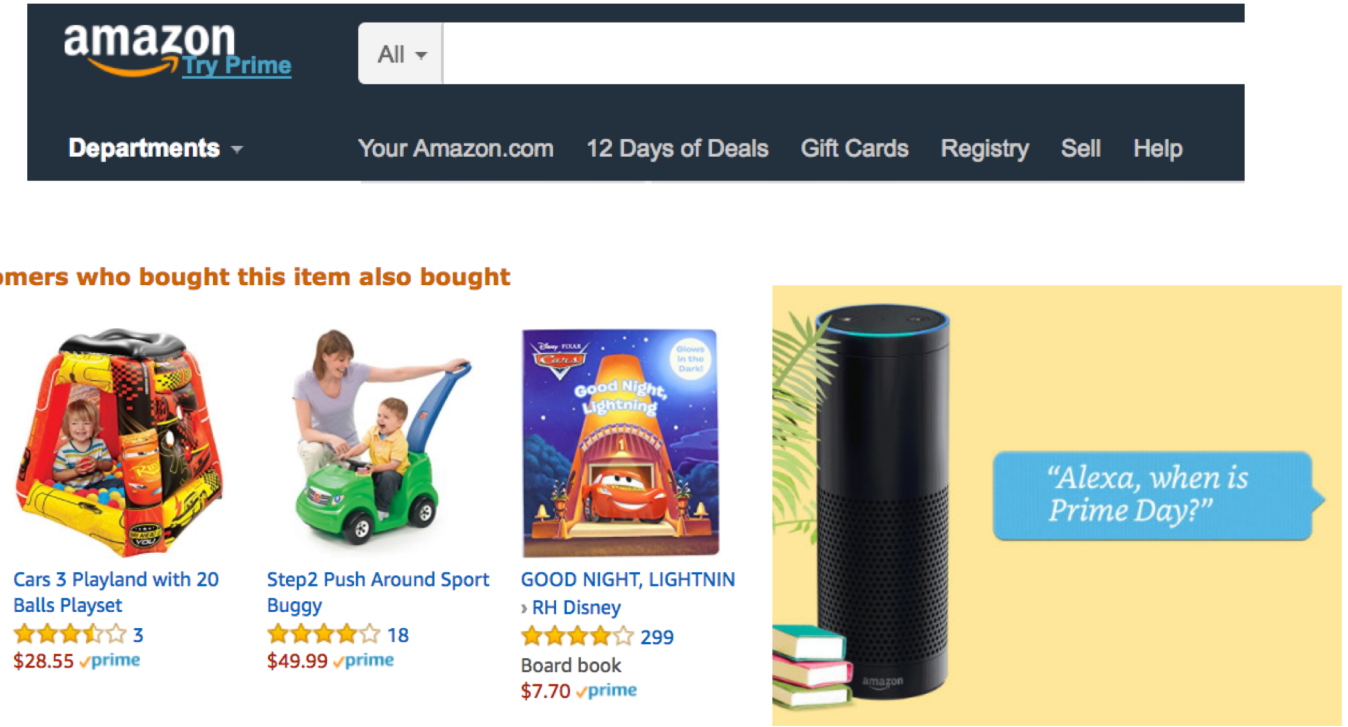
Self-Driving Product Understanding for Thousands of Categories

XIN LUNA DONG, AMAZON

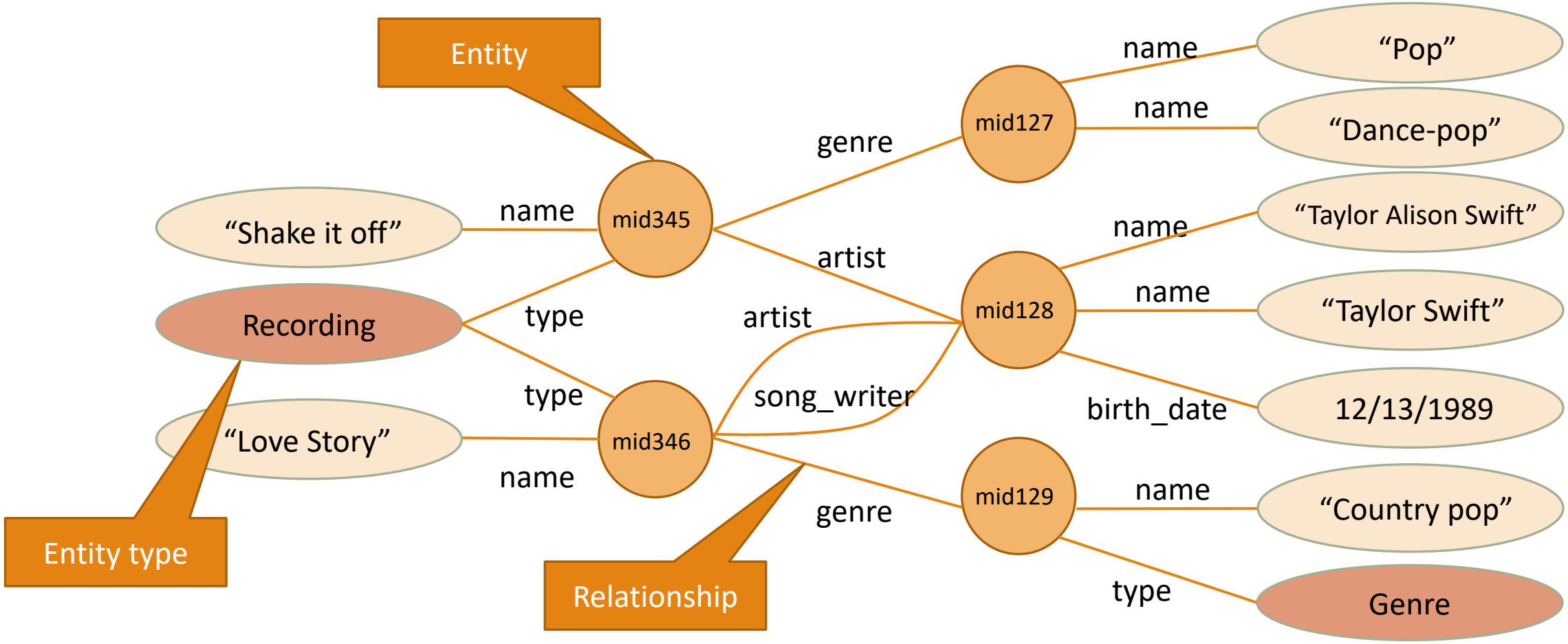
8/2020

Product Graph

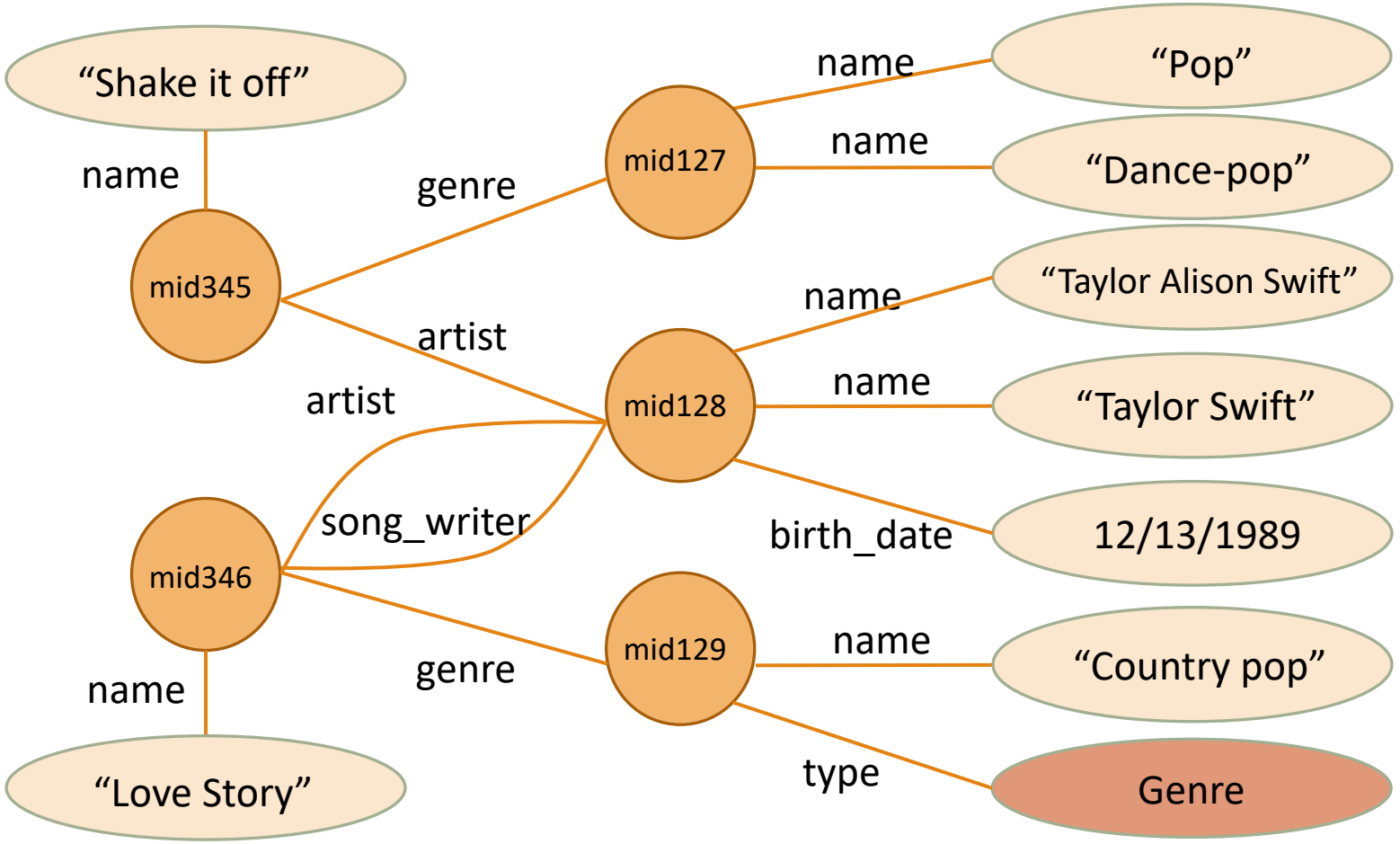
- Mission: To answer any question about products and related knowledge in the world



