

# All You Need to Know to Build a Product Knowledge Graph

---

**Nasser Zalmout**

Amazon



**Chenwei Zhang**

Amazon



**Xian Li**

Amazon



**Yan Liang**

Amazon



**Xin Luna Dong**

Amazon→Facebook



# Outline

**Overview and Introduction**

20 min



**Knowledge Extraction**

40 min



**Knowledge Cleaning**

25 min



**Break**

20 min

**Ontology Mining**

25 min



**Applications**

20 min



**Conclusion and Future Directions**

10 min



# Overview and Introduction

---

**Overview and Introduction**

**20 min**



Knowledge Extraction

Knowledge Cleaning

Q&A

Break

Ontology Mining

Applications

Conclusion and Future Directions

Q&A

# Product Graph

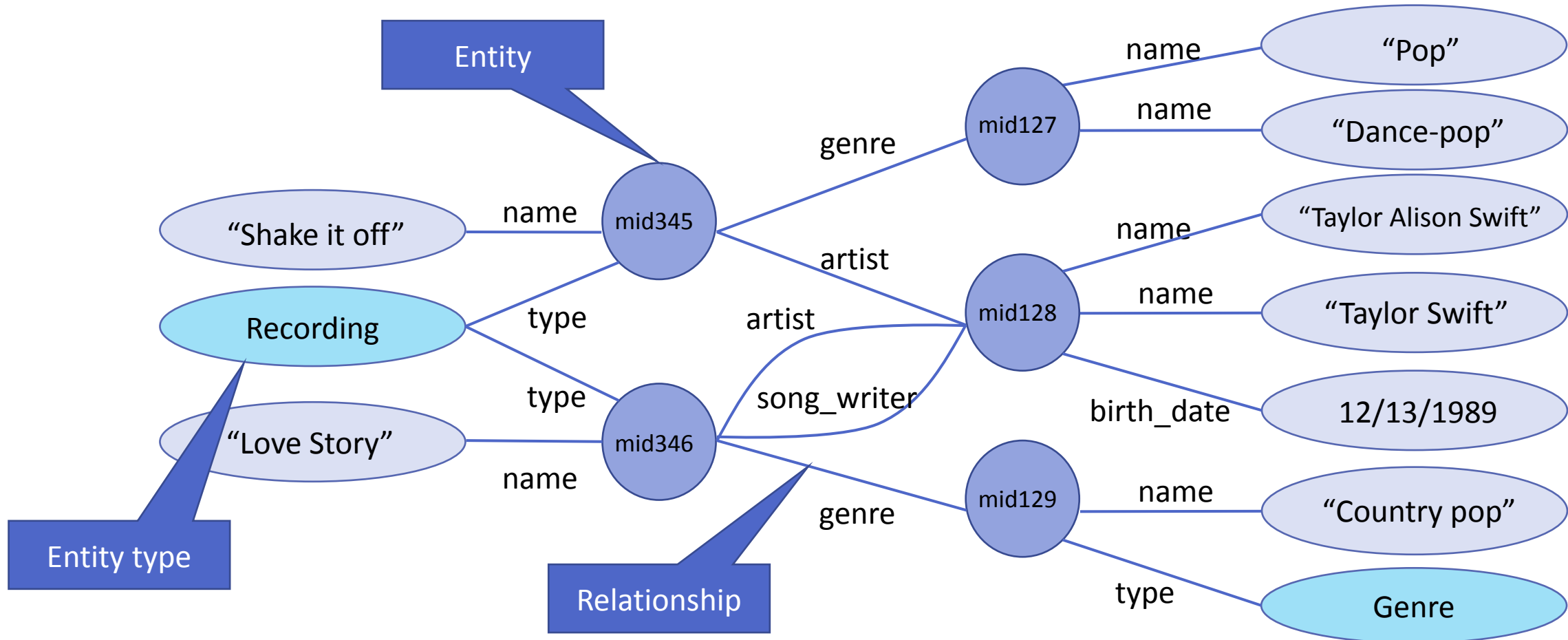
- ❑ Mission: To answer any question about products and related knowledge in the world



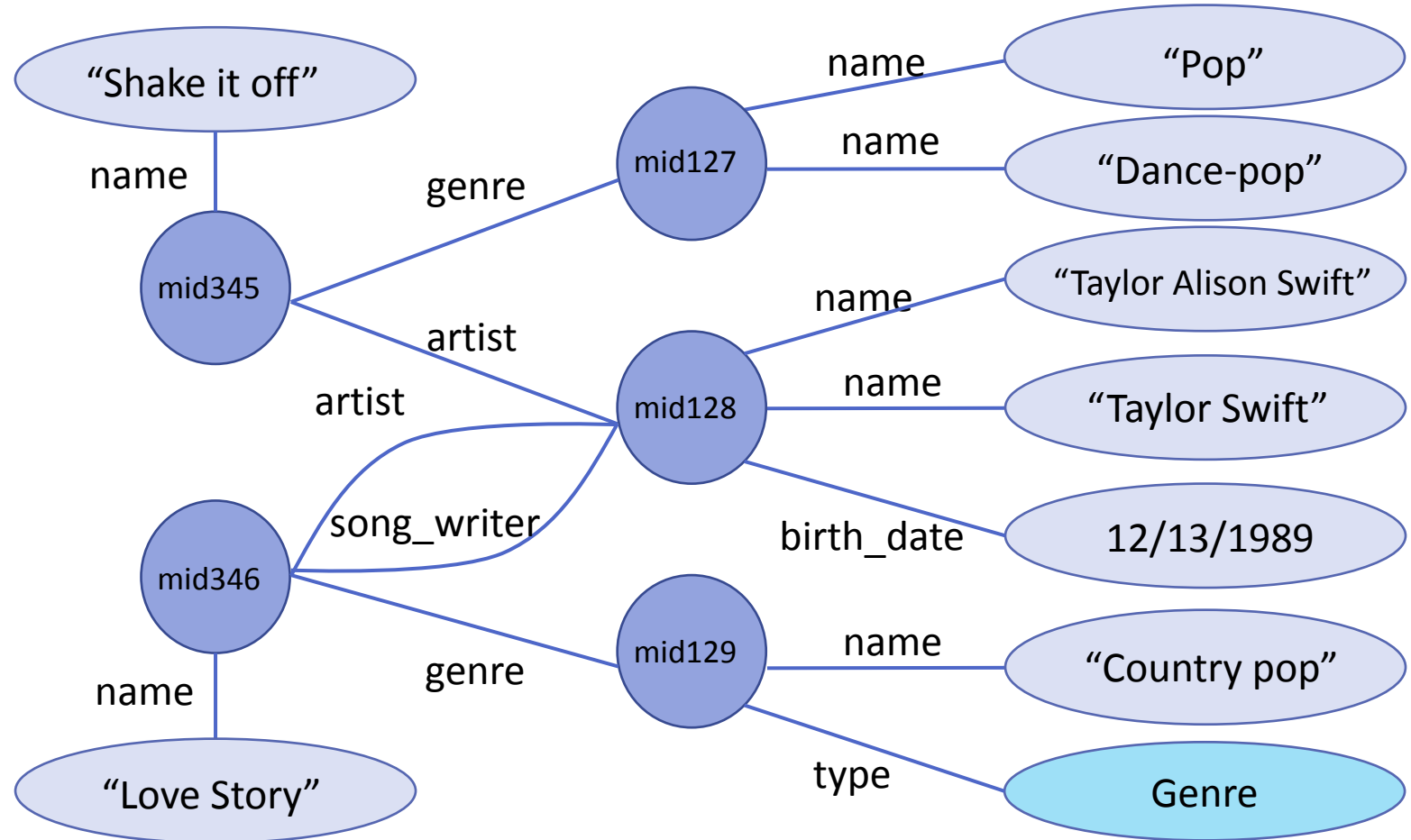
## Customers who bought this item also bought



