> <li>▪ Recipe Instruction</li> </ul> </li> <li>✚ Cuisine           <ul style="list-style-type: none"> <li>▪ Global Cuisine</li> <li>▪ Regional Cuisine</li> </ul> </li> <li>✚ Cooking Method           <ul style="list-style-type: none"> <li>▪ Dry Heat Cooking</li> <li>▪ Moist Heat Cooking</li> <li>▪ Frying</li> <li>▪ Microwaving</li> </ul> </li> <li>✚ Course Description           <ul style="list-style-type: none"> <li>▪ Appetizer</li> <li>▪ Beverage</li> </ul> </li> <li>✚ Additive           <ul style="list-style-type: none"> <li>▪ Acid</li> <li>▪ Food Flavouring</li> <li>▪ Food Colorant</li> </ul> </li> <li>✚ Preservative           <ul style="list-style-type: none"> <li>▪ Drying</li> <li>▪ Salting</li> </ul> </li> </ul>

# Step8: Knowledge formalization

- Based on Description Logics.

Entity type	Food domain elements	DL formalization
$E_1, \dots, E_m$	Entity classes	Concept TBox
$R_1, \dots, R_s$	Relation between classes	Roles
$A_1, \dots, A_u$	Attributes	Roles
is-a	Hierarchical relation	Subsumption ( $\sqsubseteq$ )
part-of	Hierarchical relations	Role
Any relations that are non-hierarchical type (e.g. ingredient, calorie content, diet)	Associative relations	Role
value-of	Hierarchical relation	Role restriction
$e_1, \dots, e_n$	Entity instances	Individuals ABox
$r_1, \dots, r_t$	Relation between entities	Role assertions
$a_1, \dots, a_v$	Attributes of entities	Role assertions
$v_1, \dots, v_u$	Attribute values	Individuals
instance-of	Hierarchical relations (between entity class and entity instances)	Concept assertions

TBox	ABox
Food $\equiv$ EdibleFood $\sqcup$ DrinkableFood	ChickenKebab(chicken_kebab)
EdibleFood $\equiv$ AnimalOriginFood $\sqcup$ PlantOriginFood $\sqcup$ MixedOriginFood	mainIngredient(chicken_kebab, chicken)
MeatProduct $\sqsubseteq$ AnimalOriginFood	preparationMethod(chicken_kabab, roasting)
BirdProduct $\sqsubseteq$ MeatProduct	taste(chicken_kebab, spicy)
ChickenKebab $\equiv$ BirdProduct $\sqcap$ $\exists$ mainIngredient.Chicken	color(chicken_kebab, golden_red)
$\exists$ preparationMethod.PreparationMethod	recipeType(chicken_kebab, non-vegetarian)
mainIngredient $\sqsubseteq$ ingredient	



# Step9: Evaluation

---

- **Aim:** evaluate the adequacy and efficacy of the ontology for its projected tasks and how well it epitomizes the domain of interest.
- **Methodology:** Manual, i.e., assessed by human users/ experts
- The evaluators were asked to do the following two tasks:
  - **Task 1:** Participants were instructed to enlist questions;
  - **Task 2:** Asked to manually navigate and annotate the concept model displayed on the white board with colored marker pens.

# Step9: Evaluation (contd...2)

---

- **Step 1:** (create a set of questions) Task 1 yielded a set of questions from the participants keeping the particular scenario in mind (i.e., visiting a restaurant).
- **Step 2:** (extraction of key terms) Key terms were extracted manually from the list of questions.
- **Step 3:** (navigate through the ontology) Participants were instructed to use colored marker pen to navigate through the designed ontology to search for the answers to the queries.
- **Step 4:** (analyse the replies) The set of questions were categorized based on the user satisfaction level, i.e. *satisfactory*, *partially satisfactory* and *unsatisfactory*.
  - Satisfactory level is identified based on the **term mapping** and **concept mapping**

# Step9: Evaluation (contd...3)

---

E.g.: (*Step 2*: Key terms were extracted manually from the list of questions.)

Questions	Key Terms
What is the price of the Banana Sundae?	<price, banana sundae>
Is the meat halal or not?	<halal, meat>
Will mushroom pepper dry be spicy?	<mushroom pepper dry, spicy>
What is the time taken to serve the food?	<time, serve>
What is the amount of food served?	<amount, food>
Do you have Chinese food?	<chinese, food>

# Step9: Evaluation (contd...4)

*Step 4:* The set of questions were categorized based on the user satisfaction level, i.e. *satisfactory*, *partially satisfactory* and *unsatisfactory*.

Evaluators	No of queries	Evaluation Parameter		
		Satisfactory	Partially satisfactory	Unsatisfactory
Participant 1	11	10	0	1
Participant 2	10	8	0	2
Participant 3	6	4	2	0
Participant 4	13	11	1	1
Participant 5	10	8	0	2
Participant 6	9	9	0	0
Participant 7	8	7	0	1
Participant 8	18	17	0	1
Participant 9	8	7	0	1
Participant 10	8	6	0	2
Participant 11	6	6	0	0
Participant 12	10	9	0	1
Participant 13	15	15	0	0
Participant 14	14	14	0	0
<b>Total</b>	<b>146</b>	<b>131</b>	<b>3</b>	<b>12</b>

# Result

---

## Before the evaluation

	Domain concepts	No of domain concepts	Terms
Before Evaluation	Entity class	263	358
	Relations	59	73
	Attributes	18	32
	Entity	+ ∞	+ ∞

## After the evaluation

	Domain Concepts	No of domain concepts	Terms
After Evaluation	Entity class	275	380
	Relations	62	76
	Attributes	23	44
	Entity	+ ∞	+ ∞

# Conclusion

---

- Proposed YAMO methodology is scalable
- Provides a step-by-step approach
- Provides a set of guiding principles
  
- Working on various domain ontologies applying the proposed approach
- Applied to the domains food, online recipe and natural disaster

# References

---

- Noy, N.F. and McGuinness, D.L. (2001). Ontology development 101: A guide to creating your first ontology, retrieved from Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.
- Vrandečić, D. and Pinto, S., Tempich, C., Sure, Y. (2005). The DELIGENT knowledge processes. In *Journal of Knowledge Management*, 9(5): 85-96.
- Gruninger, M. and Fox, M. (1995). Methodology for the design and evaluation of ontologies. Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada.
- Uschold, M. and King, M., Moralee, S., Zorgios, Y. (1995). The Enterprise Ontology. In *The Knowledge Engineering Review*, Vol. 13.
- KBSI (1994). The IDEF5 Ontology Description Capture Method Overview. KBSI Report, Texas.
- Fernandez, M. and Gomez-Perez, A., Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering*.
- Giunchiglia, F. and Dutta, B. (2011), DERA: a Faceted Knowledge Organization Framework, available at: <http://eprints.biblio.unitn.it/archive/00002104/> (accessed 20 October 2013).
- Giunchiglia, F. and Dutta, B., Maltese, V. (2014). From Knowledge Organization to Knowledge representation. *Knowledge Organization*, 41(1): 44-56.
- Ranganathan, S.R. (1967). *Prolegomena to library classification*. Asia Publishing House, New York.

---

Thank you for your kind attention!

Question?

Dr. Biswanath Dutta  
Email: [bisu@drtc.isibang.ac.in](mailto:bisu@drtc.isibang.ac.in)