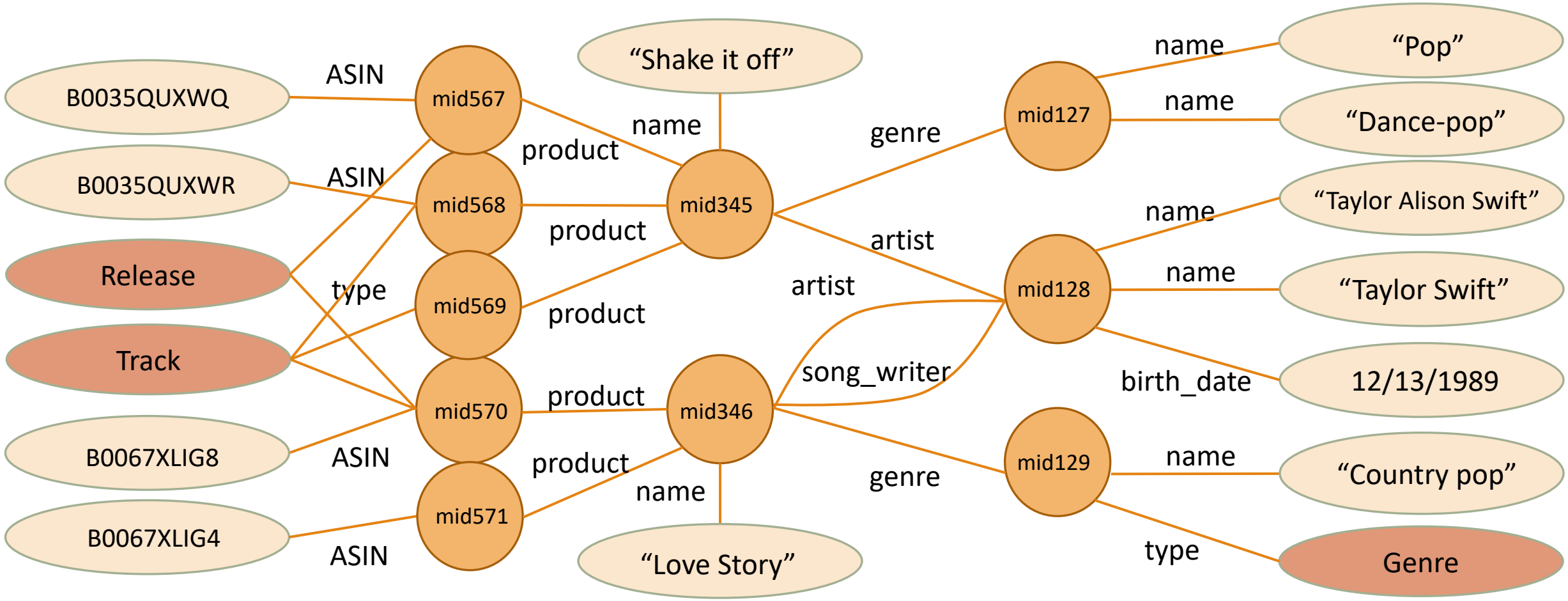
Knowledge Graph Example for 2 Songs



Product Graph Example for 2 Songs

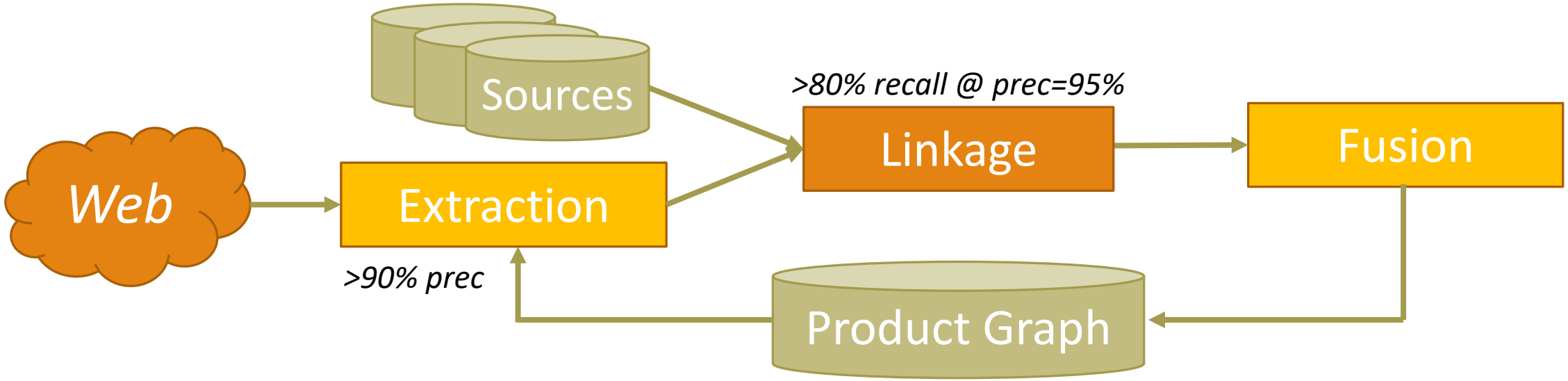


Product Graph Example for 2 Songs



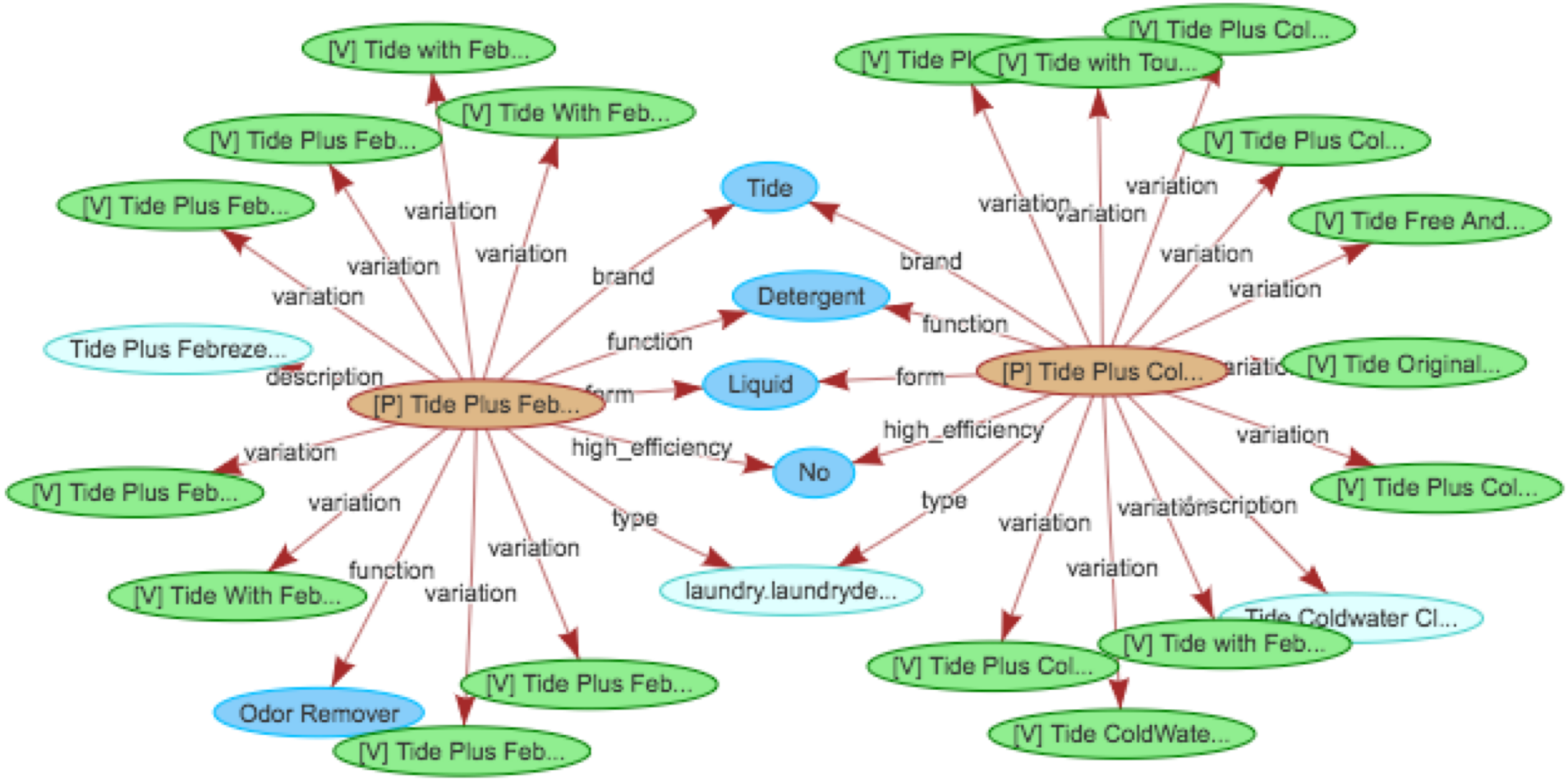
Product Graph for Media Products

- Main challenge: **Heterogeneous** Data everywhere
- Key techniques



- Support ~20 Amazon Music applications

Product Graph for Retail Products



Product Graph for Retail Products

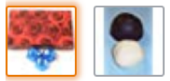
- Main challenge: **Sparse & Noisy** Data everywhere

Product Graph for Retail Products

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…	1 lb. Daisy Box	1 lb. Red Rose Box
$19.95	$19.95	$19.95

19 Sizes: Dark Blue Foils

Aqua Foils	Burgandy Foils	Dark Blue Foils
$19.95	$19.95	$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.

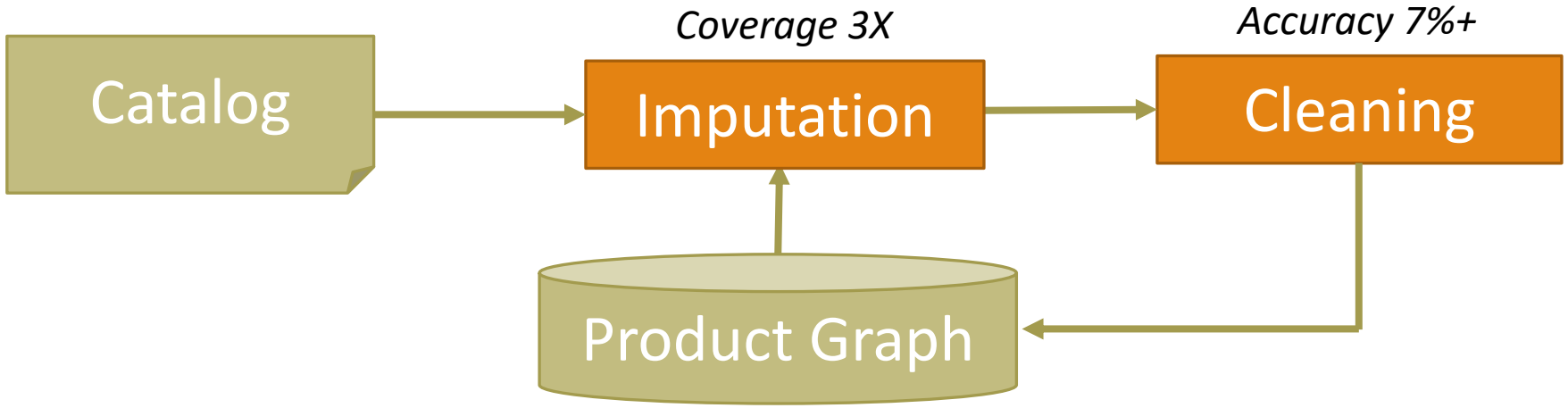
Deliver to Yaqing - Seattle 98109

Qty: 1

Turn on 1-click ordering

Product Graph for Retail Products

- Main challenge: **Sparse & Noisy** Data everywhere
- Key techniques



- Live on Alexa Shopping, Detail Page, and Amazon Search

How to Scale Up?

Scale Up in 3 Dimensions

Hierarchy of 100K categories

Focus of this talk

One vertical,
A few sources

Millions of sources

Behavior info from tens of applications over years

Big challenge: Limited training labels for large-scale, rich data

Challenges for 100K Categories (I)

- Require 100K+ models
  - Different value vocabularies
  - Different patterns to express the same attribute

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)

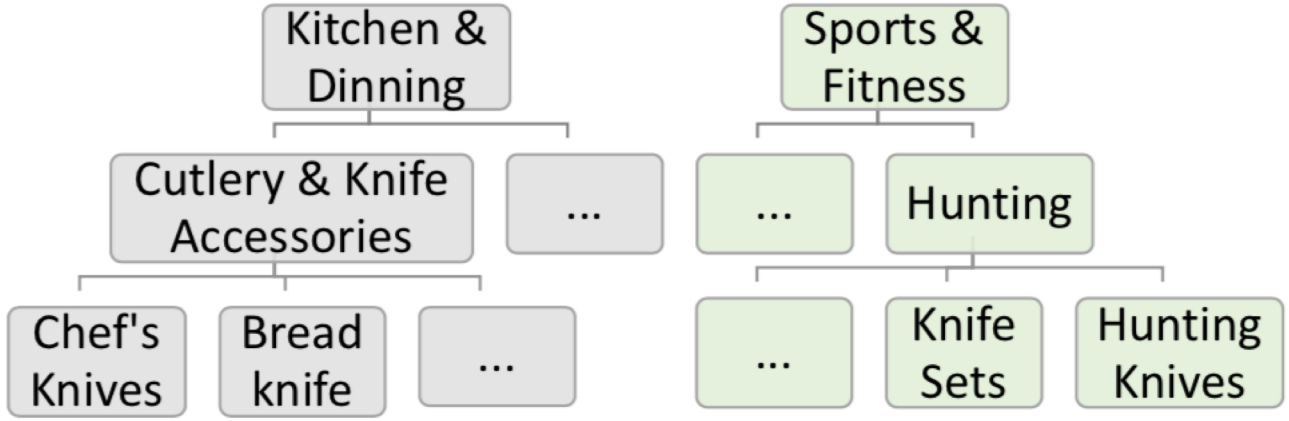
Caribou Coffee Caribou Blend, Medium Roast Ground Coffee, 20 Ounce Bag, Rainforest Alliance Certified

Challenges for 100K Categories (II)

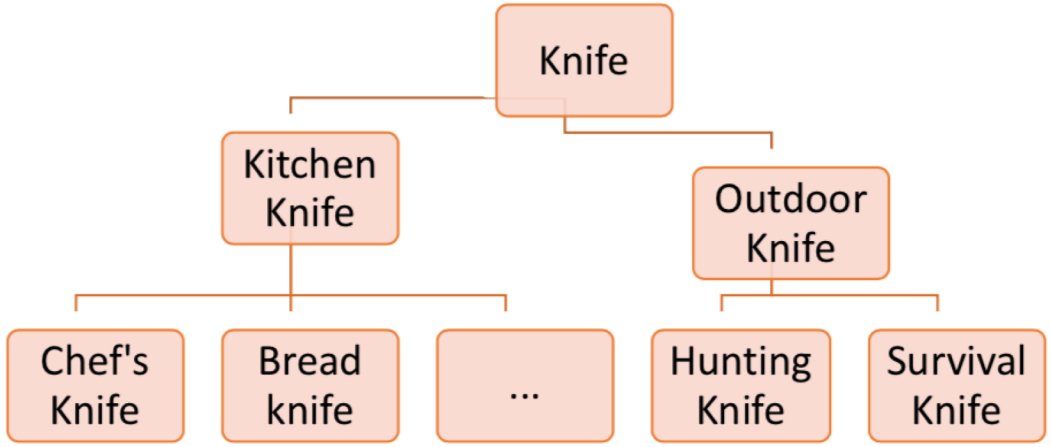
- Different attributes apply to different product categories
  - A lot of work to manually define ontology
  - Hard to catch the trend of new product categories and properties

Challenges for 100K Categories (III)

- It is hard to build a product taxonomy
  - What are product types in customers' language?
  - How to organize types into a taxonomy tree?