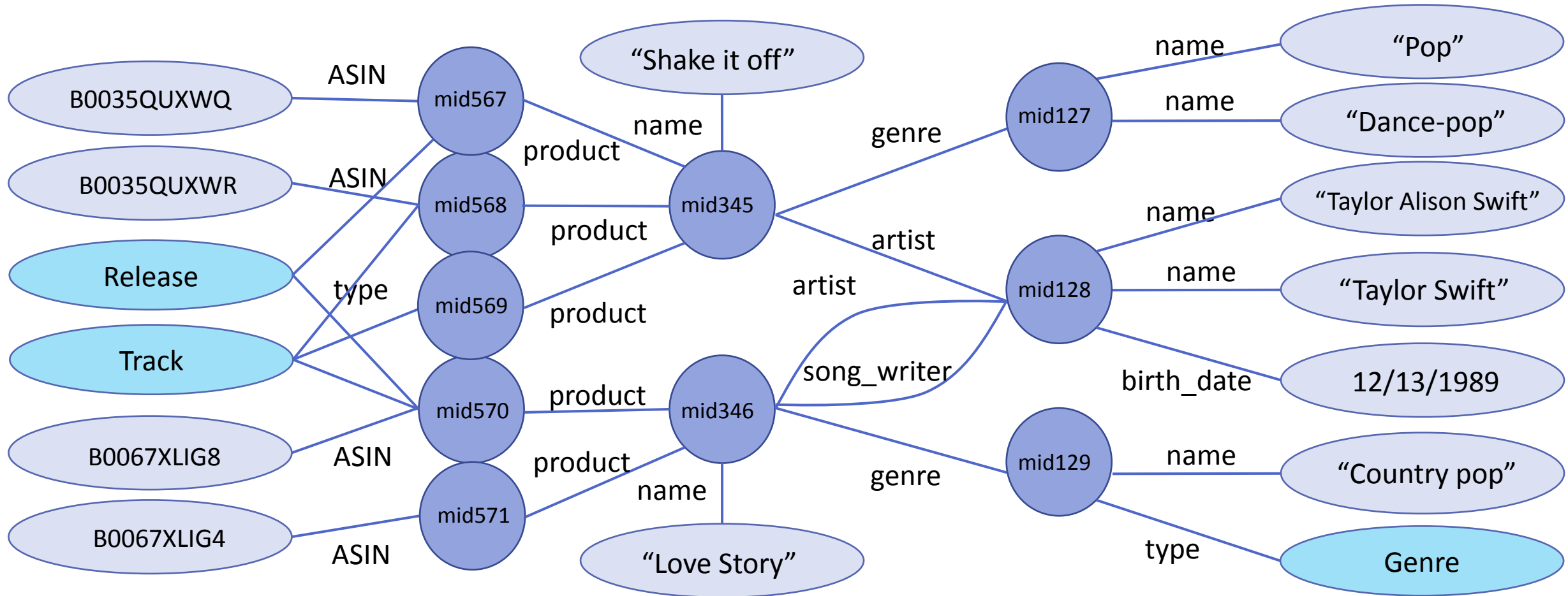
# Knowledge Graph Example for 2 Songs



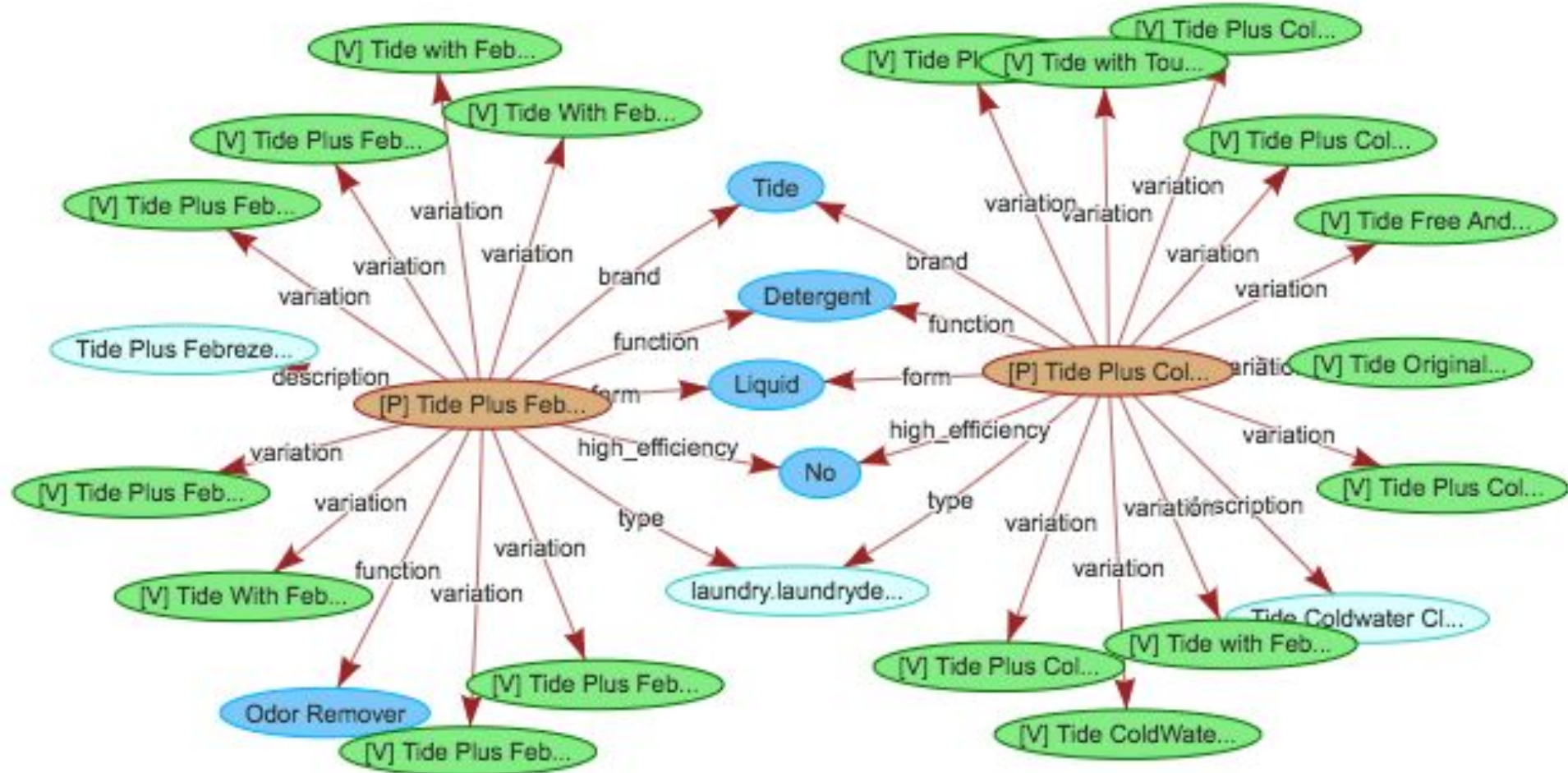
# Product Graph Example for 2 Songs



# Product Graph Example for 2 Songs



# Product Graph Example for 2 Products





# Use Case I: Providing Information



Roll over image to zoom in

<b>Brand</b>	Cetaphil
<b>Ingredients</b>	Water, Cetyl Alcohol, Propylene Glycol, Iodopropynyl Butylcarbamate, 2-Bromo-2-Nitropropane-1, 3-Diol, Sodium Lauryl Sulfate, Stearyl Alcohol, Methylparaben, Propylparaben, Sodium Citrate, Butylparaben, Allantoin, Zinc Gluconate.
<b>Scent</b>	Fragrance free
<b>Additional Information</b>	Non-Comedogenic, Fragrance-free, Natural
<b>Skin Type</b>	Sensitive

## About this item


- Gentle for everyday use; Cetaphil gentle skin cleansing cloths will leave your skin feeling clean, refreshed and balanced after every use
- Removes makeup & dirt: Thoroughly remove makeup and dirt, leaving skin clean
- Mild & non irritating: Soap free formulation won't strip skin of its natural protective oils and emollients

# Use Case II: Providing Choices

Visit the Bigmuscles Nutrition Store ★★★★★ 6,270

Bigmuscles Nutrition Crude Whey 1kg, Whey Protein Concentrate 80%, 24g Protein, 5.5g BCAA, 4 g Glutamine

40% off



Style name:  
**Crude Whey [Unflavored]**

Flavor:  
**Unflavored**

Size:  
**1Kg[2.2lbs]**

₹ 779 (₹ 82.50 / 100 g) Lightning Deal


MRP: ₹ 1,299.00 Save ₹ 520

prime

Visit the Bigmuscles Nutrition Store ★★★★★ 6,270

Bigmuscles Nutrition Crude Whey 1kg, Whey Protein Concentrate 80%, 24g Protein, 5.5g BCAA, 4 g Glutamine

40% off



Style name:  
**Crude Whey [Unflavored]**

Flavor: **Unflavored**


<b>Unflavored</b>	Caffe Latte	Isola
₹ 779.00 prime (₹ 77.90 / 100 g) Deal is 34% Claimed	See available options	See a option

Size:

# Use Case III: Improving Search

All ▾

k-cups dunkin donuts dark



Dunkin' Donuts Original Blend Medium Roast Coffee, 88 K Cups for Keurig Coffee Makers


★★★★☆ ▾ 3,395

**\$38<sup>50</sup>** (\$0.44/Count)

Save 5% more with Subscribe & Save

✓prime FREE Delivery Sun, May 10

88 Count



Dunkin Donuts K-cups Dark Roast - 48 K-cups


★★★★☆ ▾ 112

**\$38<sup>69</sup>** (\$38.69/Count)

✓prime FREE Delivery Fri, May 8

More Buying Choices

\$28.00 (7 new offers)



Dunkin' Donuts Dark K Cup Pods, Dark Roast Coffee, for Keurig Brewers, 60Count


★★★★☆ ▾ 55

**\$35<sup>99</sup>** (\$0.60/Count)

Save 5% more with Subscribe & Save

✓prime FREE Delivery Sun, May 10

60 Count (Pack of 1)



Dunkin Donuts Dunkin Dark Coffee K-Cups For Keurig K Cup Brewers (96 Count) - Packaging May Vary


★★★★☆ ▾ 79

**\$70<sup>57</sup>** (\$0.74/Count)

✓prime FREE Delivery Sat, May 9

More Buying Choices

\$66.95 (8 new offers)



The Original Donut Shop Keurig Single-Serve K-Cup Pods, Regular Medium Roast Coffee, 72 Count

★★★★☆ ▾ 9,914

**\$29<sup>99</sup>** (\$0.42/Count)


Save 5% more with Subscribe & Save

✓prime FREE One-Day Get it Tomorrow, May 5


72 Count

# Use Case III: Improving Search


All ▾ k-cups dunkin donuts dark 🔍




**Dunkin' Donuts Original Blend**  
Medium Roast Coffee, 48 K Cups for Keurig Coffee Makers  
★★★★★ 3,395  
\$38<sup>50</sup> (\$0.44/Count)  
Save 5% more with Subscribe & Save  
✓prime FREE Delivery Sun, May 10  
88 Count




**Dunkin Donuts K-cups Dark Roast - 48 K-cups**  
★★★★★ 112  
\$38<sup>69</sup> (\$38.69/Count)  
✓prime FREE Delivery Fri, May 8  
More Buying Choices  
\$28.00 (7 new offers)



**Dunkin' Donuts Dark K Cup Pods, Dark Roast Coffee, for Keurig Brewers, 60Count**  
★★★★★ 55  
\$35<sup>99</sup> (\$0.60/Count)  
Save 5% more with Subscribe & Save  
✓prime FREE Delivery Sun, May 10  
60 Count (Pack of 1)



**Dunkin Donuts Dunkin Dark Coffee K-Cups For Keurig K Cup Brewers (96 Count) - Packaging May Vary**  
★★★★★ 79  
\$70<sup>57</sup> (\$0.74/Count)  
✓prime FREE Delivery Sat, May 9  
More Buying Choices  
\$66.95 (8 new offers)




**The Original Donut Shop Keurig Single-Serve K-Cup Pods, Regular Medium Roast Coffee, 72 Count**  
★★★★★ 9,914  
\$29<sup>99</sup> (\$0.42/Count)  
Save 5% more with Subscribe & Save  
✓prime FREE One-Day Get It Tomorrow, May 5  
72 Count



# Use Case III: Improving Search

All ▾

k-cups dunkin donuts dark



Dunkin Donuts Dunkin Dark, Dark Roast Coffee K-Cups For Keurig K Cup Brewers (96 Count)

★★★★★ ▾ 18


**\$73<sup>92</sup>** (\$73.92/Count)

✓prime FREE Delivery Tue, May 12

More Buying Choices

\$69.98 (7 new offers)

96 Count



Dunkin Donuts K-cups Dark Roast - 48 K-cups


★★★★★ ▾ 112

**\$38<sup>69</sup>** (\$38.69/Count)

✓prime FREE Delivery Fri, May 8

More Buying Choices

\$28.00 (7 new offers)



Dunkin' Donuts Dark K Cup Pods, Dark Roast Coffee, for Keurig Brewers, 60Count


★★★★★ ▾ 55

**\$35<sup>99</sup>** (\$0.60/Count)

Save 5% more with Subscribe & Save

✓prime FREE Delivery Sun, May 10

60 Count (Pack of 1)



Dunkin Donuts Dunkin Dark Coffee K-Cups For Keurig K Cup Brewers (96 Count) - Packaging May Vary


★★★★★ ▾ 79

**\$70<sup>57</sup>** (\$0.74/Count)

✓prime FREE Delivery Sat, May 9

More Buying Choices

\$66.95 (8 new offers)



Dunkin Donuts K-cups Dark Roast - 24 Kcups for Use in Keurig Coffee Brewers

★★★★★ ▾ 140


**\$20<sup>15</sup>** (\$0.84/Count)

✓prime FREE One-Day Get it Tomorrow, May 5

More Buying Choices

\$13.20 (8 new offers)

# Use Case IV: Improving Recommendation



**KitchenAid KSM150PSE Artisan Tilt-Head Stand Mixer with Pouring Shield, 5-Quart, Empire Red**  
by KitchenAid  
★★★★★ 6,331 customer reviews | 972 answered questions


List Price: \$429.99  
Price: **\$249.99 & FREE Shipping**  
You Save: **\$180.00 (42%)**




*i* Item is eligible for **6 Month Special Financing** with your Amazon.com Store Card. [Learn more](#)

**Note:** Not eligible for Amazon Prime. Available with free Prime shipping from [other sellers on Amazon](#).

**Only 14 left in stock.**  
**Estimated Delivery Date:** July 28 - Aug. 2 when you choose Standard at checkout.  
Ships from and sold by [GE Premier](#) in [easy-to-open packaging](#).

Color: **Empire Red**




Share   

Qty:

**\$249.99** + Frt  
**Only 14 left in sto**  
**Prem**

☐ Include **2-Year I**  
**\$14.99**

 **Add**

[Turn on 1-Click order](#)


**Ship to:**  
KEVIN DAVENPO

[Add to List](#)

[Add to Wedding f](#)

**Similar But Different**


**More Capacity**



[See more choices](#)

[KitchenAid KSM150PSE Certified Refurbished](#)


**More Attachments**



[See Color Options](#)

[KitchenAid KP26M1XER 6](#)


**Different Brand**




[See Color Options](#)

[Hamilton Beach 63232 Electric](#)

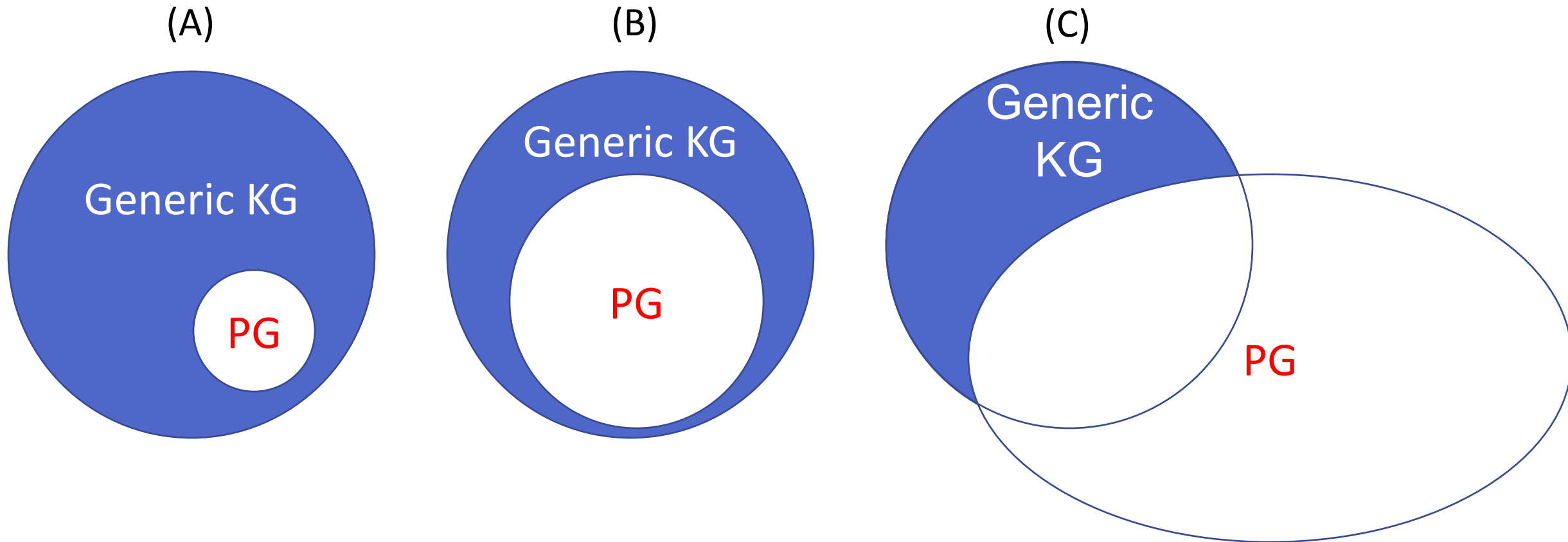
**Other Sellers**

**\$264.99**  
  
Sold by: Amazon.com

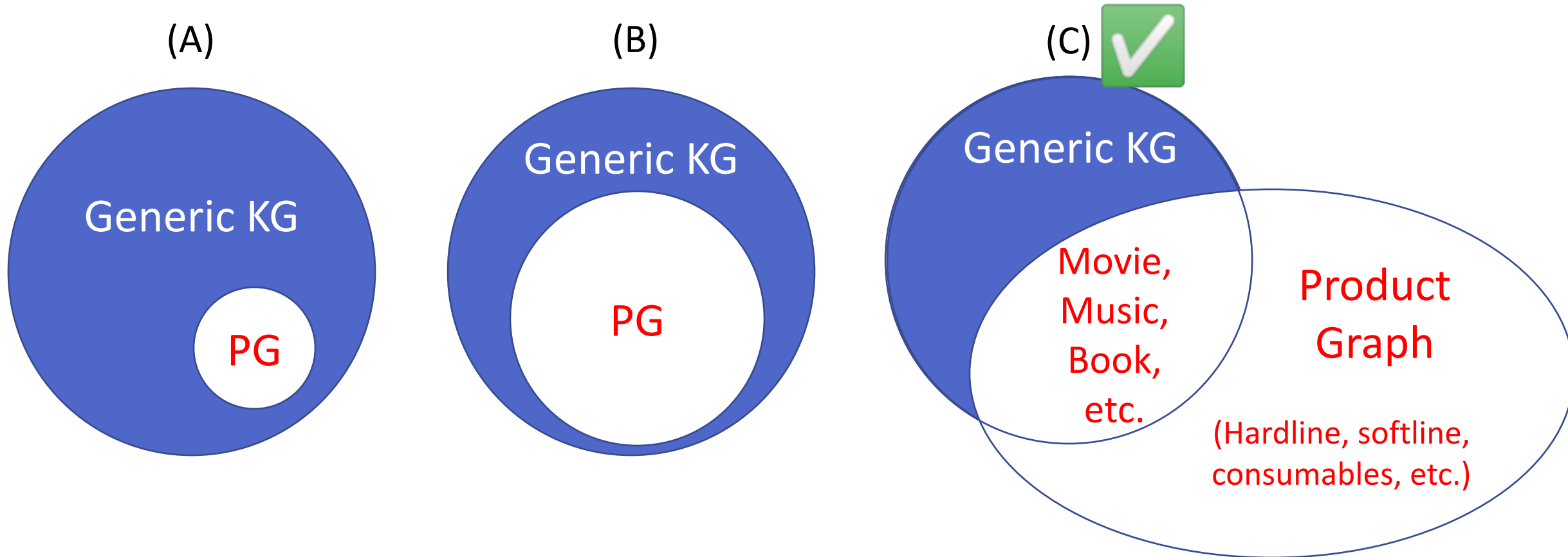
**\$264.99**  
+ Free Shipping  
Sold by: Marcus AV

**\$289.00**  
  
Sold by: goldentech

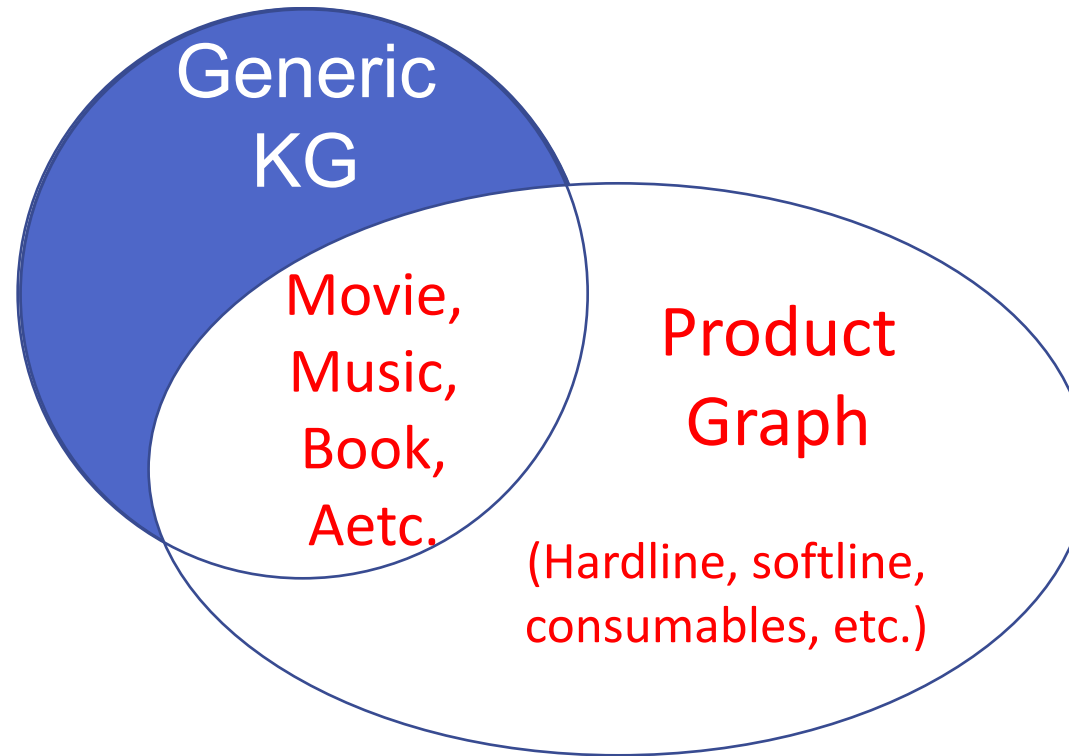
# Product Graph vs. Knowledge Graph



# Product Graph vs. Knowledge Graph





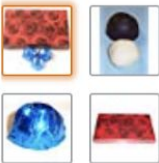


# But, Is The Problem Harder?

# Challenges in Building Product Graph I

## ❑ Sparse and noisy structured data

Scott's Cakes Dark Chocolate Toffee Cream Filling Candies with Dark Blue Foils in a 1 Pound Red Roses Box  
by [Scott's Cakes](#)  
[Be the first to review this item](#)



Price: **\$19.95** + \$14.95 shipping  
You can get 5% back on all Amazon.com purchases with the Amazon Prime Store Card. No annual fee.  
**Note:** Not eligible for Amazon Prime.  
**In Stock.** Ships from and sold by [Scott's Cakes](#).

**38 Flavors:** 1 lb. Red Rose Box

1 lb. Christma... \$19.95	1 lb. Daisy Box \$19.95	1 lb. Red Rose Box \$19.95	▼
------------------------------	----------------------------	-------------------------------	---

**19 Sizes:** Dark Blue Foils

Aqua Foils \$19.95	Burgandy Foils \$19.95	Dark Blue Foils \$19.95	▼
-----------------------	---------------------------	----------------------------	---

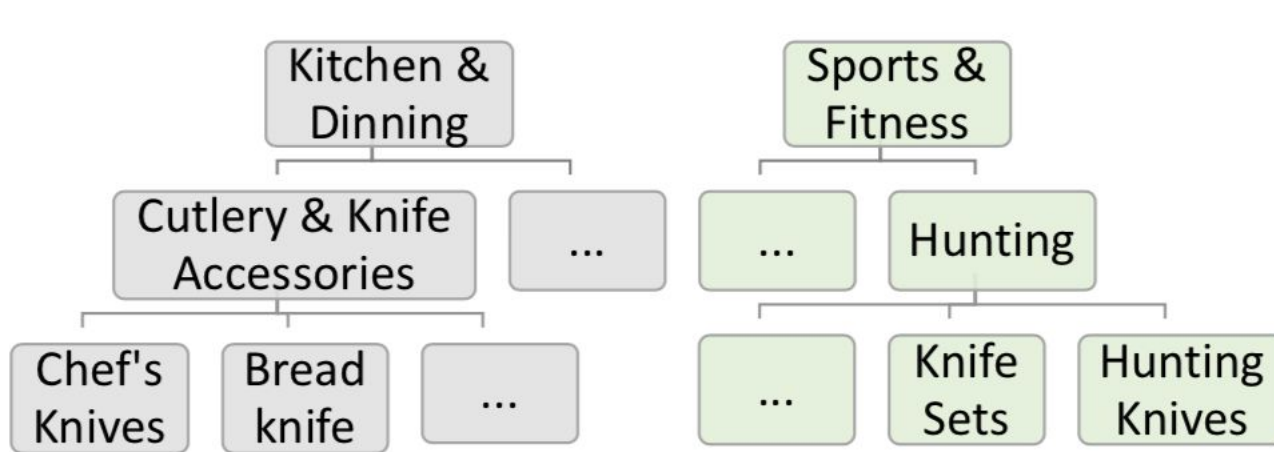
Get it Thu, Aug 22 - Tue, Aug 27  
Get it Tue, Aug 20 - Fri, Aug 23 if you choose paid shipping at checkout.

📍 Deliver to Yaqing - Seattle 98109

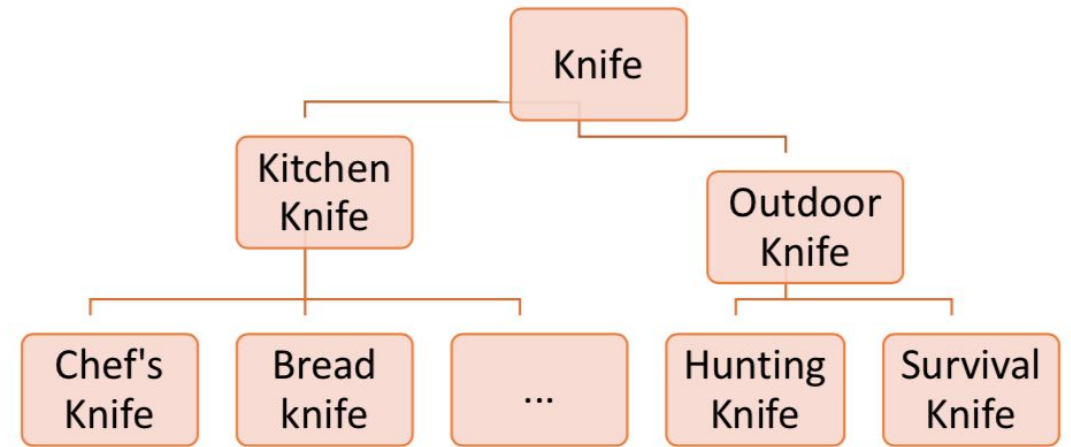
Qty: 1 ▼ [Turn on 1-click ordering](#)

# Challenges in Building Product Graph II

- ❑ Extremely complex domains
- ❑ How to identify the millions of product types?
- ❑ How to organize types into a taxonomy tree?