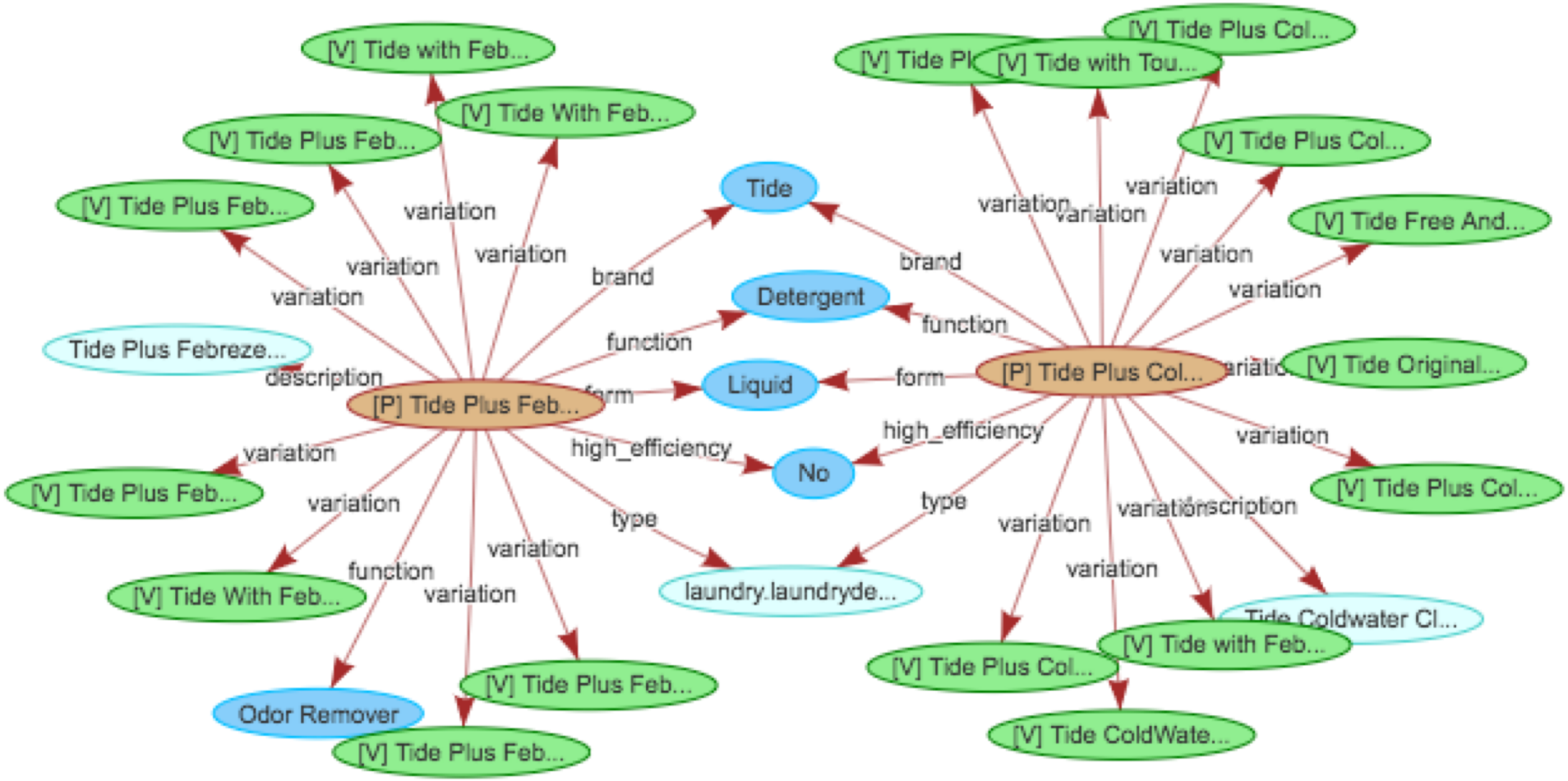


Sellers' view



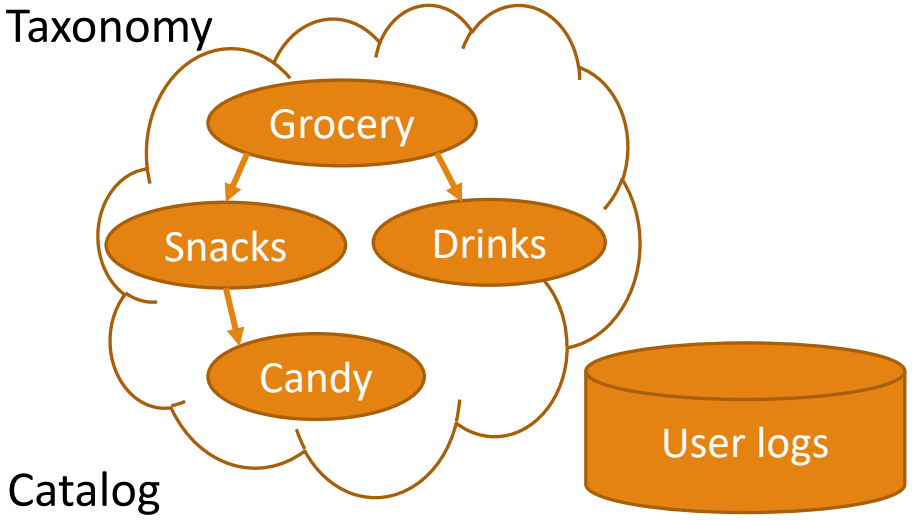
Buyers' view

A 100-Year Project

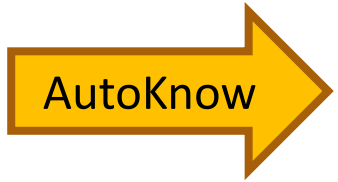


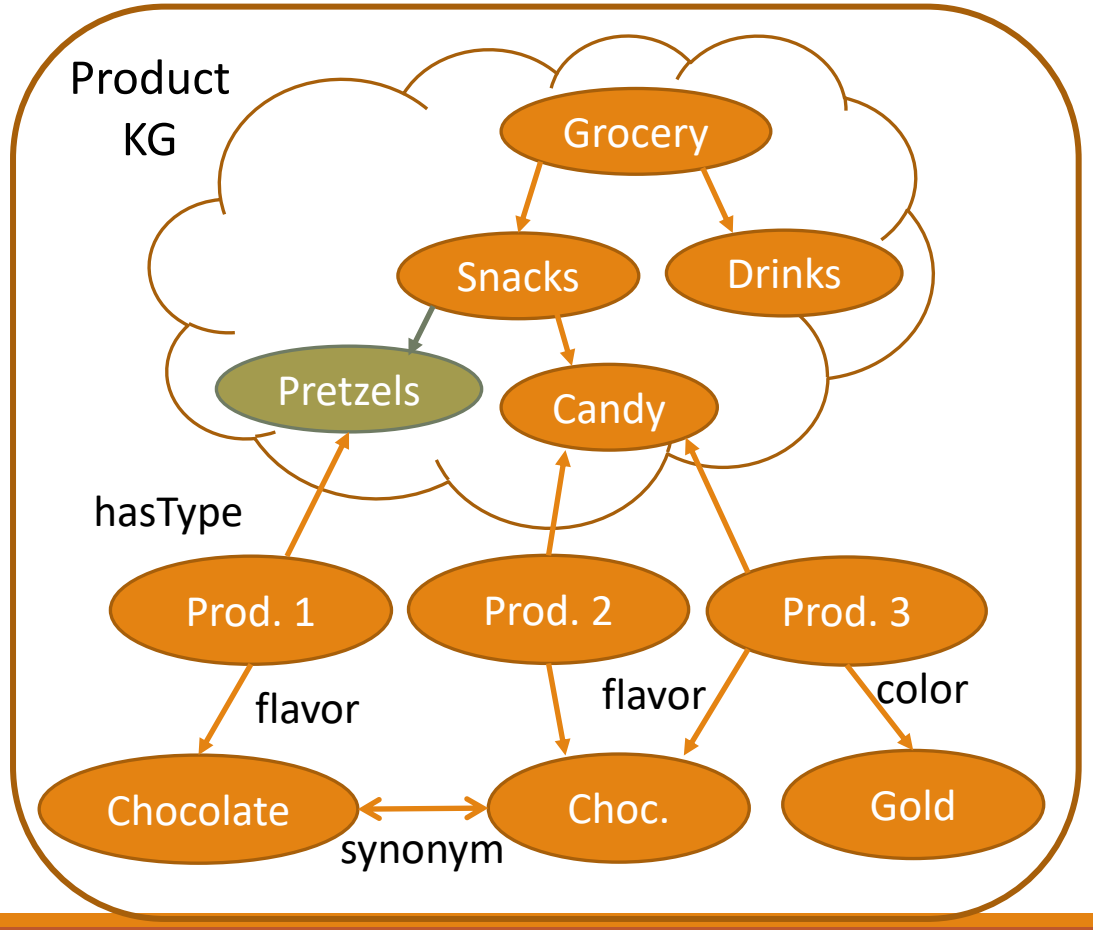
Our Solution: Self-Driving Product Understanding

AutoKnow: Self-Driving Product Knowledge Collection



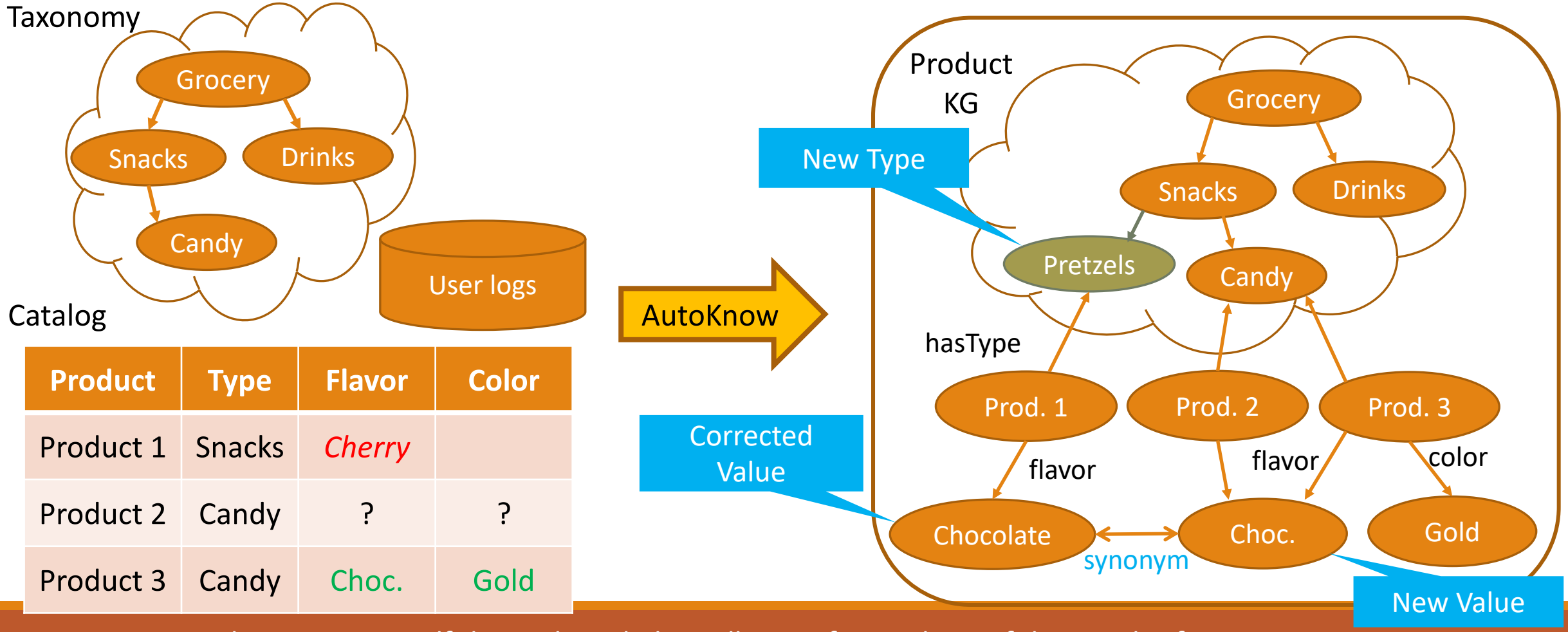
Product	Type	Flavor	Color
Product 1	Snacks	*Cherry*	
Product 2	Candy	?	?
Product 3	Candy	Choc.	Gold





Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

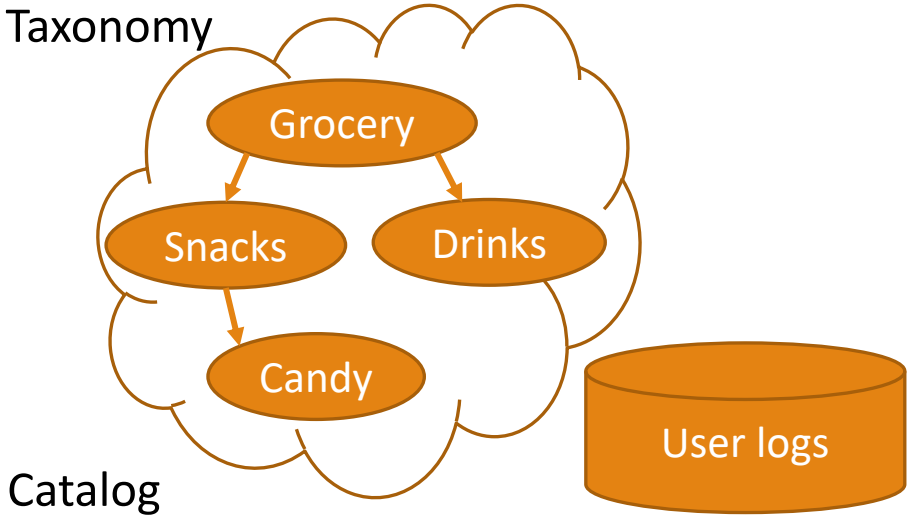
AutoKnow: Self-Driving Product Knowledge Collection



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

Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection

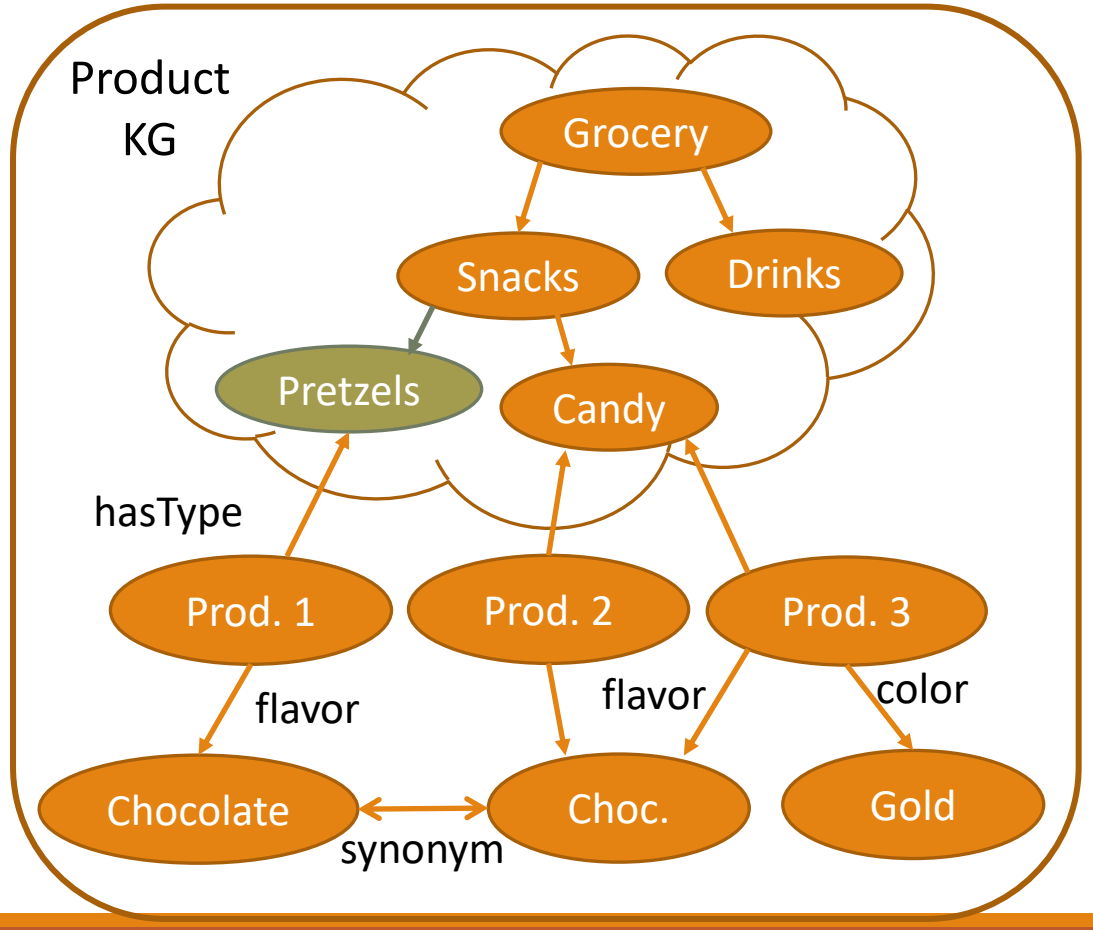


Product	Type	Flavor	Color
Product 1	Snacks	*Cherry*	
Product 2	Candy	?	?
Product 3	Candy	Choc.	Gold

AutoKnow

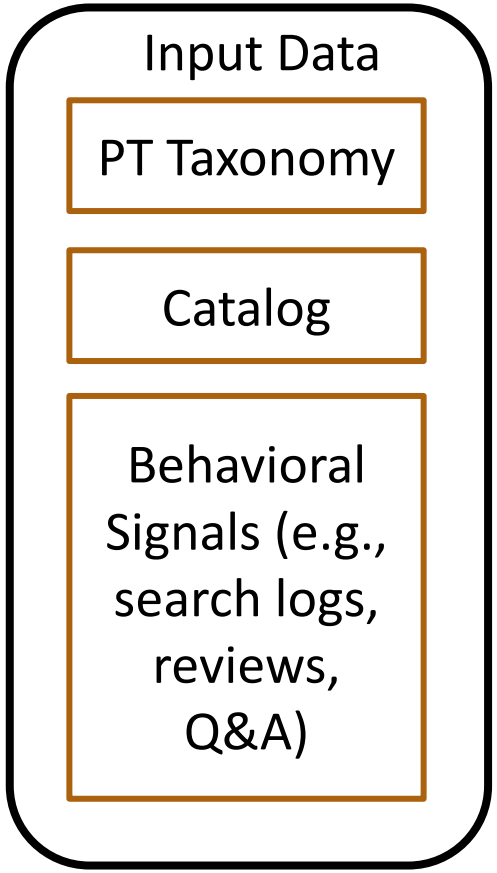
- **#Types ↑ 3X**
- **Defect rate ↓**

up to 68 percent points



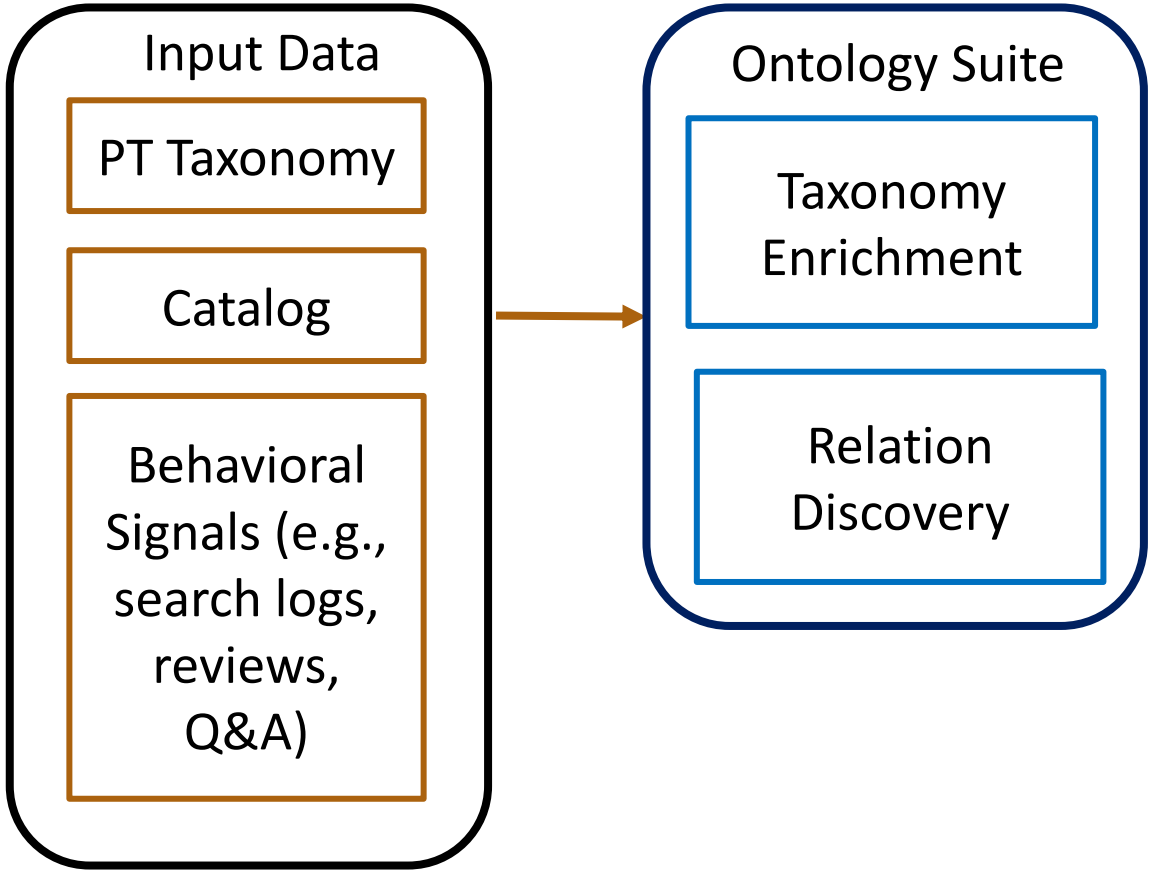
Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection



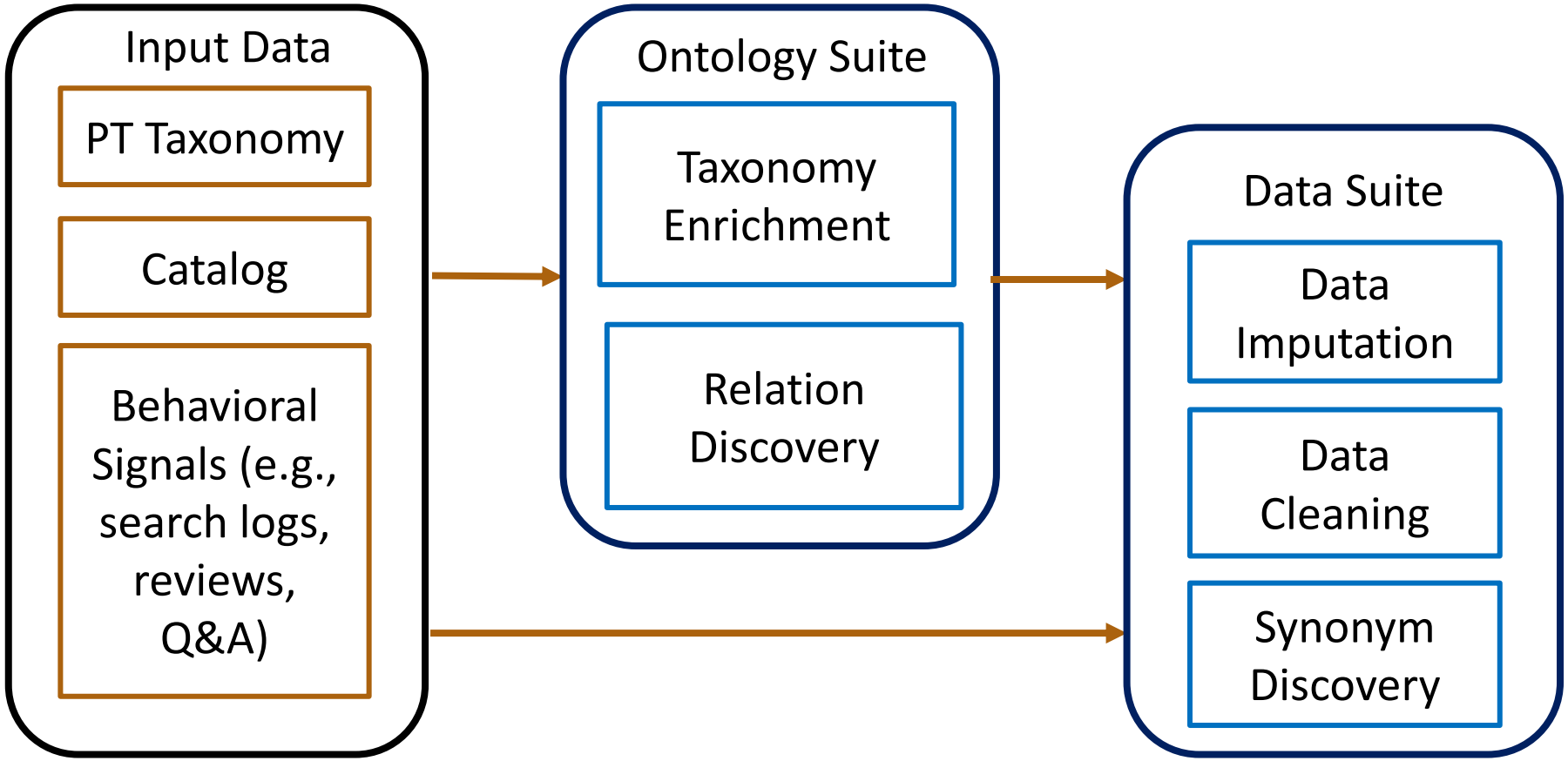
Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection



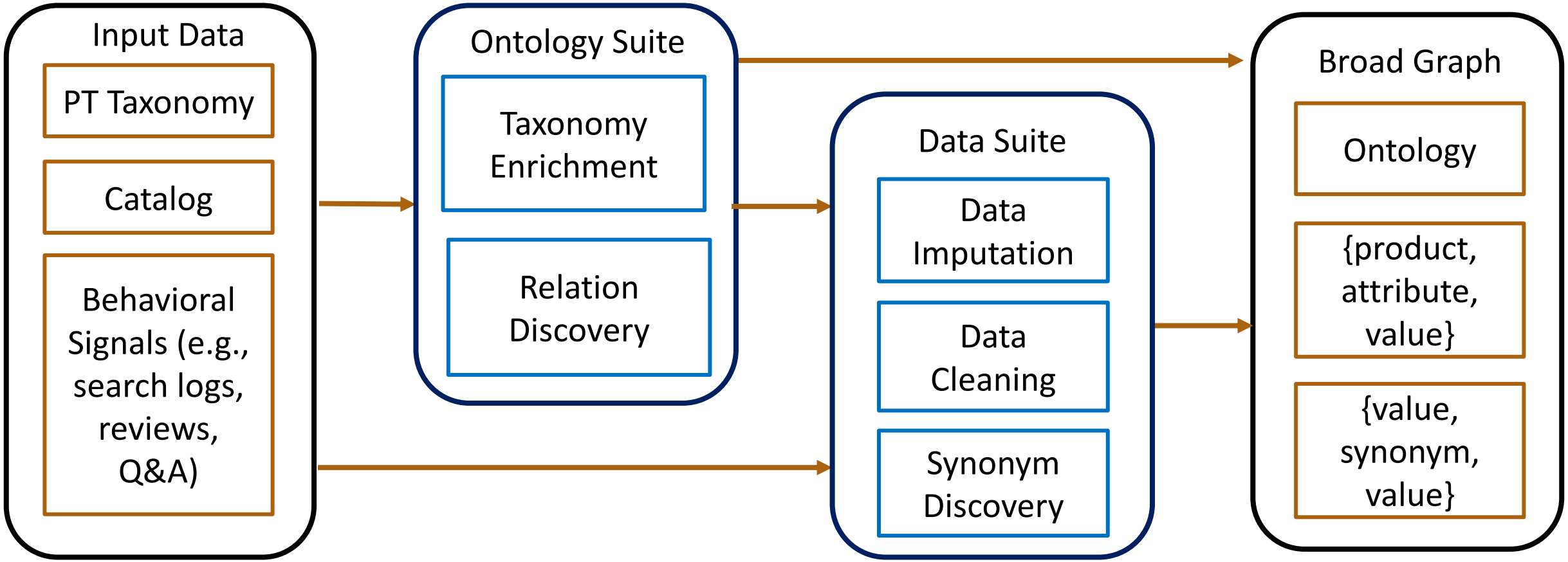
Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection



Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection



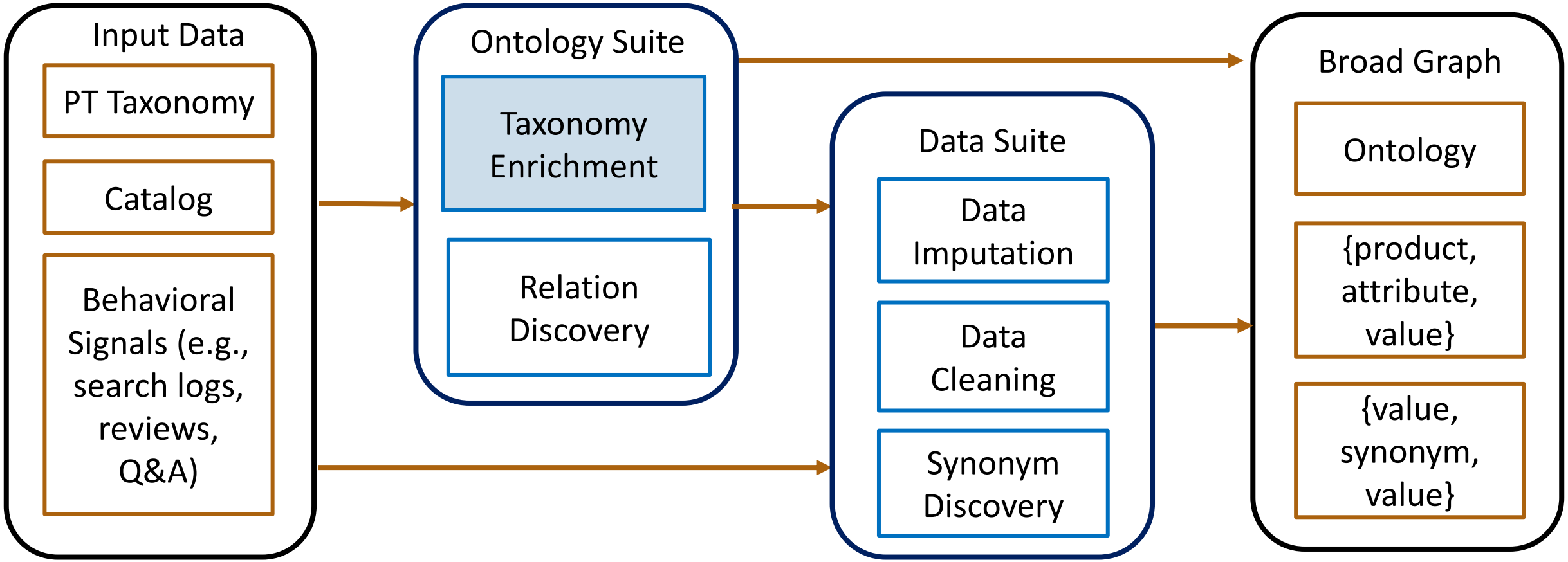
Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

Self-Driving to Navigate 100K Categories

- **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

Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection



Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

I. GNN-Based Taxonomy Enrichment

- Step 1. Type extraction from product titles and queries by OpenTag tagging system

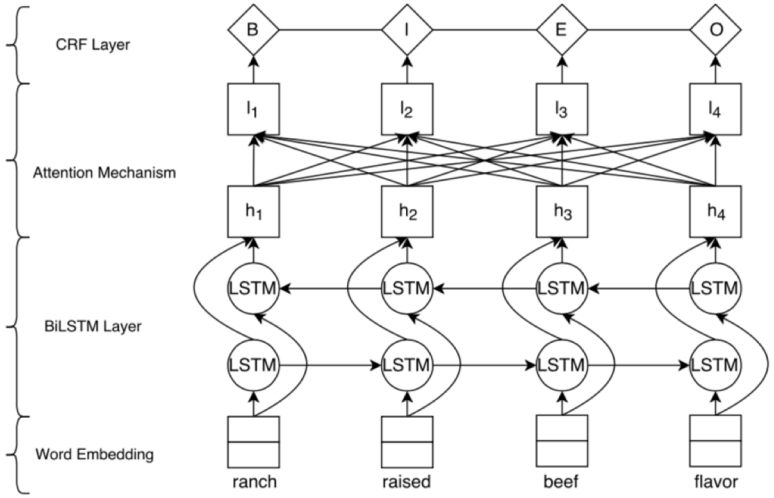


Figure 2: OpenTag **Architecture: BiLSTM-CRF with Attention.**

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

Mao et al., OCTET: Online catalog taxonomy enrichment with self-supervision, SigKDD, 2020.

I. GNN-Based Taxonomy Enrichment

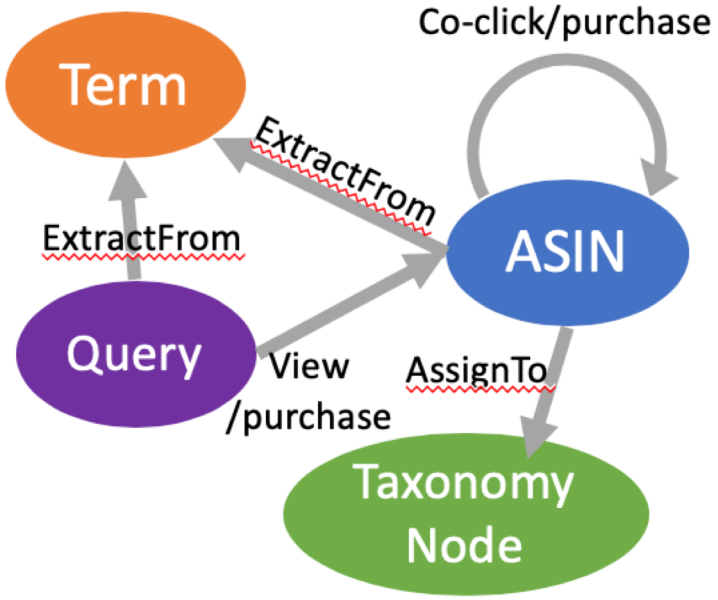
- Consumables taxonomy **size increases by 2.9X**

Source	Text	Product Type
Product	*4 Country Pasta Homemade Style Egg Pasta - 16-oz bag*	*Egg Pasta*
Product	*Hamburger Helper Lasagna Pasta, Four Cheese, 10.3 Ounce (Pack of 6)*	*Lasagna Pasta*
Product	*COFFEE MATE The Original Powder Coffee Creamer 35.3 Oz. Canister Non-dairy, Lactose Free, Gluten Free Creamer*	*Coffee Creamer*
Query	*mccormick paprika 8.5 ounce*	*paprika*
Query	*flax seeds raw*	*flax seeds*

Mao et al., OCTET: Online catalog taxonomy enrichment with self-supervision, SigKDD, 2020.

I. GNN-Based Taxonomy Enrichment

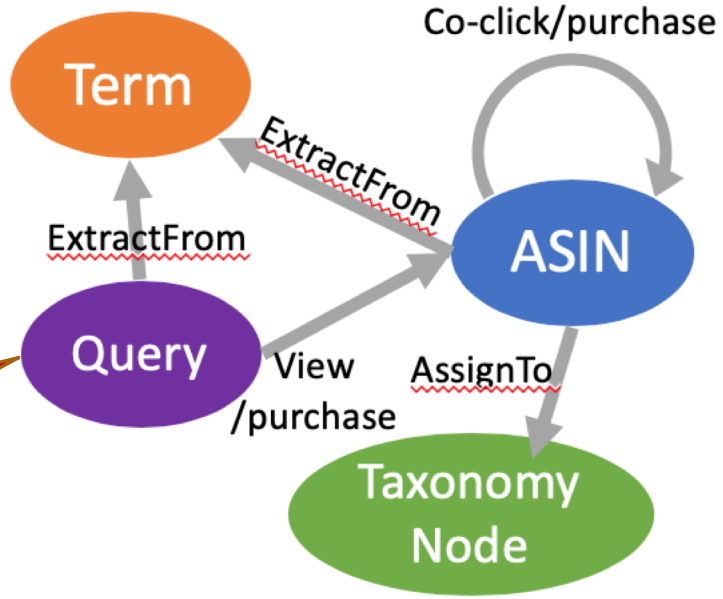
- Step 2. Taxonomy expansion with hypernym discovery by Graph Neural Network



Mao et al., OCTET: Online catalog taxonomy enrichment with self-supervision, SigKDD, 2020.