Sellers' view



Buyers' view

# Challenges in Building Product Graph III

- ❑ Big variety across product types
- ❑ Different attributes apply to different product types
- ❑ Different value vocabularies and different patterns



Cute Rascals

Cute Rascals Blue Red Cranberry Taped Neck Girl Ruffle Cotton Long  
Sleeve Sunflower Dress

[Be the first to review this item](#)

Price: \$30.99

Color: Soft Pink



Flavor

Ingredient

# Challenges in Building Product Graph III

- ❑ Big variety across product types
- ❑ Different attributes apply to different product types
- ❑ Different value vocabularies and different patterns



Samsung UN58RU7100FXZA Flat **58-Inch**  
4K UHD 7 Series Ultra HD Smart TV with  
HDR and Alexa Compatibility (2019 Model)

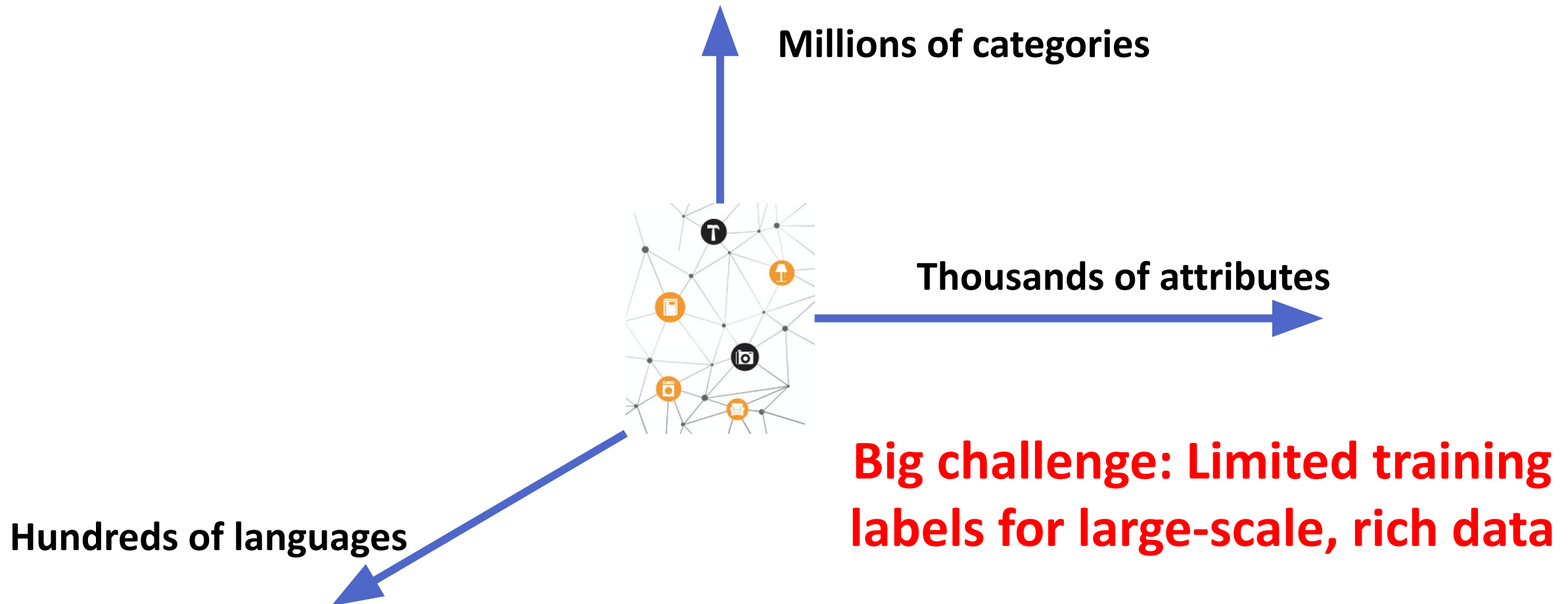


Taylors of Harrogate Classic Tea Variety Box,  
**48 Count (Pack of 1)**



Caribou Coffee Caribou Blend, Medium  
Roast Ground Coffee, **20 Ounce** Bag,  
Rainforest Alliance Certified

# Scale Up in 3 Dimensions

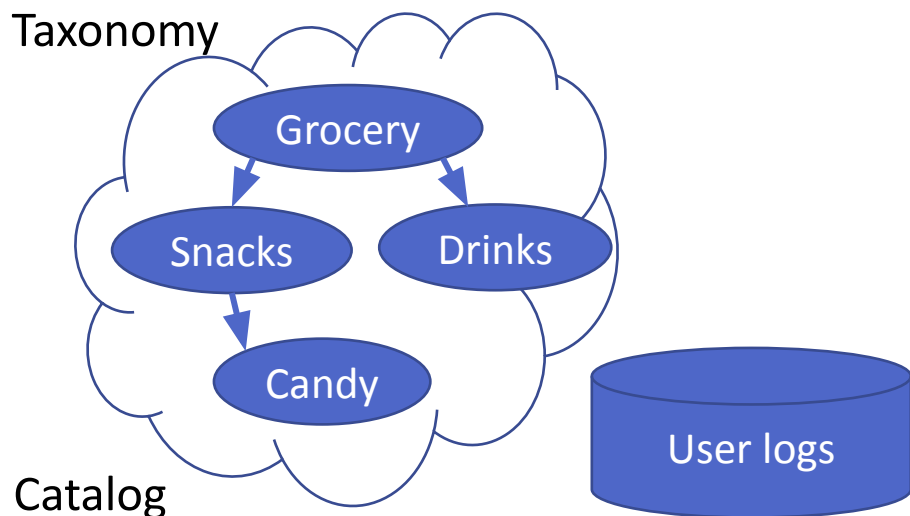






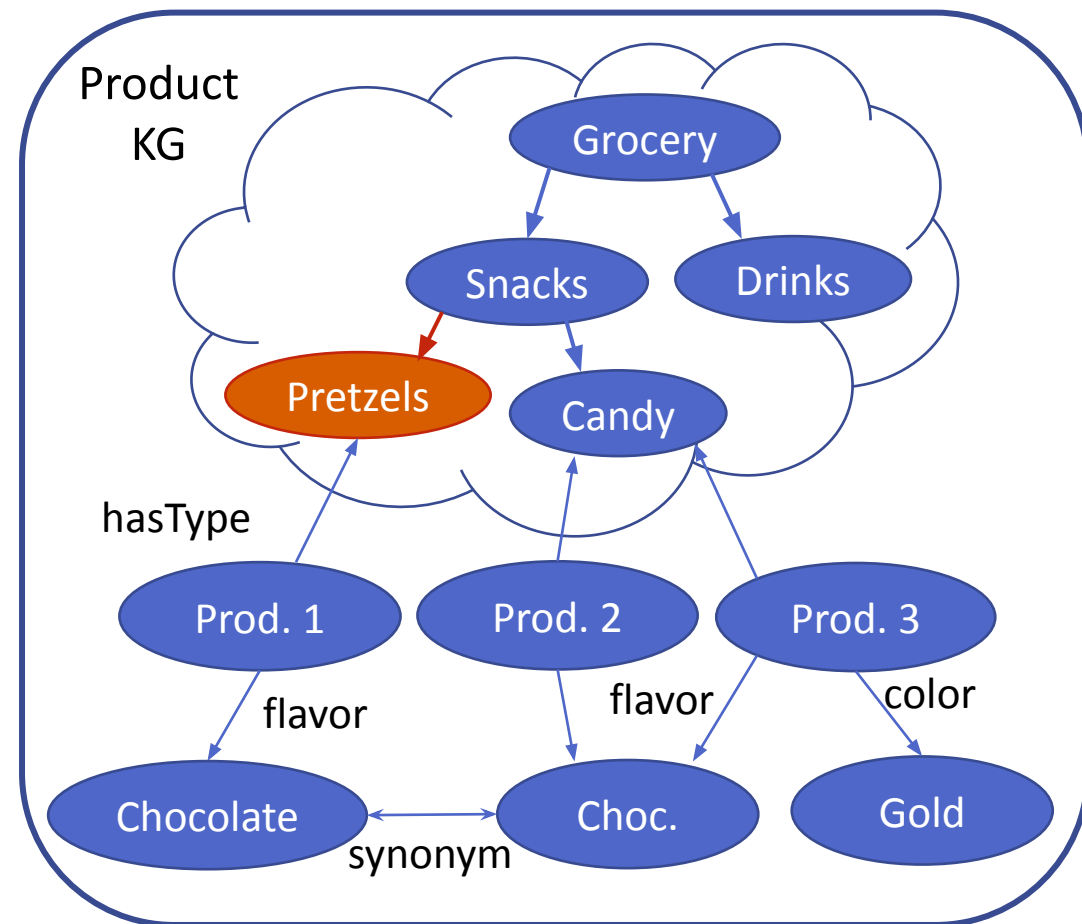
Can We Build A **Self-Driving**  
Product Understanding System?

# Our Goal: Self-Driving Product Knowledge Collection



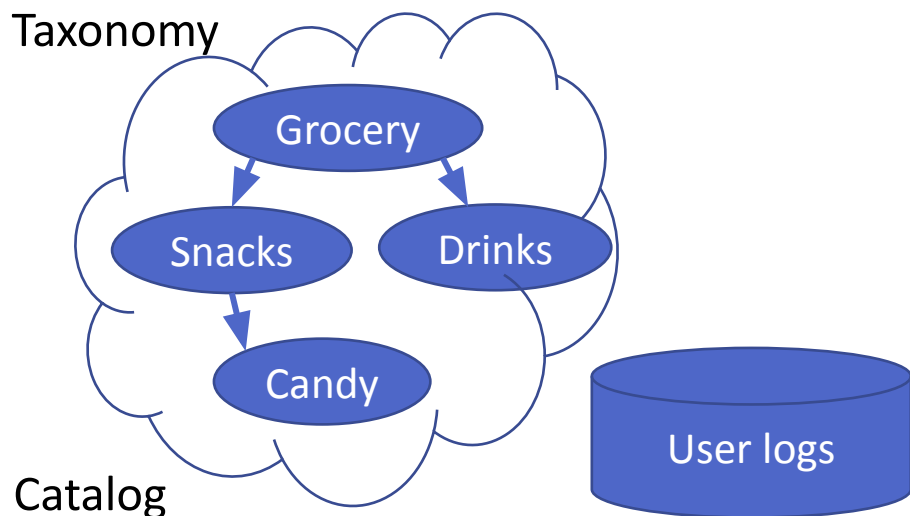
Product	Type	Flavor	Color
Product 1	Snacks	Cherry	
Product 2	Candy	?	?
Product 3	Candy	Choc.	Gold

AutoKnow

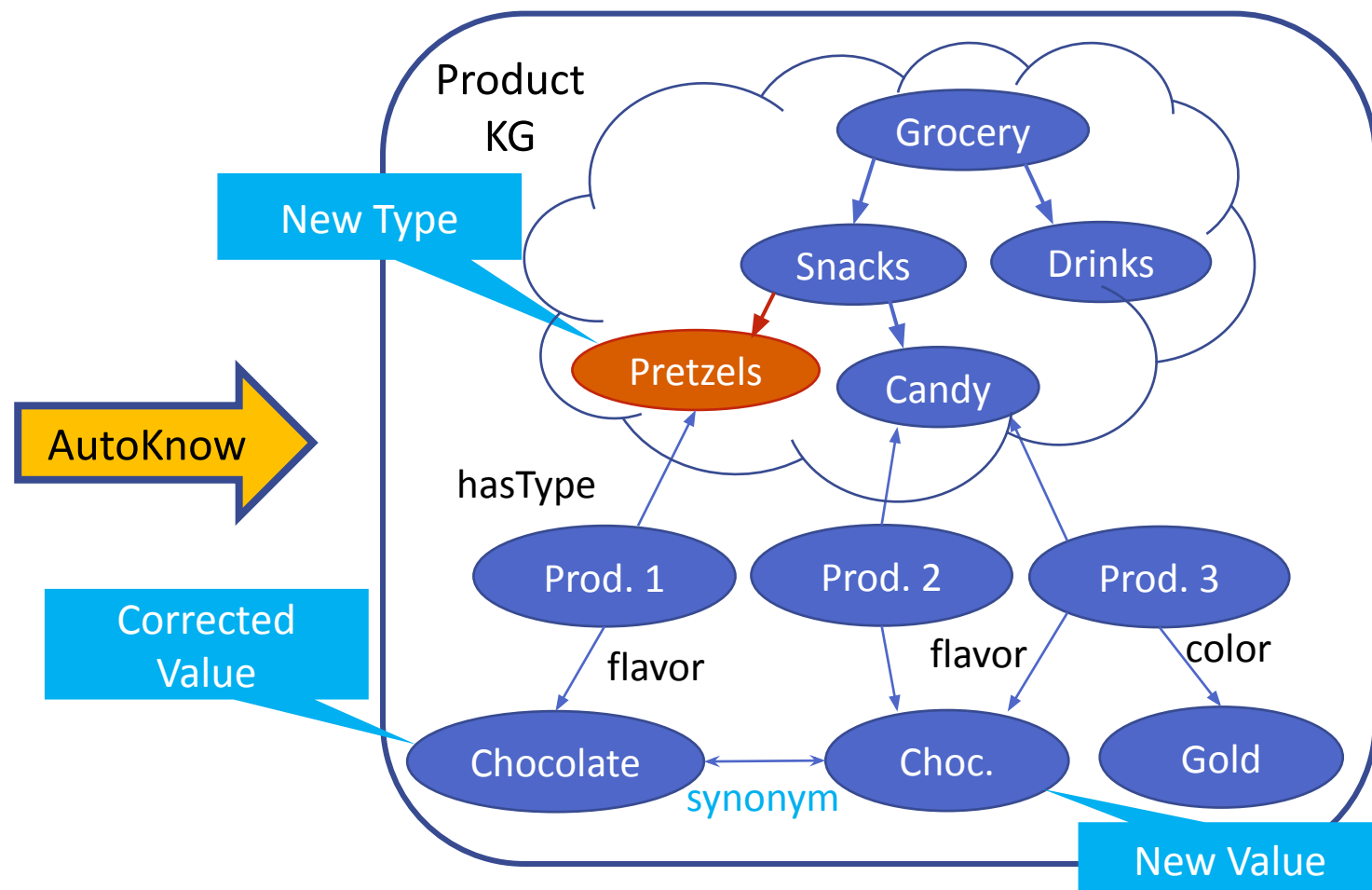




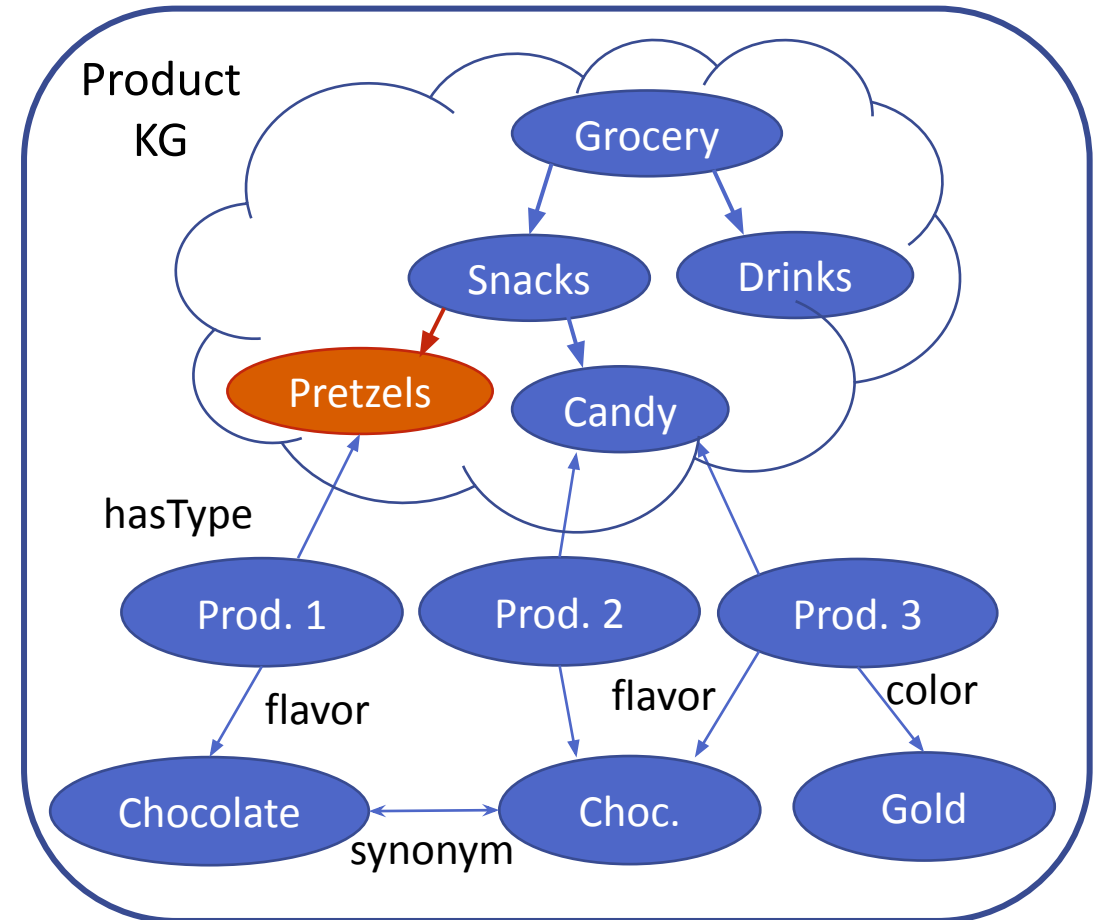
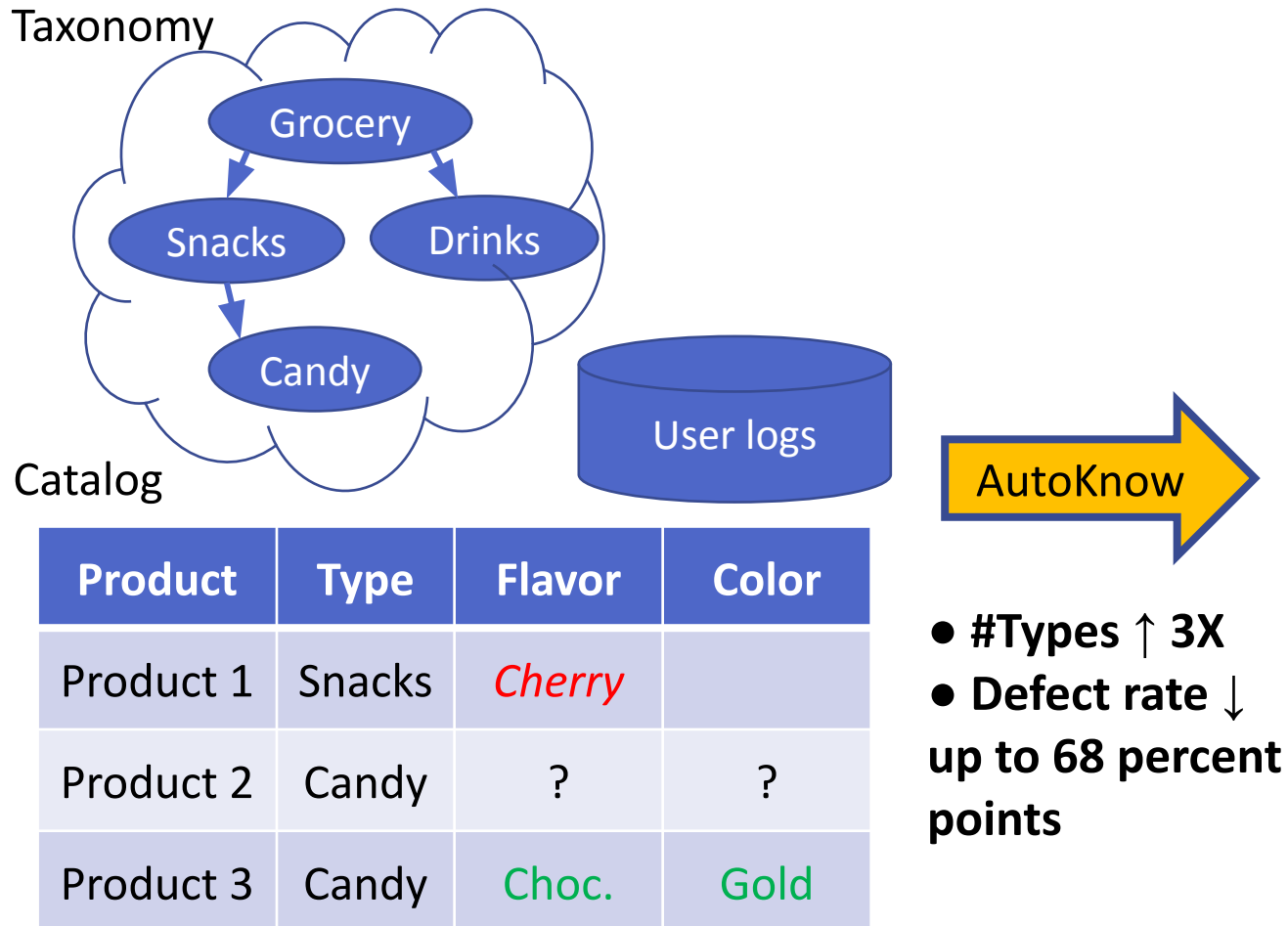
# Our Goal: Self-Driving Product Knowledge Collection



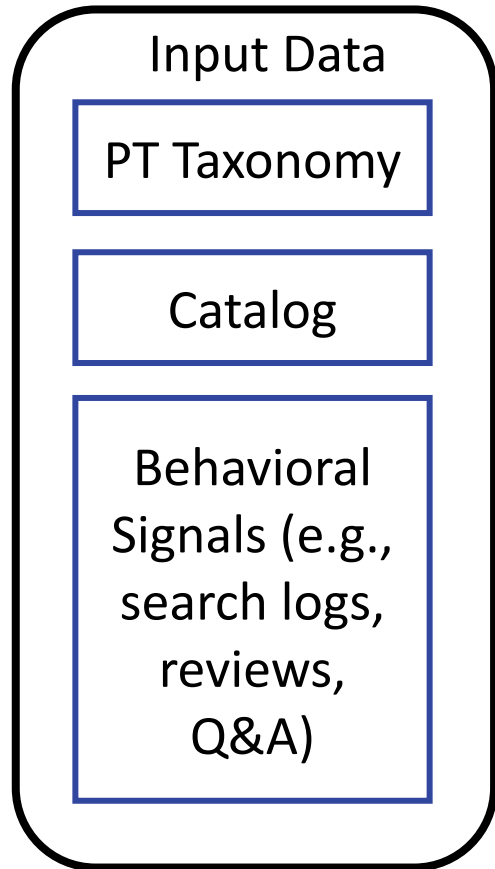
Product	Type	Flavor	Color
Product 1	Snacks	Cherry	
Product 2	Candy	?	?
Product 3	Candy	Choc.	Gold