I. GNN-Based Taxonomy Enrichment

- Step 2. Taxonomy expansion with hypernym discovery by Graph Neural Network



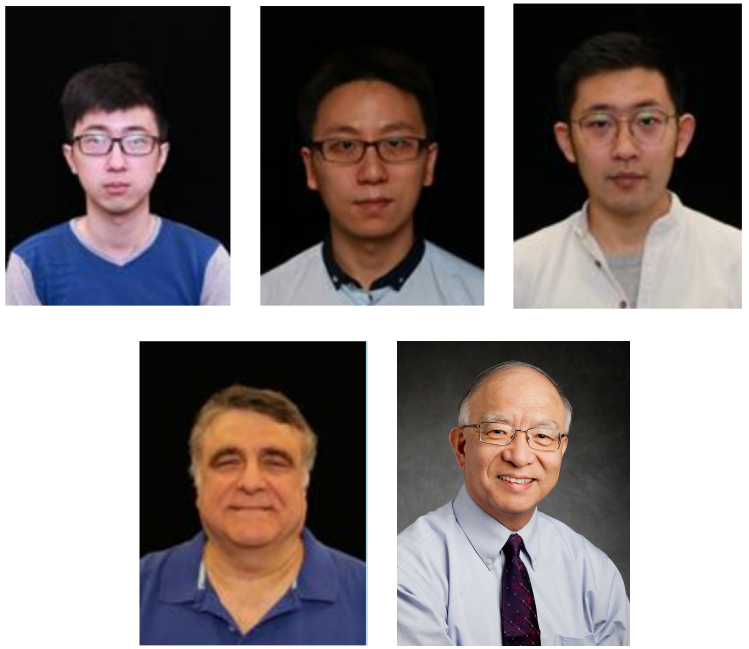
Leveraging search behavior with GNN

Mao et al., OCTET: Online catalog taxonomy enrichment with self-supervision, SigKDD, 2020.

I. GNN-Based Taxonomy Enrichment

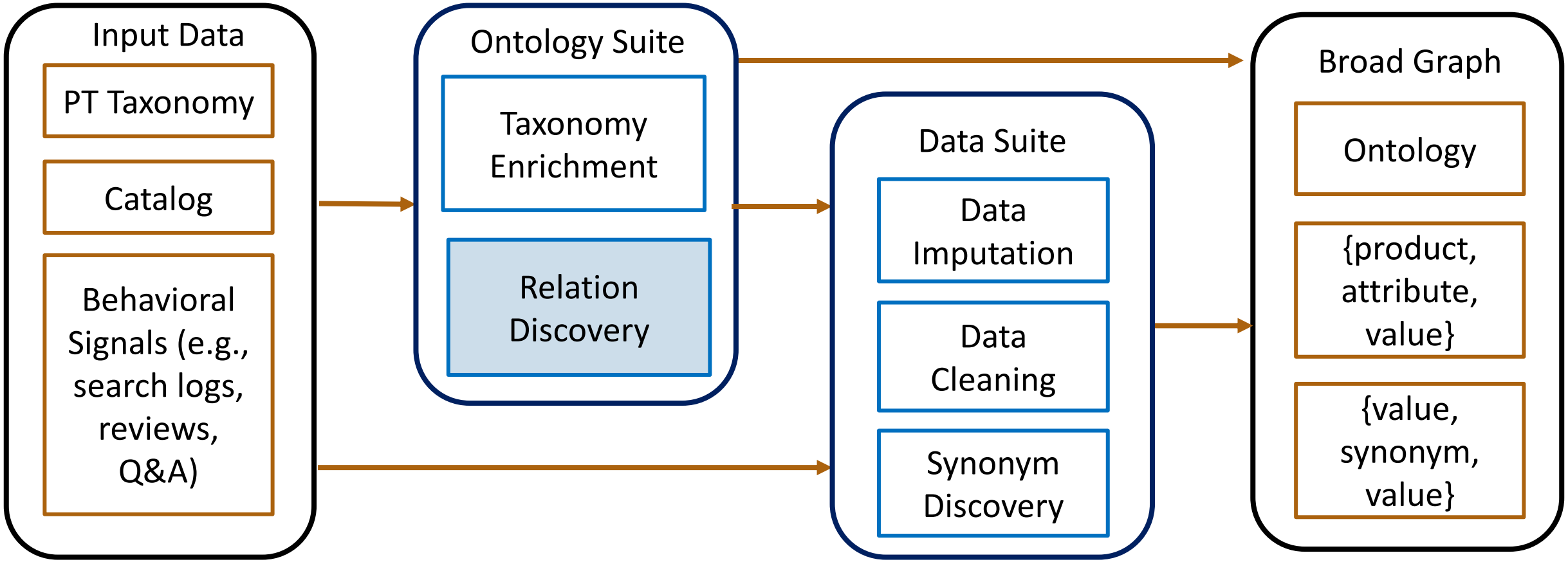
- Improve hypernym identification **F1 by 18% over state-of-the-art**

Child Type	Parent Type
Coconut flour	*Baking flours & meals*
Tilapia	*Fresh fish*
Fresh cut carnations	*Fresh cut flowers*
Bock beers	*Lager & pilsner beers*
Pinto beans	*Dried beans*



Mao et al., OCTET: Online catalog taxonomy enrichment with self-supervision, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection



Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

II. Behavior-Based Attribute Importance

Child safety car seats	Almond
brand	flavor
style	brand
size or capacity	ingredients
maximum weight recommendation	size or capacity
color	number of items
model	energy content per serving
number of items	bar code
usage	color
bar code	scent
flavor	usage
ingredients	style
scent	model
energy content per serving	maximum weight recommendation

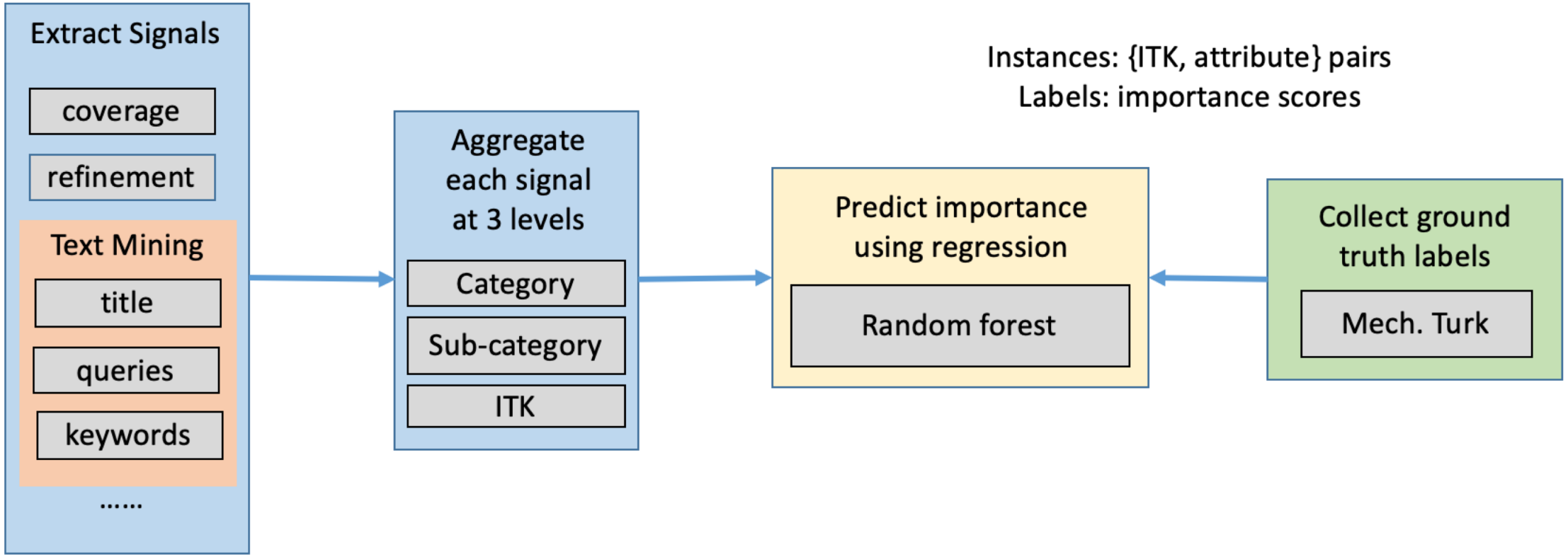
Important (ranked by score)

Unimportant

Unapplicable

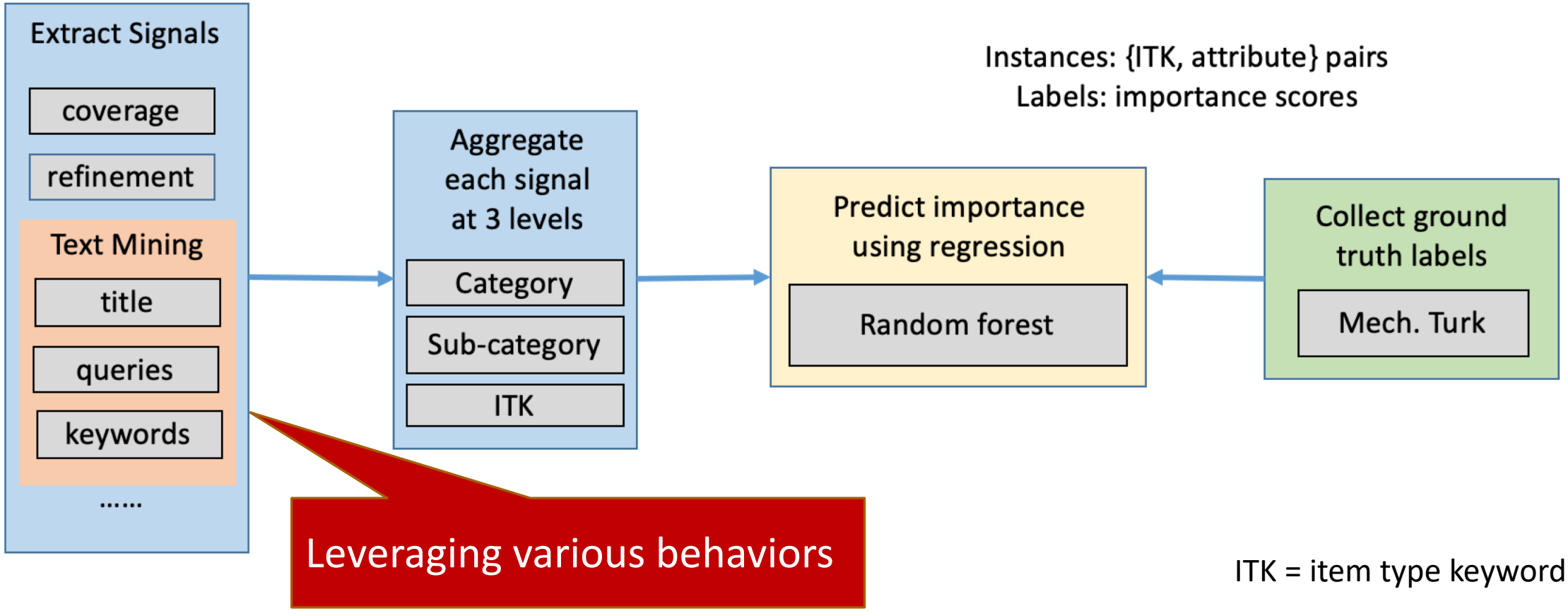
Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

II. Behavior-Based Attribute Importance



ITK = item type keyword

II. Behavior-Based Attribute Importance



Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

II. Behavior-Based Attribute Importance

- Predict for **7.5K categories** the applicability of **3K attributes** with **P=0.94/R=0.84**, and importance with 0.74 correlation (Spearman)

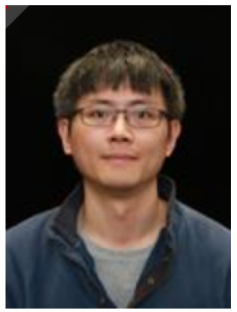
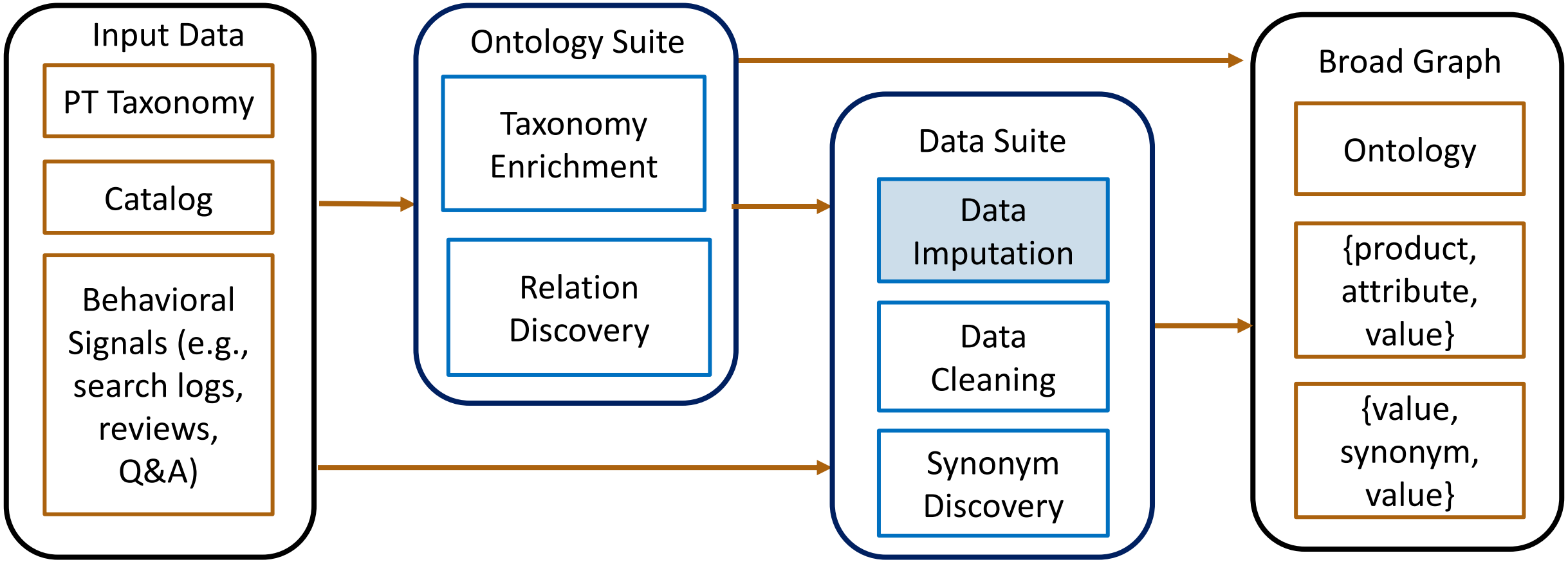


Table 15: Attributes identified as most important for two example types.

Cereals	Shampoo
brand	*brand*
ingredients	*hair type*
flavor	*number of items*
number of items	*ingredients*
energy content	*liquid volume*

Dong et al., AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD, 2020.