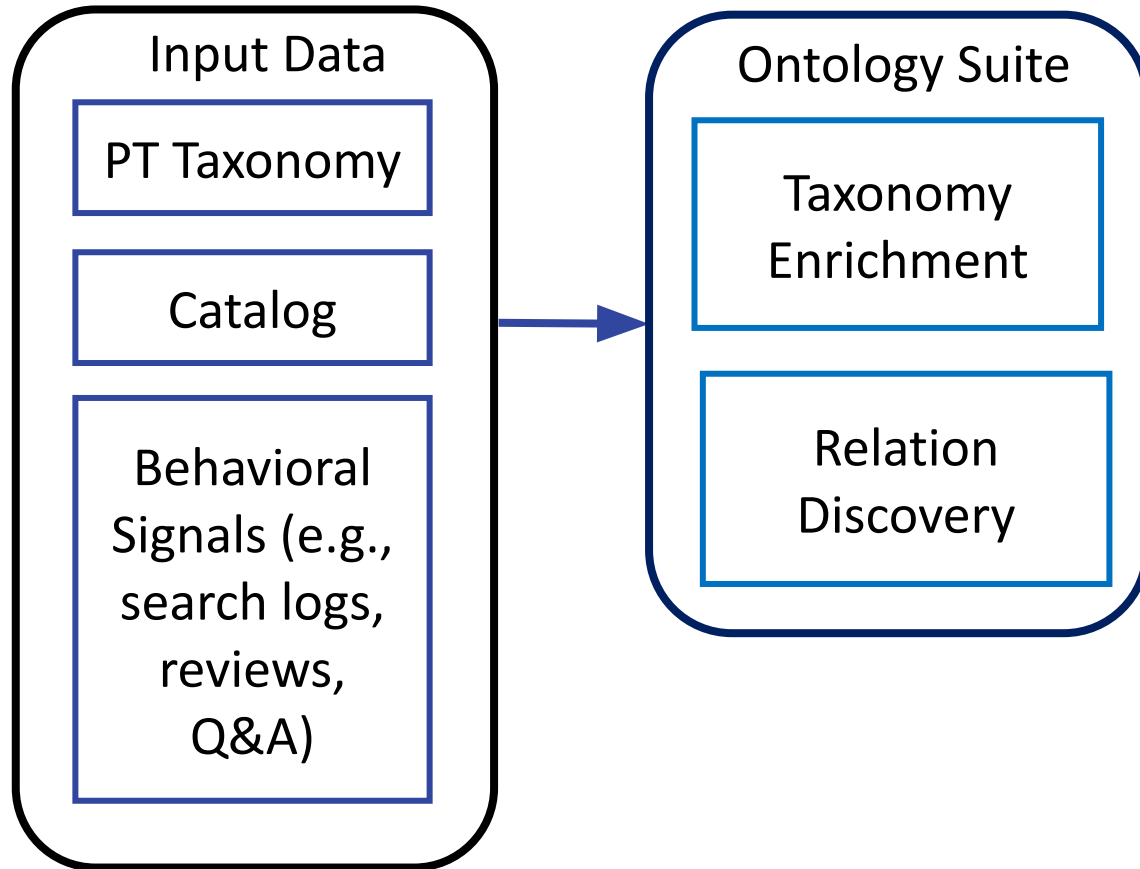
# Amazon AutoKnow: Self-Driving Product Knowledge Collection



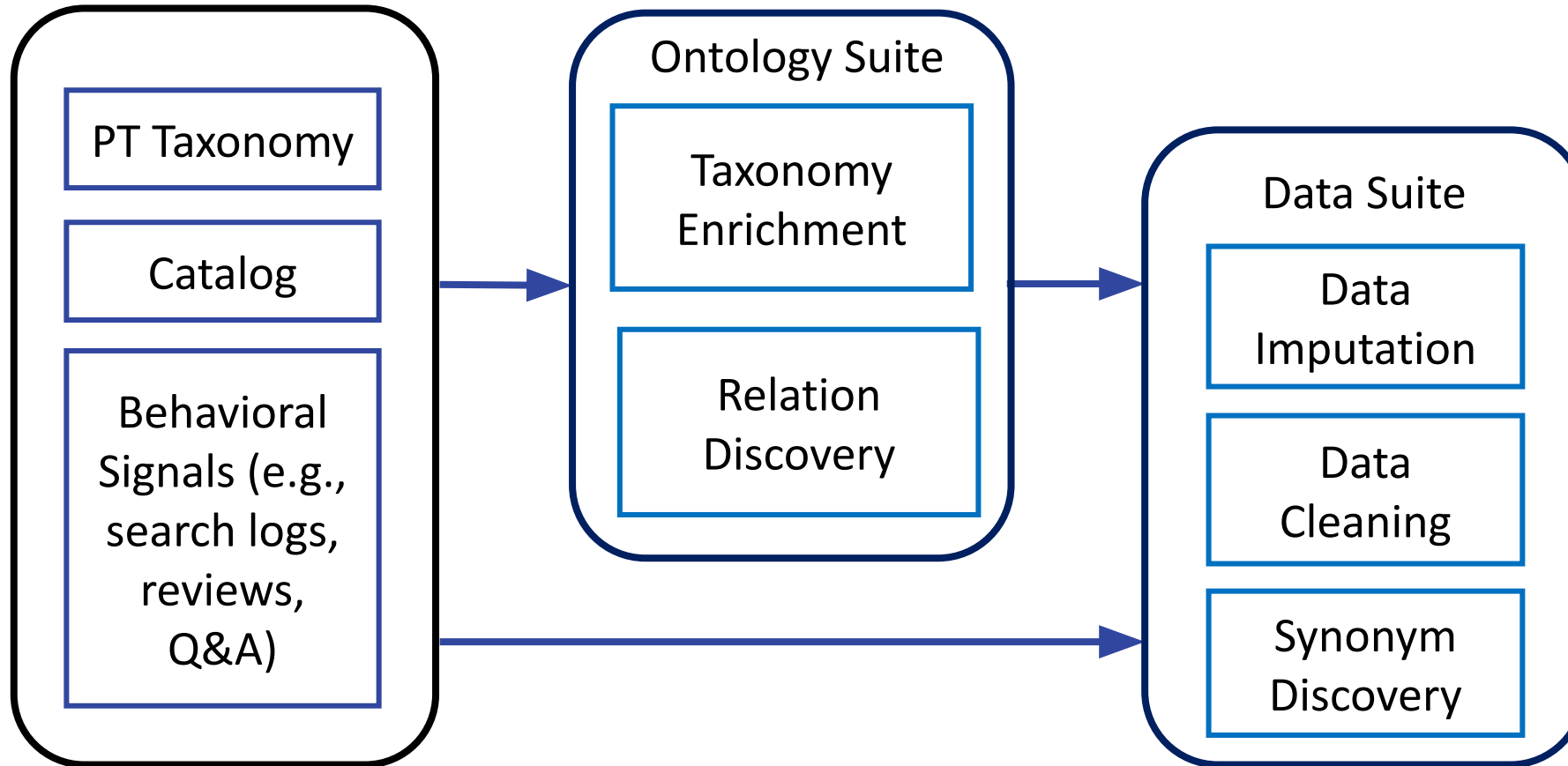
# Amazon AutoKnow: Self-Driving Product Knowledge Collection



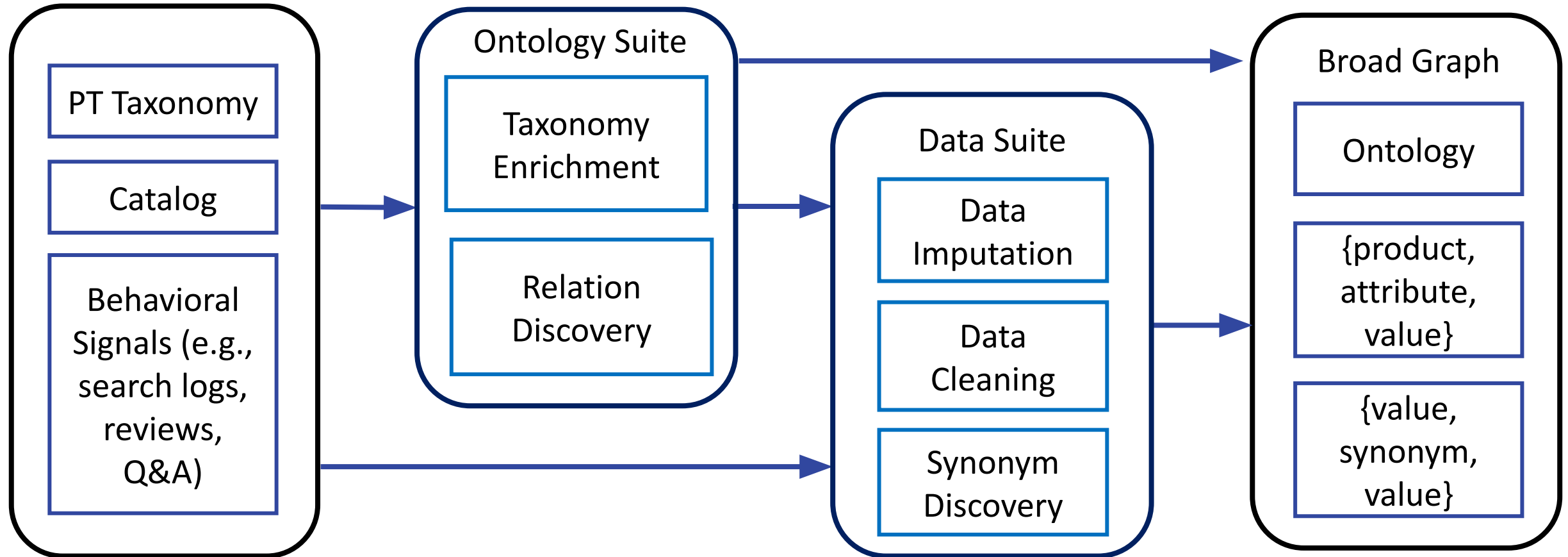
# Amazon AutoKnow: Self-Driving Product Knowledge Collection



# Amazon AutoKnow: Self-Driving Product Knowledge Collection



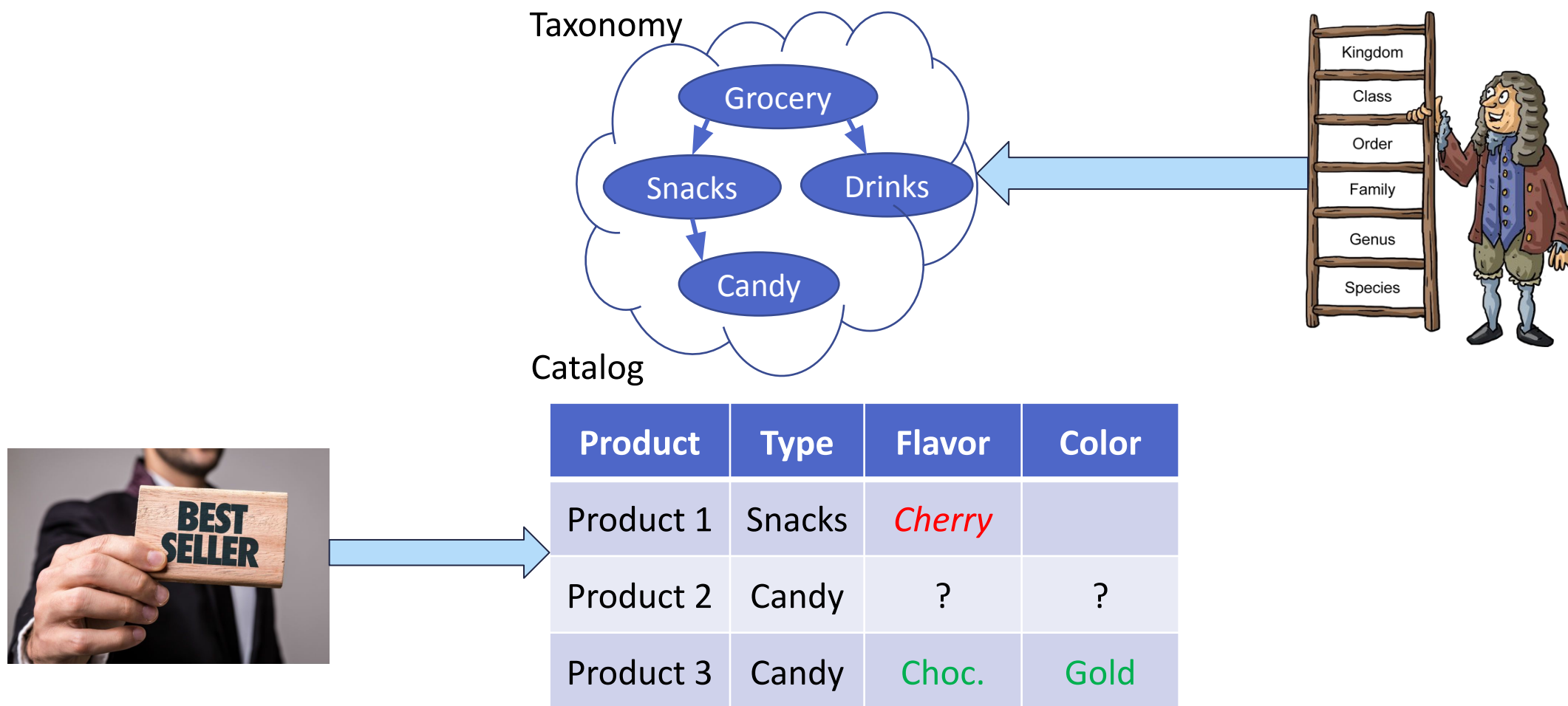
# Amazon AutoKnow: Self-Driving Product Knowledge Collection



# Self Driving to Navigate a Large Space

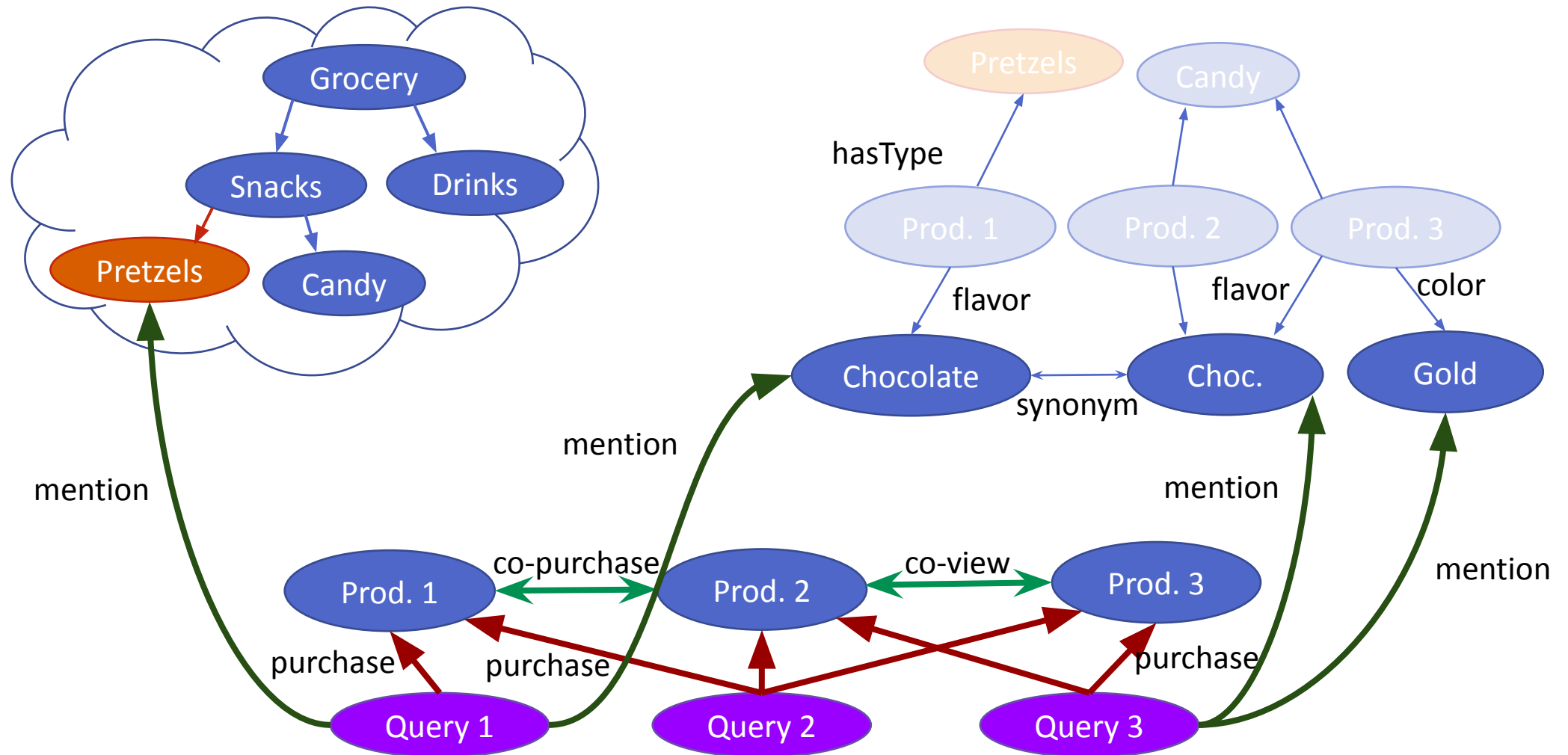
- **Automatic:** Fully ML-based
- **Annotation free:** Weak learning based on existing Catalog data and user behavior
- **One-size-fits-all:** Few taxonomy-aware models
- **Self guidance:** Identify important attributes and categories to focus efforts

# Key Intuition I. Learning w. Limited Labels Generated from Existing Catalog Data

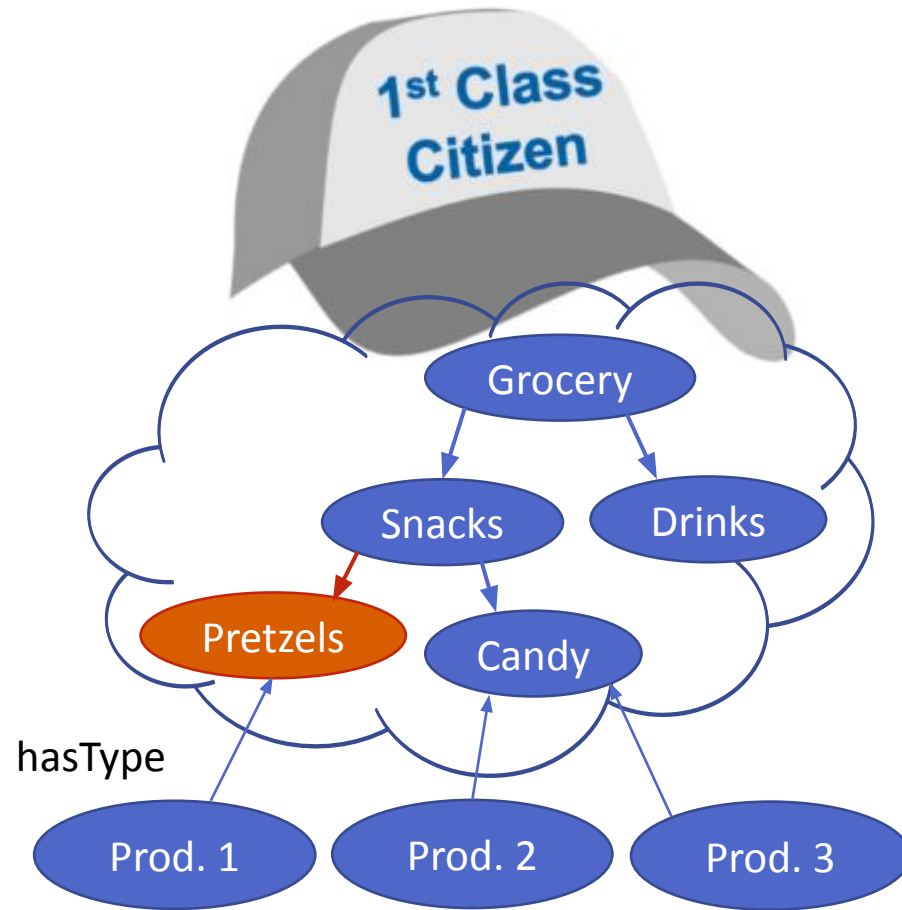




# Key Intuition II. Rich Customer Behavior

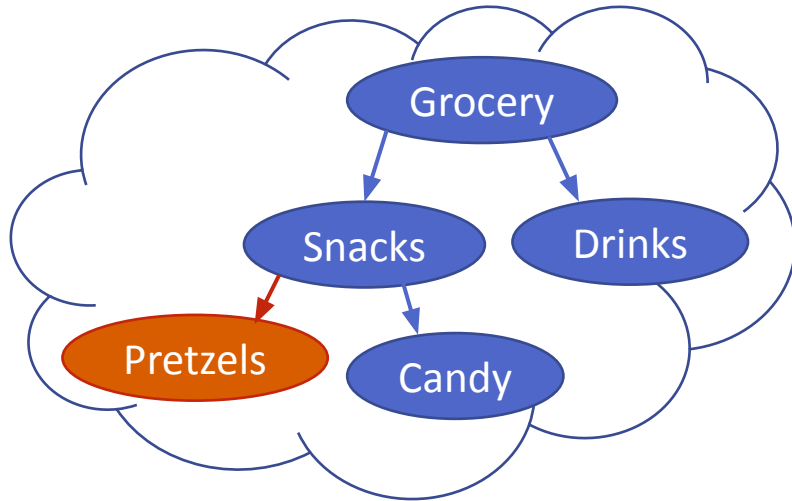


# Key Intuition III. Product Categories as First-Class Citizen in Modeling

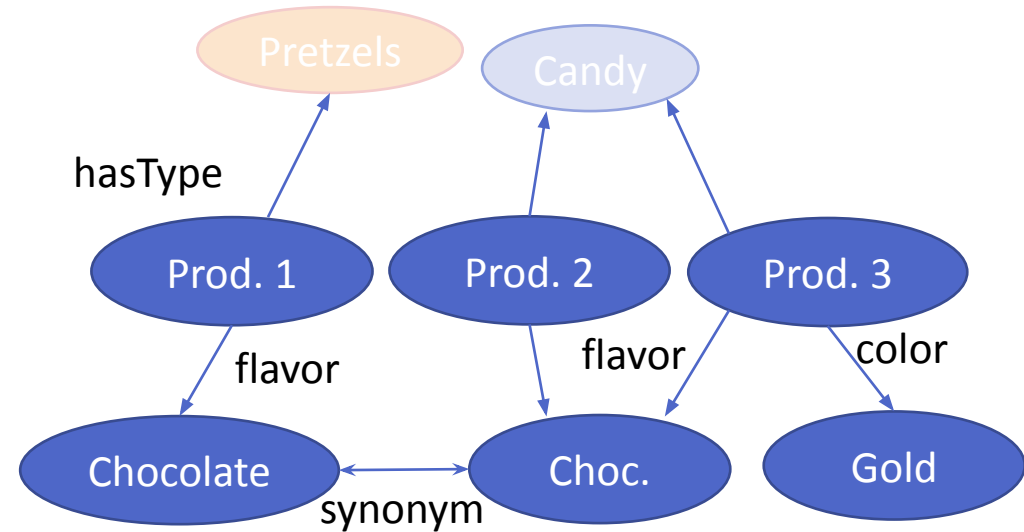


# Key Intuition IV. Leverage Graph Structure

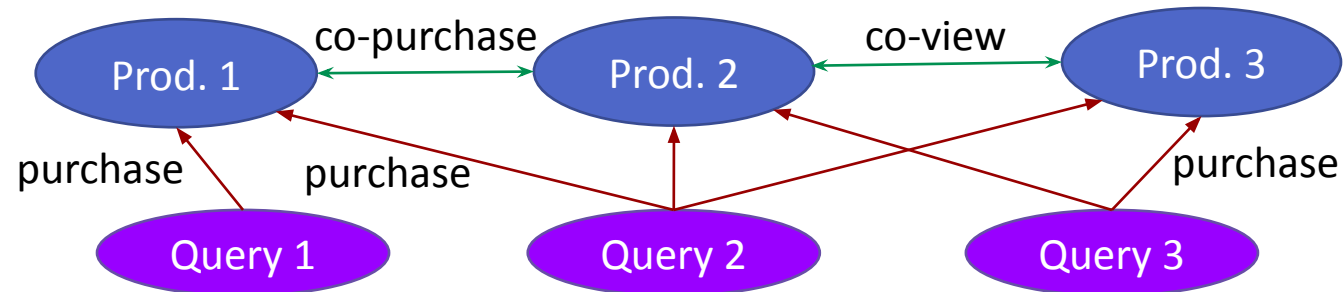
**Taxonomy is a tree**



**KG is a graph**



**Customer behavior forms a graph**



# Key Techniques in AutoKnow

Techniques \ Component	<i>AK-Taxonomy</i>	<i>AK-Relations</i>			<i>AK-Imputation</i>	<i>AK-Cleaning</i>	<i>AK-Synonyms</i>
Graph structure	X				X		
Taxonomy signal					X	X	
Distant supervision	X				X	X	
Behavior information	X	X					X

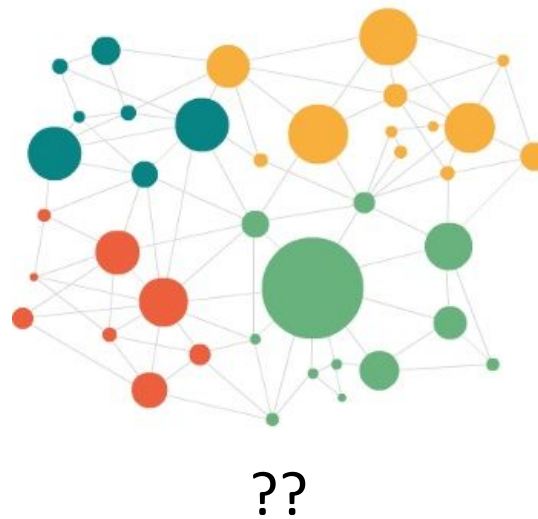
# Which ML Model Works Best?