AutoKnow: Self-Driving Product Knowledge Collection



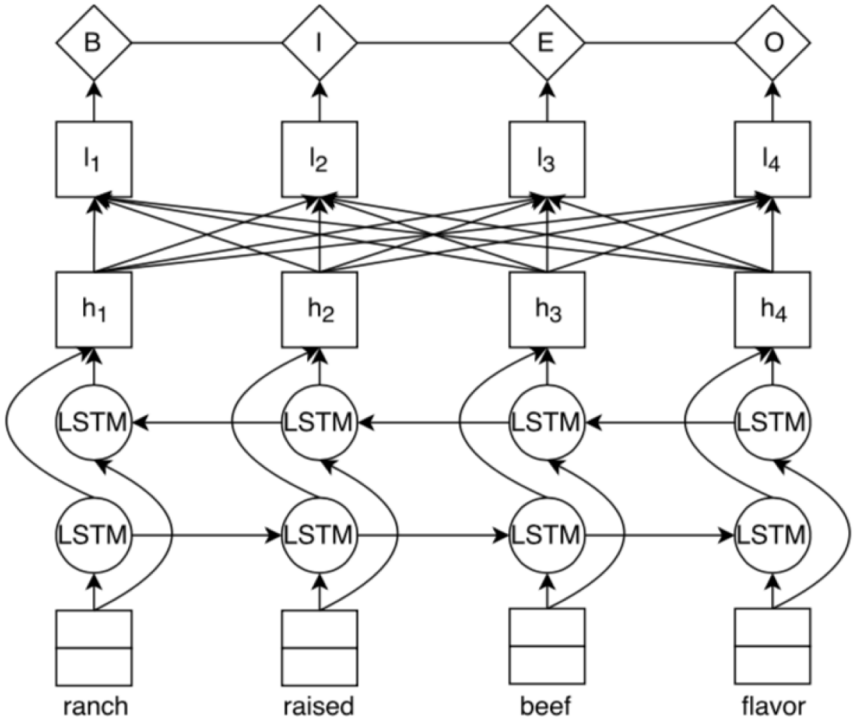
III. Value Extraction—Traditional

CRF
Captures correlations between BIOE tags

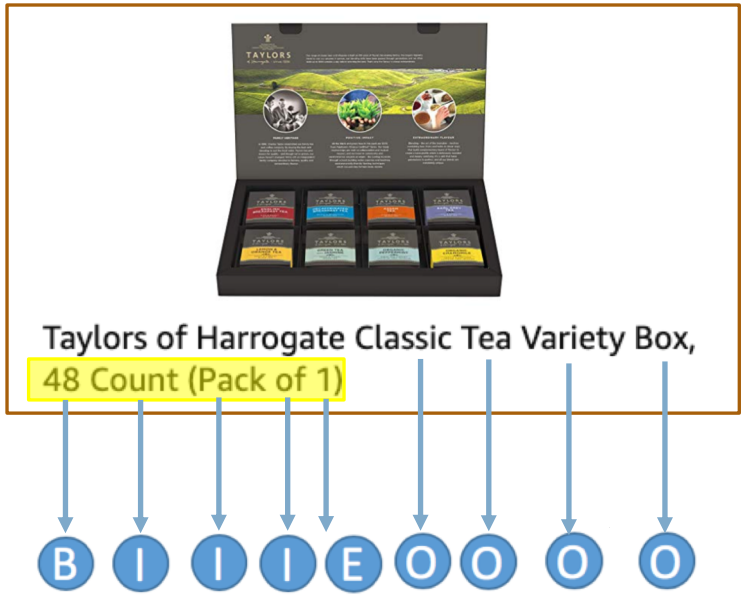
Attention
Identifies important terms leading to attribute values

Bi-LSTM
Captures sequence info

Word Embedding
Captures semantics of each token



OpenTag



Zheng et al., OpenTag: Open attribute value extraction from product profiles, KDD 2018.

III. Value Extraction—100K Categories

Option 1. Train a single model? *Train/Test Distribution shift -> Invalid predictions*

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

Karamanolakis et al., TXtract: Taxonomy-aware knowledge extraction for thousands of product categories, ACL 2020.

III. Value Extraction—100K Categories

Option 1. Train a single model? *Train/Test Distribution shift -> Invalid predictions*

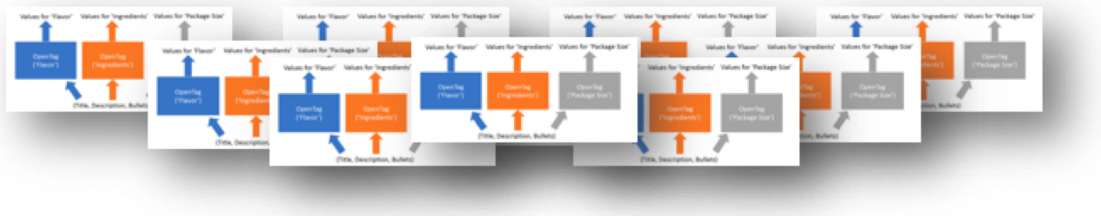
Samsung UN58RU7100FXZA Flat 58-Inch 4K UHD 7 Series Ultra HD Smart TV with HDR and Alexa Compatibility (2019 Model)

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

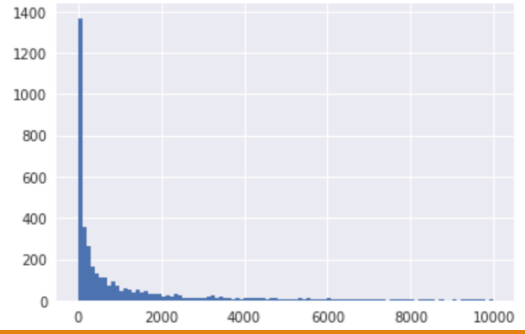
Caribou Coffee Caribou Blend, Medium Roast Ground Coffee, 20 Ounce Bag, Rainforest Alliance Certified

Option 2. Train a model for each category?

Store/orchestrate 100K+ OpenTag models



Most categories are very sparse



Karamanolakis et al., TXtract: Taxonomy-aware knowledge extraction for thousands of product categories, ACL 2020.

III. Value Extraction—Taxonomy-Aware

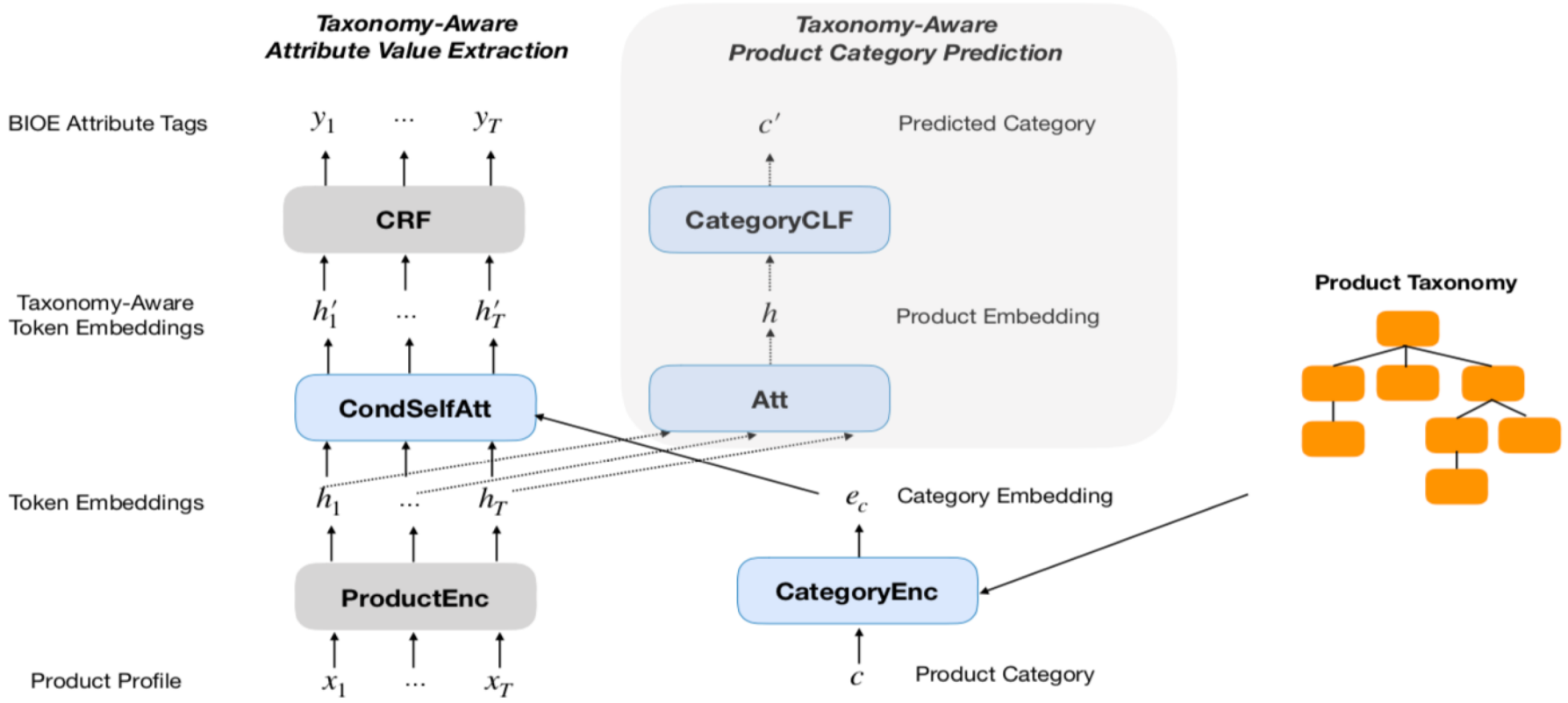


Figure 2: Our TXtract architecture for hierarchical multi-task learning.

III. Value Extraction—Taxonomy-Aware

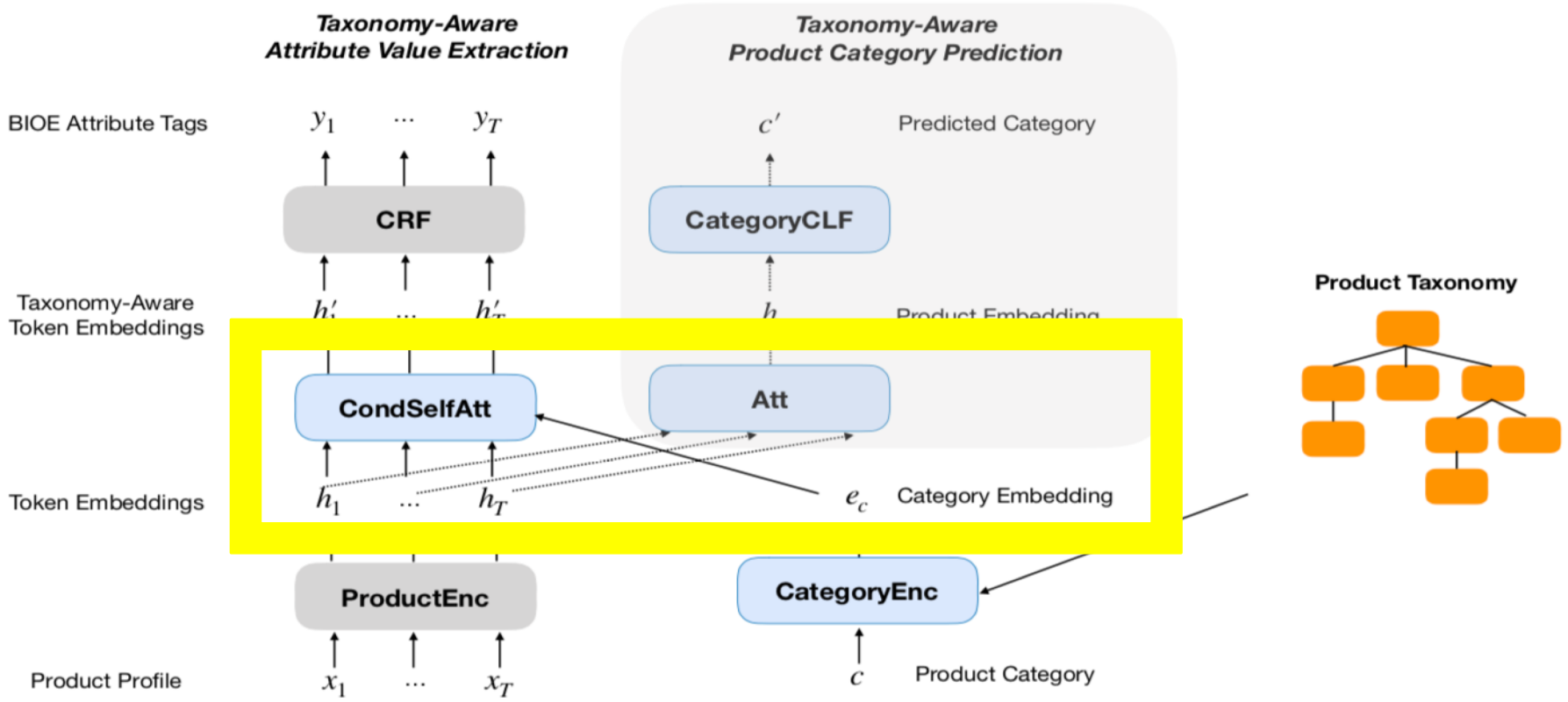


Figure 2: Our TXtract architecture for hierarchical multi-task learning.

III. Value Extraction—Taxonomy-Aware

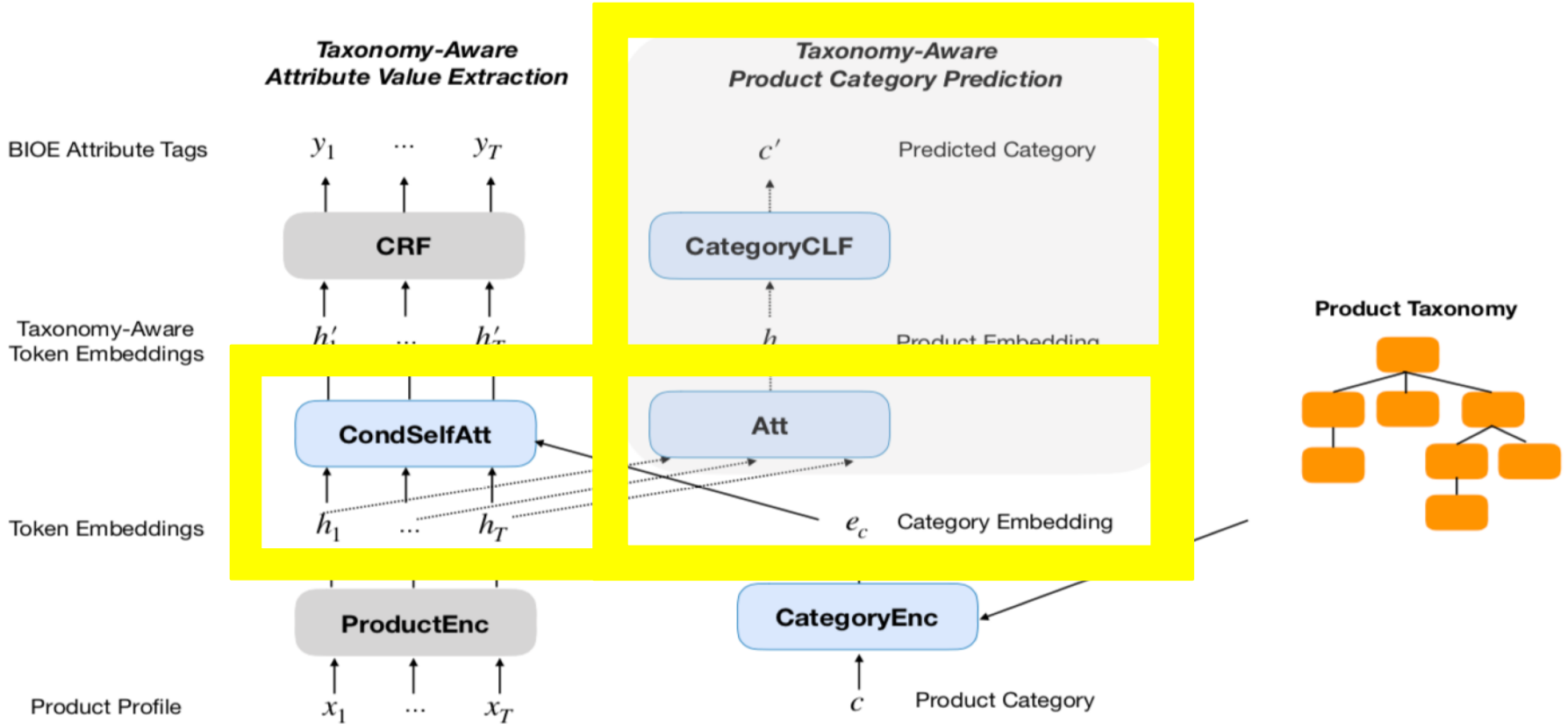


Figure 2: Our TXtract architecture for hierarchical multi-task learning.

III. Value Extraction—Taxonomy-Aware

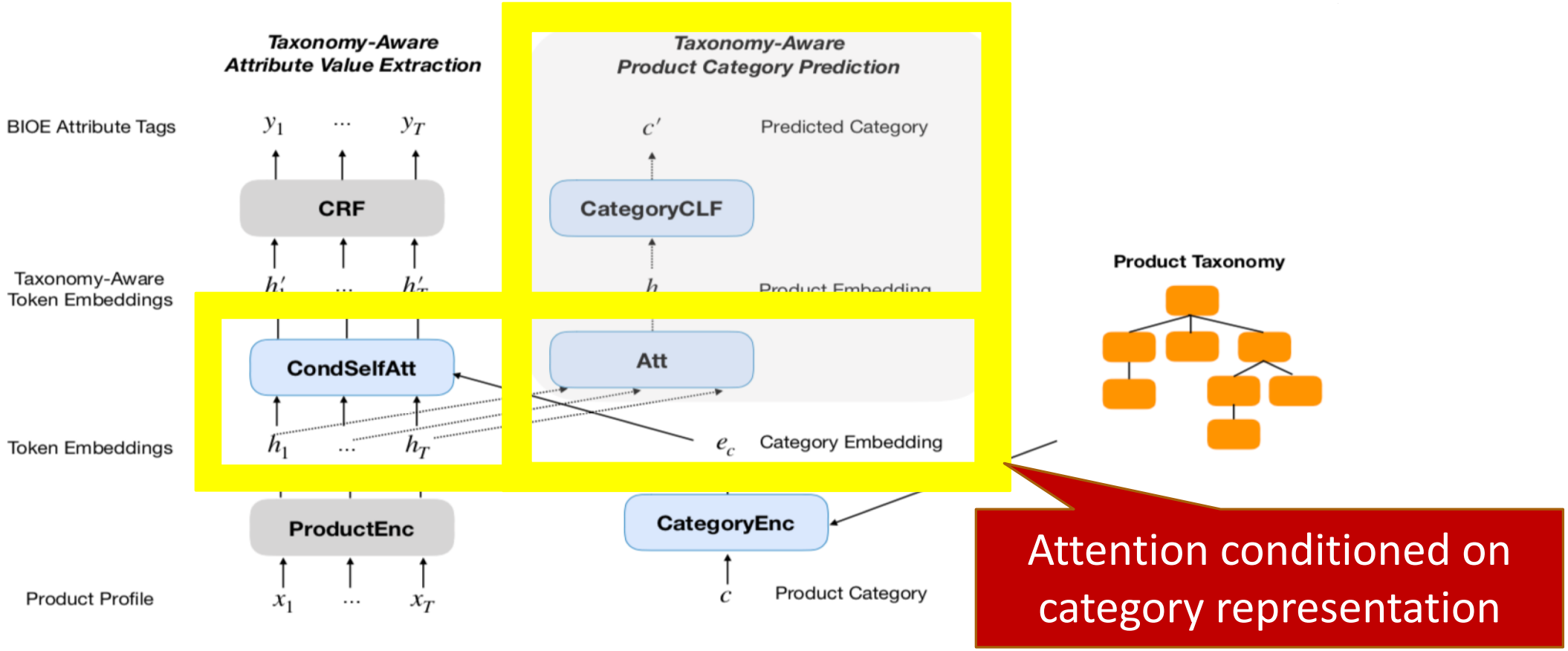
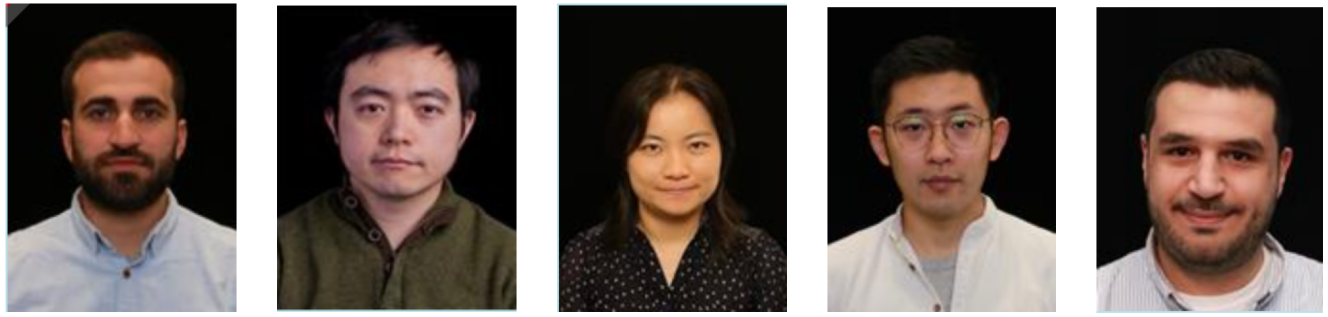


Figure 2: Our TXtract architecture for hierarchical multi-task learning.

III. Value Extraction—Taxonomy-Aware

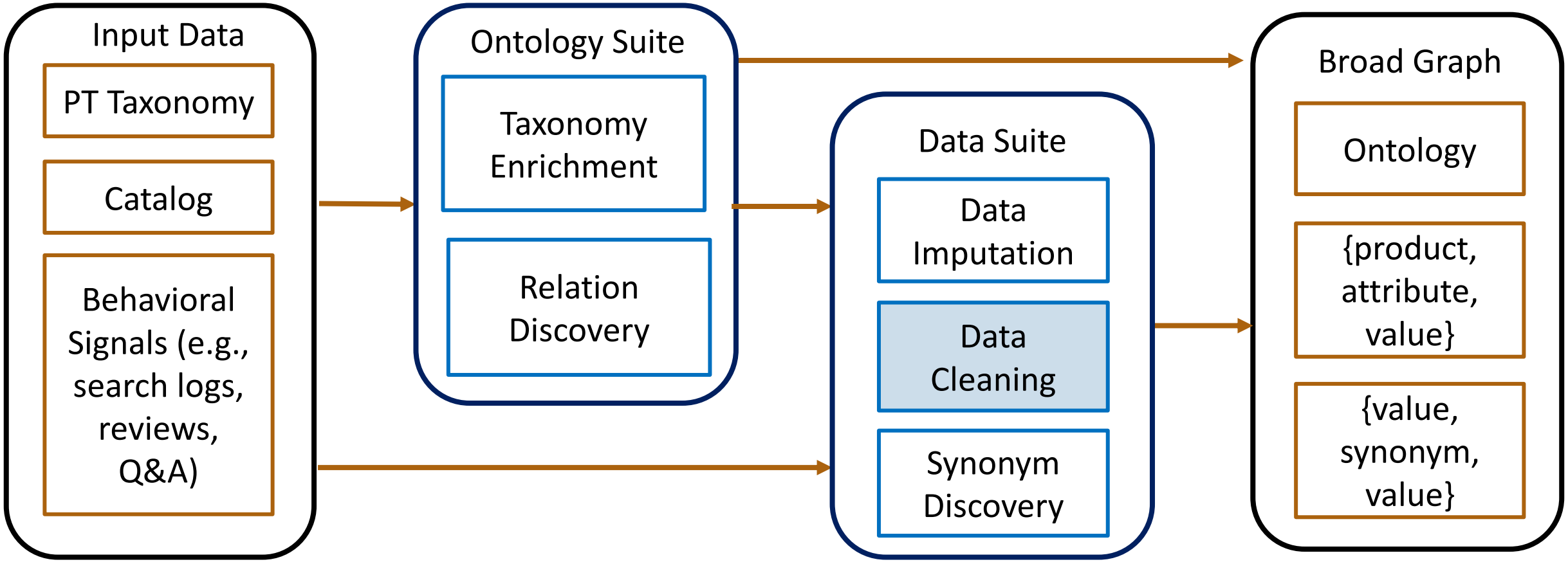
- Train one model on 4K categories, and improve state-of-the-art by 10.4% in F1, and by 11.7% in coverage

	Title	OpenTag	TXtract
1	Controlled Labs Purple Wraath 90 Servings - Purple Lemonade	*flavor*: -	*flavor*: purple lemonade
2	Click - Espresso Protein Drink Vanilla Latte - 16 oz.	*flavor*: espresso	*flavor*: vanilla latte
3	Mason Vitamins Melatonin 500 mcg Fast Meltz Tablets, Fruit, 60 Count	*flavor*: -	*flavor*: fruit
4	Fashion Glitter Matte Eye Shadow Powder Palette Single Shimmer Eyeshadow	*scent*: palette	*scent*: -
5	Baby car seat cover, Nursing covers Breastfeeding cover carseat canopy (Style5)	*scent*: style5	*scent*: -

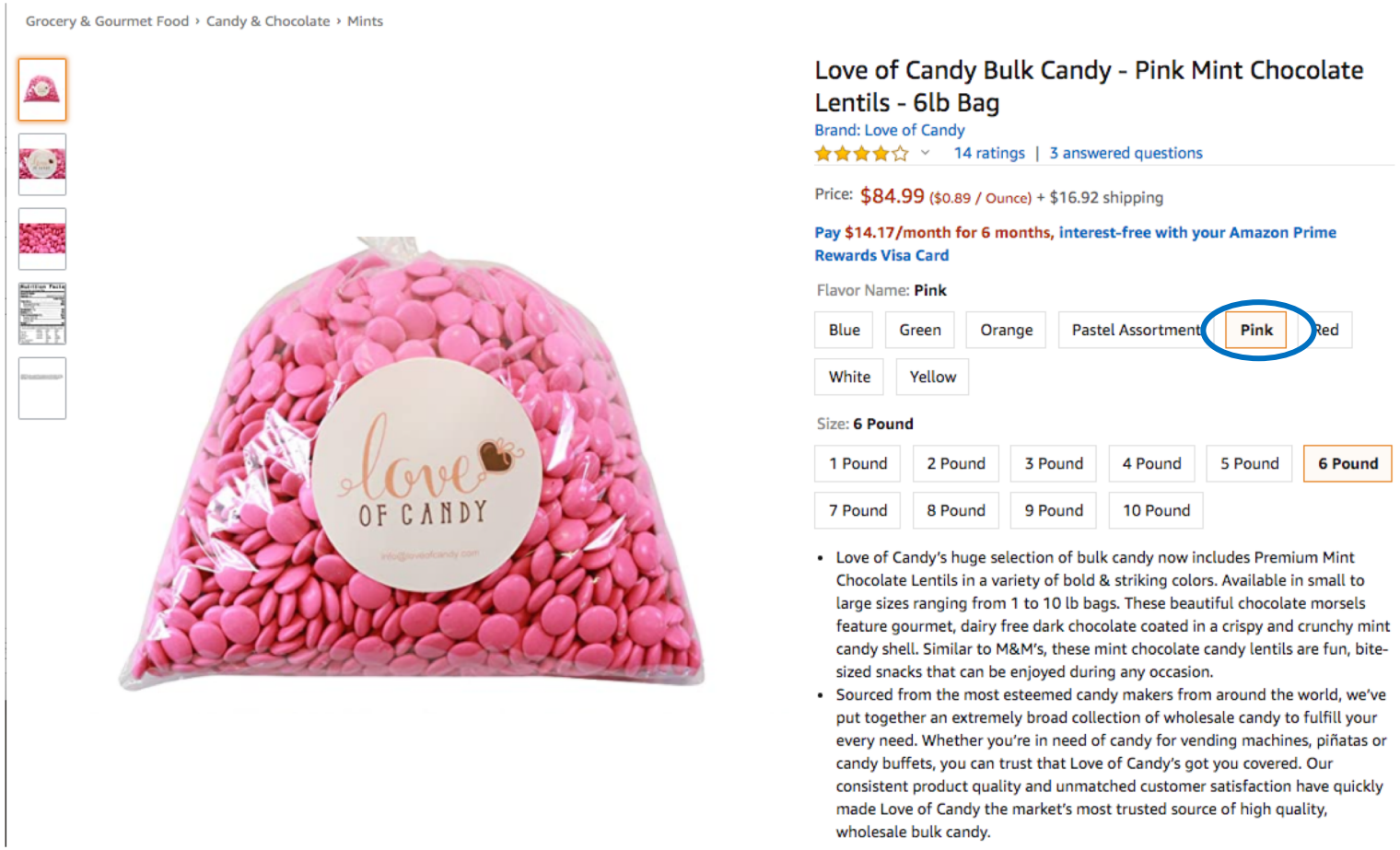


Karamanolakis et al., TXtract: Taxonomy-aware knowledge extraction for thousands of product categories, ACL 2020.

AutoKnow: Self-Driving Product Knowledge Collection



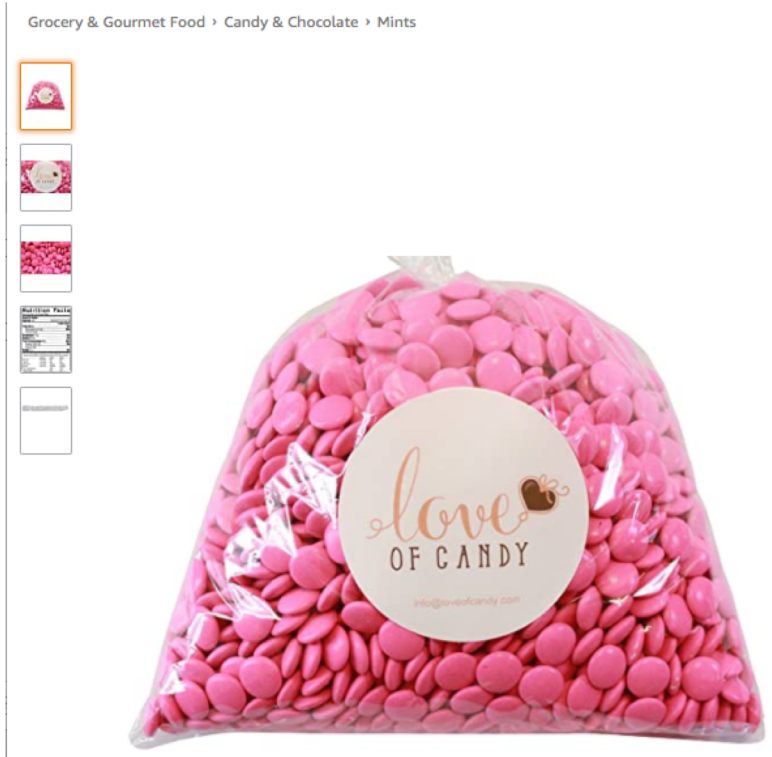
IV. Taxonomy-Aware Anomaly Detection



Wang et al., Automatic validation of textual attribute values in eCommerce Catalog by learning with limited labeled data, KDD'20

IV. Taxonomy-Aware Anomaly Detection

Grocery & Gourmet Food › Candy & Chocolate › Mints



Love of Candy Bulk Candy - Pink Mint Chocolate Lentils - 6lb Bag

Brand: Love of Candy

14 ratings | 3 answered questions

Price: $84.99 ($0.89 / Ounce) + $16.92 shipping

Pay $14.17/month for 6 months, interest-free with your Amazon Prime Rewards Visa Card