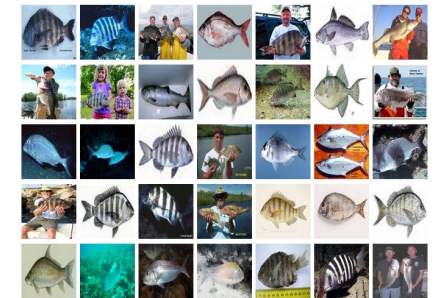
# Which ML Model Works Best?

ID	NAME	CLASS	MARK	SEX
1	John Deo	Four	75	female
2	Max Ruin	Three	85	male
3	Arnold	Three	55	male
4	Krish Star	Four	60	female
5	John Mike	Four	60	female
6	Alex John	Four	55	male
7	My John Rob	Fifth	78	male
8	Asruid	Five	85	male
9	Tes Qry	Six	78	male
10	Big John	Four	55	female

Tree-based models



SCENE FROM "DAN'L DRUCE."  
This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, about two months ago. Our readers will probably not need to be reminded that the hero of the story, Dan'l Druce, the blacksmith, is a solitary recluse dwelling on the coast of Norfolk, where his lone cottage is visited by fugitives from party vengeance during the civil wars of the Commonwealth. His hoard of money is stolen; but a different sort of treasure, a helpless female infant, is left by some mysterious agency, and may be accepted, as in George Eliot's tale of "Silas Marner," for a Divine gift to the sad-hearted misanthrope, far better than riches. In this spirit, at least, he is content to receive the precious human charge; and so to those who would remove it from his home, Dan'l Druce here makes answer with the solemn exclamation, "Touch not the Lord's gift!" This character is well acted by Mr. Hermann Vezin.



Neural network



# Which ML Model Works Best?

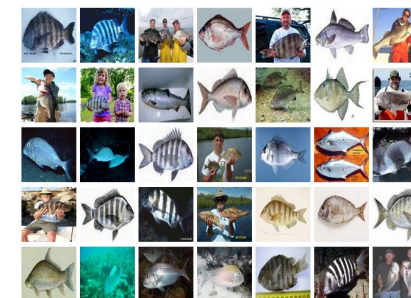
ID	NAME	CLASS	MARK	SEX
1	John Deo	Four	75	female
2	Max Ruin	Three	85	male
3	Arnold	Three	55	male
4	Krish Star	Four	60	female
5	John Mike	Four	60	female
6	Alex John	Four	55	male
7	My John Rob	Fifth	78	male
8	Asruid	Five	85	male
9	Tes Qry	Six	78	male
10	Big John	Four	55	female

Tree-based models



Knowledge  
embedding,  
GNN, etc.

SCENE FROM "DAN'L DRUCE."  
This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, about two months ago. Our readers will probably not need to be reminded that the hero of the story, Dan'l Druce, the blacksmith, is a solitary recluse dwelling on the coast of Norfolk, where his lone cottage is visited by fugitives from party vengeance during the civil wars of the Commonwealth. His hoard of money is stolen; but a different sort of treasure, a helpless female infant, is left by some mysterious agency, and may be accepted, as in George Eliot's tale of "Silas Marner," for a Divine gift to the sad-hearted misanthrope, far better than riches. In this spirit, at least, he is content to receive the precious human charge; and so to those who would remove it from his home, Dan'l Druce here makes answer with the solemn exclamation, "Touch not the Lord's gift!" This character is well acted by Mr. Hermann Vezin.



Neural network

# Deliver the Data Business

1,000,000,000



# Deliver the Data Business

1

High  
precision  
models

# Deliver the Data Business

1,000

High  
precision  
models

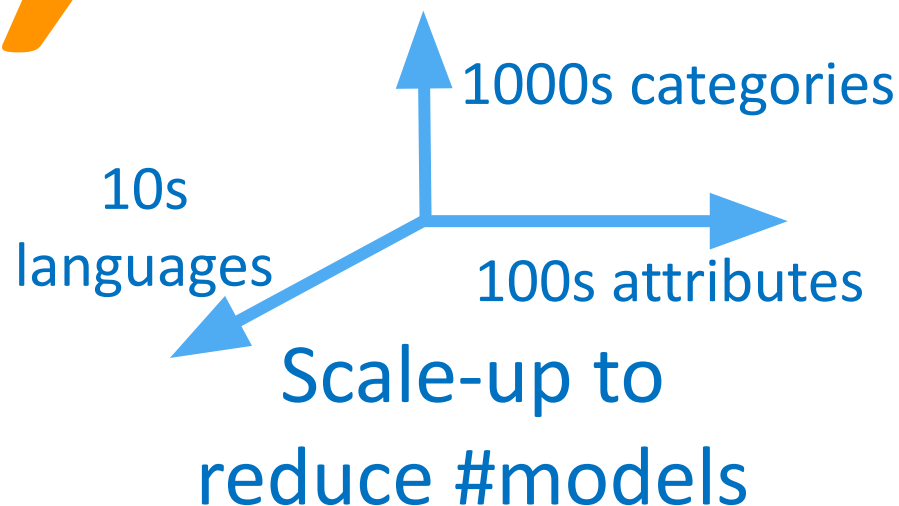
E2E pipeline  
+ AutoML  
to reduce  
modeling cost

# Deliver the Data Business

1,000,000

High  
precision  
models

E2E pipeline  
+ AutoML  
to reduce  
modeling cost

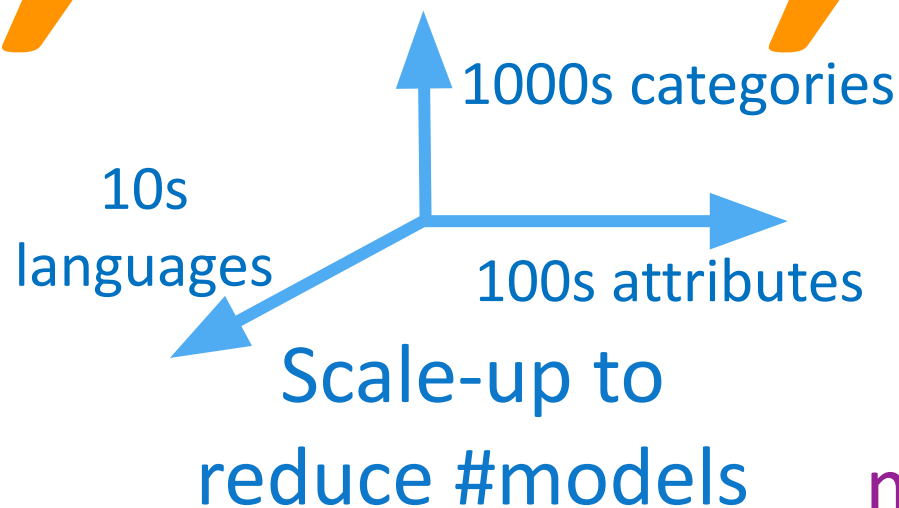


# Deliver the Data Business

1,000,000,000

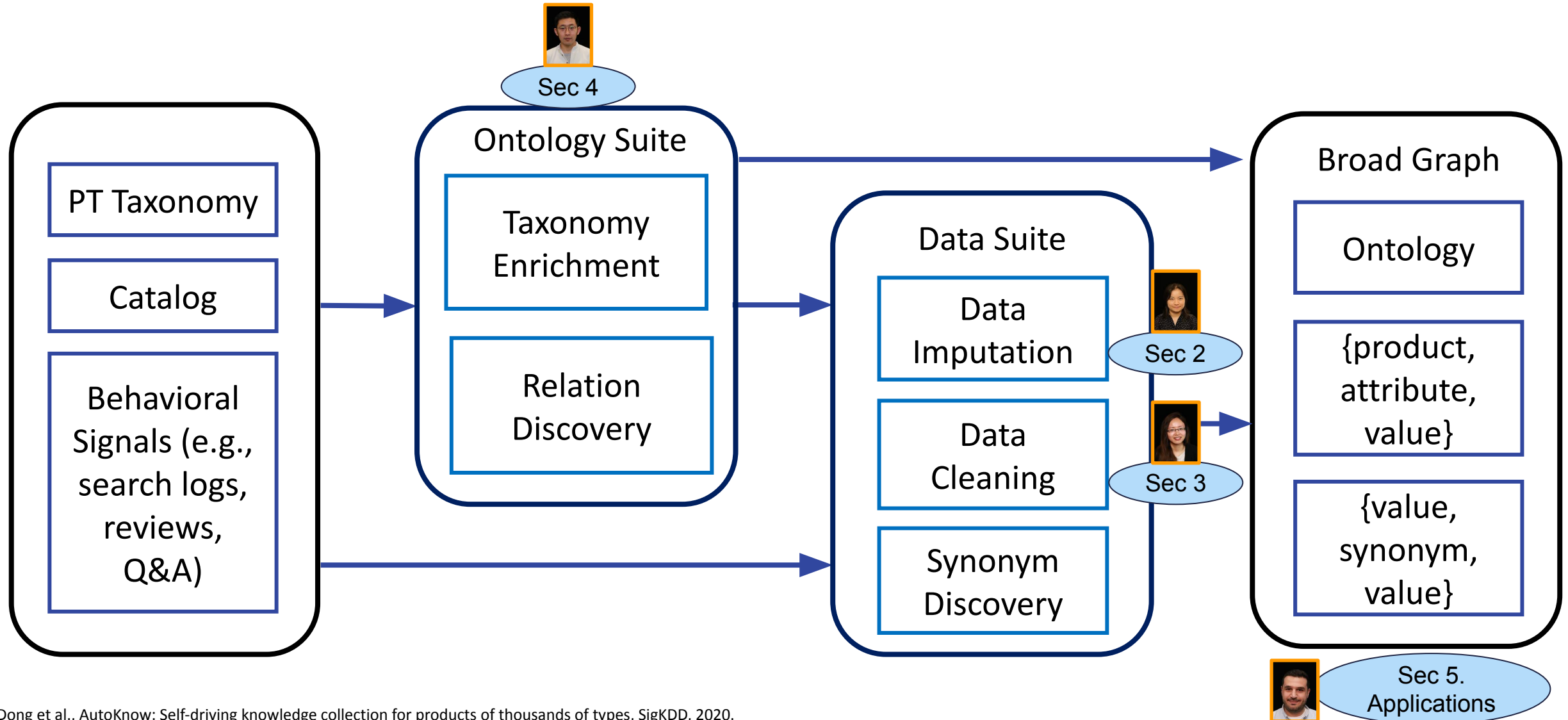
High  
precision  
models

E2E pipeline  
+ AutoML  
to reduce  
modeling cost



Higher yield from  
multi-modal models

# Tutorial Structure



# Section Structure

- Problem Definition

*What is needed beyond techniques for building generic KGs?*

- Short answer -- key intuition

*What are key intuitions for building product KGs?*

- Long answer -- details

*What are practical tips?*

- Reflection/short-answer

*Can we apply the techniques to other domains?*

# Key Questions We Answer in This Tutorial

- Q1. What are unique challenges to build a product knowledge graph and what are solutions?
- Q2. Are these techniques applicable to building other domain knowledge graphs?
- Q3. What are practical tips to make this to production?