Flavor Name: **Pink**

Blue | Green | Orange | Pastel Assortment | **Pink** | Red | White | Yellow

Size: **6 Pound**

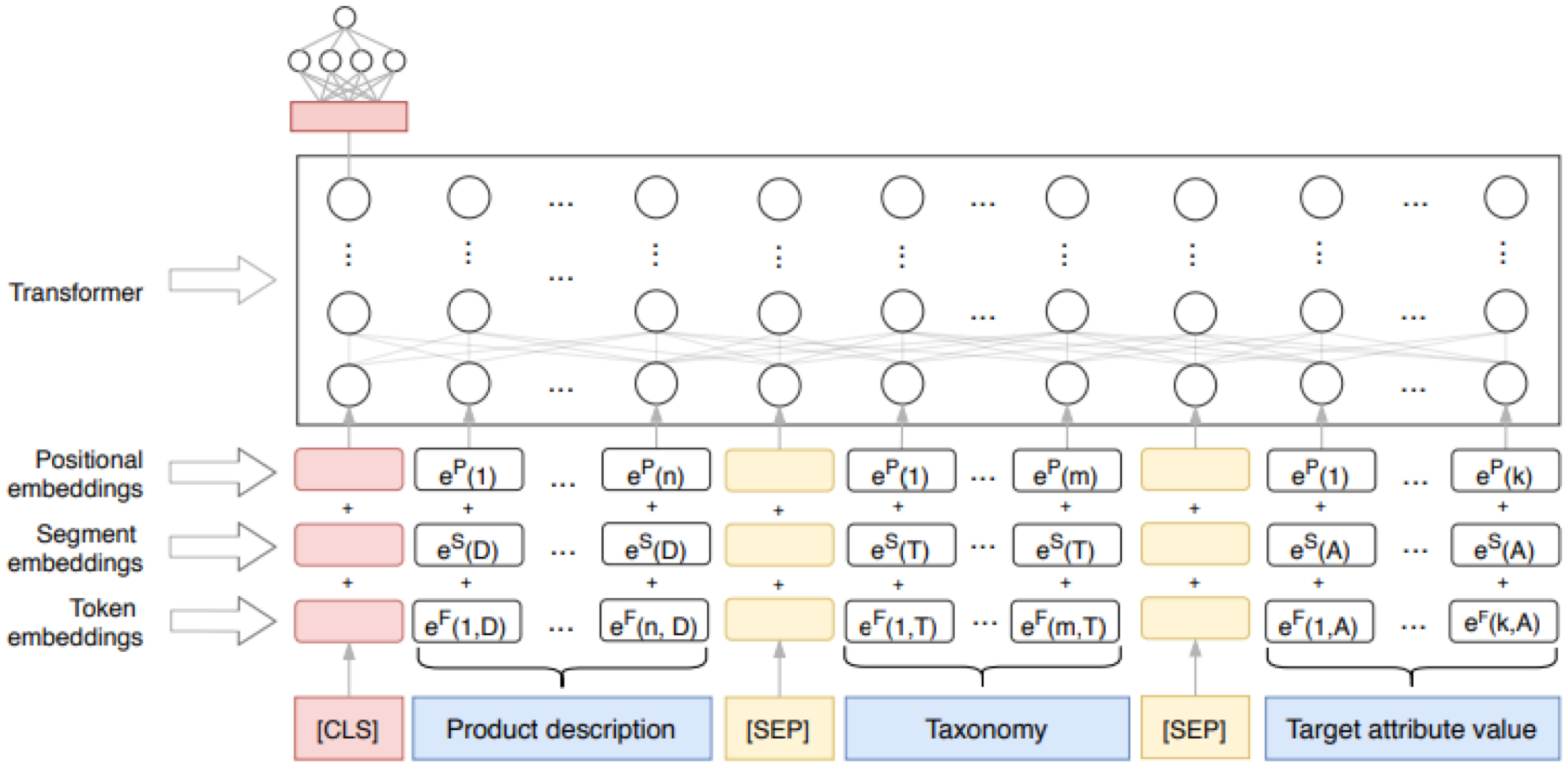
1 Pound | 2 Pound | 3 Pound | 4 Pound | 5 Pound | **6 Pound** | 7 Pound | 8 Pound | 9 Pound | 10 Pound

- Love of Candy's huge selection of bulk candy now includes Premium Mint Chocolate Lentils in a variety of bold & striking colors. Available in small to large sizes ranging from 1 to 10 lb bags. These beautiful chocolate morsels feature gourmet, dairy free dark chocolate coated in a crispy and crunchy mint candy shell. Similar to M&Ms, these mint chocolate candy lentils are fun, bite-sized snacks that can be enjoyed during any occasion.
- Sourced from the most esteemed candy makers from around the world, we've put together an extremely broad collection of wholesale candy to fulfill your every need. Whether you're in need of candy for vending machines, piñatas or candy buffets, you can trust that Love of Candy's got you covered. Our consistent product quality and unmatched customer satisfaction have quickly made Love of Candy the market's most trusted source of high quality, wholesale bulk candy.

Is the flavor "Pink"?

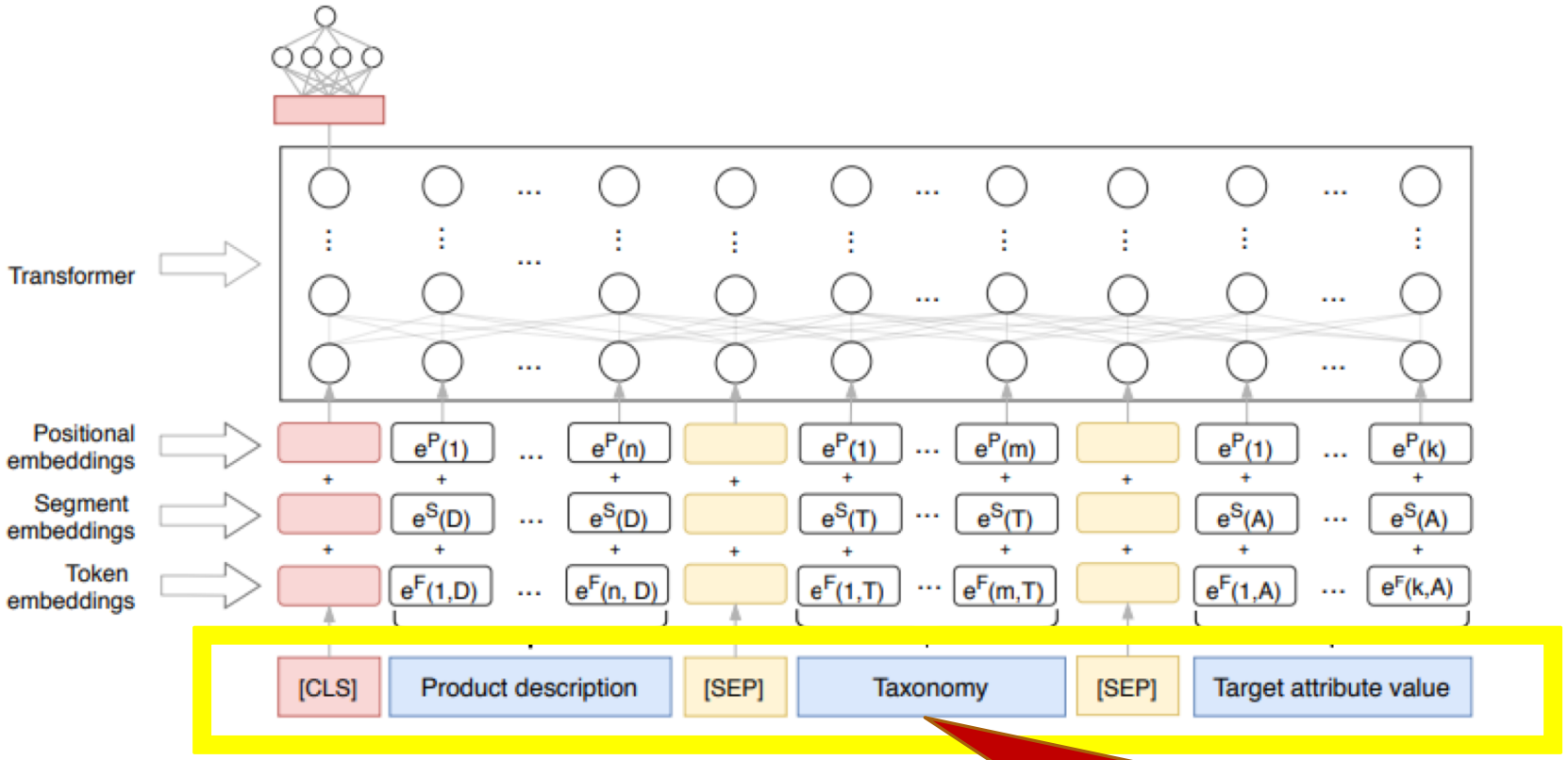
Wang et al., Automatic validation of textual attribute values in eCommerce Catalog by learning with limited labeled data, KDD'20

IV. Taxonomy-Aware Anomaly Detection



Wang et al., Automatic validation of textual attribute values in eCommerce Catalog by learning with limited labeled data, KDD'20

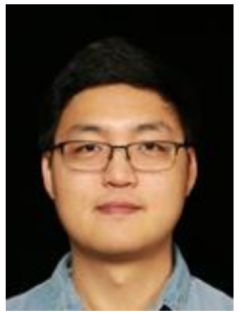
IV. Taxonomy-Aware Anomaly Detection



Category as input for model training

Dong et al., AutoKnow: Self-driving knowledge collection fo , 2020.

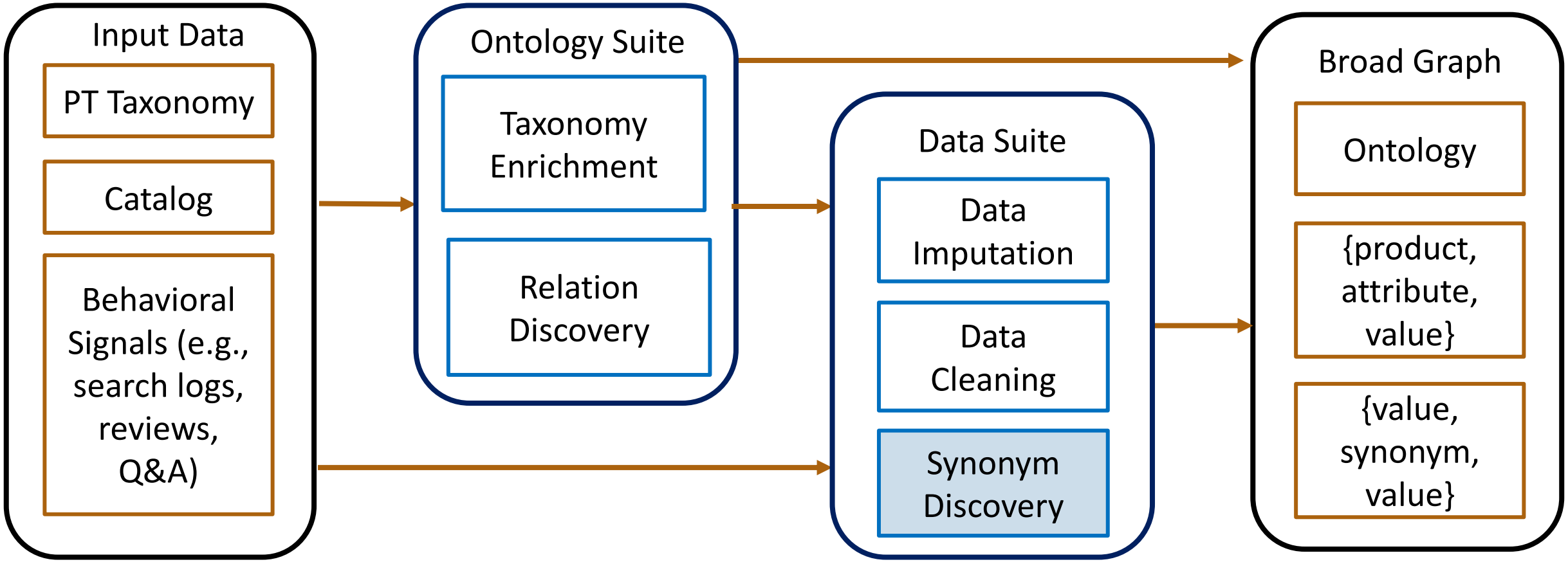
IV. Taxonomy-Aware Anomaly Detection



- Identify **1.77MM** incorrect values for Flavor and Scent for Consumables with **90% precision**

Product	Attr	Value
Love of Candy Bulk Candy - Pink Mint Chocolate Lentils - 6lb Bag	Flavor	Pink
Scott's Cakes Dark Chocolate Fruit & Nut Cream Filling Candies with Burgandy Foils in a 1 Pound Snowflake Box	Flavor	1 lb. snowflake box
Lucky Baby - Baby Blanket Envelope Swaddle Winter Wrap Coral Fleece Newborn Blanket Sleeper Infant Stroller Wrap Toddlers Baby Sleeping Bag (color 1)	Flavor	color 1
ASUTRA Himalayan Sea Salt Body Scrub Exfoliator + Body Brush (Vitamin C), 12 oz | Ultra Hydrating, Gentle, Moisturizing | All Natural & Organic Jojoba, Sweet Almond, Argan Oils	Scent	vitamin c body scrub - 12oz & body brush
Folgers Simply Smooth Ground Coffee, 2 Count (Medium Roast), 31.1 Ounce	Scent	2Packages (Breakfast Blend, 31.1 oz)

AutoKnow: Self-Driving Product Knowledge Collection



V. Synonym Discovery

- Identify **7.6K** synonym pairs for flavor and scent with **92.6% precision**

	Examples
Type synonyms	spices : seasonings / spices : herbs / flowers : live plants / drink : beverage / pasta : noodles / icing : frosting / beans : peas / foods : meals / salad dressings : condiments / nut : seeds
Attribute synonyms	[Flavor] bbq : bar-b-que / pasta : linguine / sausage : chorizo / Seaweed : nori / latte : mocha / fresh mint : clean mint / supreme dark : extra dark / vanilla vortex : very vanilla
Phrase synonyms	cr2032 battery : lithium coin / toys 6 months : engage baby / minnie dress : short cocktail / mtm 50 : 50 caliber / coq10 200mg : energizing antioxidant

V. Synonym Discovery

- Identify **7.6K** synonym pairs for flavor and scent with **92.6% precision**

	Examples
Type synonyms	spices : seasonings / spices : herbs / flowers : live plants / drink : beverage / pasta : noodles / icing : frosting / beans : peas / foods : meals / salad dressings : condiments / nut : seeds
Attribute synonyms	[Flavor] bbq : bar-b-que / pasta : linguine / sausage : chorizo / Seaweed : nori / latte : mocha / fresh mint : clean mint / supreme dark : extra dark / vanilla vortex : very vanilla
Phrase synonyms	cr2032 battery : lithium coin / toys 6 months : engage baby / minnie dress : short cocktail / mtm 50 : 50 caliber / coq10 200mg : energizing antioxidant

Search log mining

Key Techniques in AutoKnow

Component / Techniques	*AK-Taxonomy*	*AK-Relations*	*AK-Imputation*	*AK-Cleaning*	*AK-Synonyms*
Graph structure	X		X		
Taxonomy signal			X	X	
Distant supervision	X		X	X	
Behavior information	X	X			X

Take Aways

- We need to scale up for 100K categories for building an authoritative product knowledge graph
- We are investigating techniques for self-driving end-to-end product understanding
- Graph Neural Network and Learning w. limited labels play an important role in various aspects

Thank You!