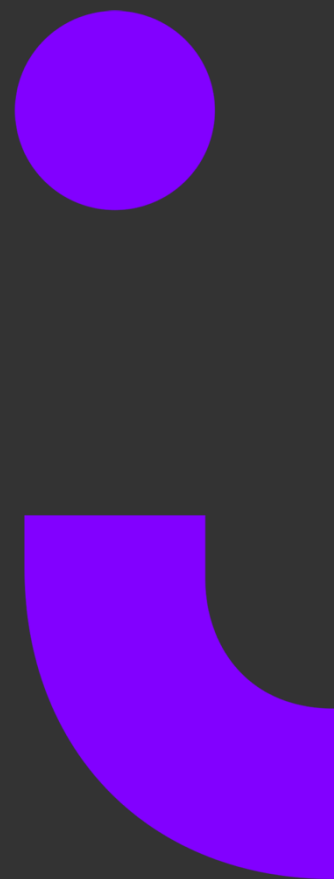




Towards a Simplified Ontology for Better eCommerce Search

Aliasgar Kutiyawala, Prateek Verma, Zheng (John) Yan
Walmart Labs



Motivation: Search with BOW Token Matching

Query:

Samsung LED TV

SKU 0:

Samsung LED TV

SKU 1:

Remote for Samsung LED LCD TV

SKU 2:

Samsung LCD TV

SKU 3:

VIZIO LED TV

Motivation: Search with Query & SKU Understanding

Query:

Samsung

BRAND

LED

DISPLAY TYPE

TV

PRODUCT

SKU 0:

Samsung

BRAND

LED

DISPLAY TYPE

TV

PRODUCT

SKU 1:

Samsung

BRAND

LCD

DISPLAY TYPE

TV

PRODUCT

SKU 2:

VIZIO

BRAND

LED

DISPLAY TYPE

TV

PRODUCT

SKU 3:

Remote

PRODUCT

for

PREPOSITION

Samsung

BRAND

LED

DISPLAY
TYPE

LCD

DISPLAY
TYPE

TV

ATTRIBUTE

Ontology

- One approach to query understanding is to use NER, which requires identification of named entities for the domain (eCommerce)
- A ontology is a formal explicit description of a domain by identifying classes (or concepts), slots and slot restrictions between classes for a particular domain



Named Entities for Search

- Product
 - Atomic entity that describes an item sold on the site
- Attribute
 - Atomic entity that provides more information about a product
- Brand
 - Entity that denotes the manufacturer of the item



Existing E-Commerce Ontologies Are Not Optimized for Search

- Supply Side Ontologies (ecl@ss, NAICS/NAPCS, etc):
 - Products (Concepts) like: “Shirts, underwear, men’s and boys’, cut and sewn from purchased fabric (except apparel contractors)”
 - Slots like: GLN of manufacturer, GLN of supplier, product article number of supplier
- Catalog Side Ontologies:
 - Baby
 - Baby Food and Formula
 - Toddler Juices and Milk
 - Baby and Toddler Snacks



Atomicity

What is the most basic term that describes the product of this SKU?

- Wooden Dining Chair
- Dining Chair
- Wooden Chair
- Chair



Atomicity Enables Easier NER

What is the most basic term that describes the product (or attribute) of this SKU?



Chair

Query: Wooden Dining Chair

- Wooden MATERIAL
- Dining FEATURE
- Chair PRODUCT

Query: Wooden Chair for dining table

- Wooden MATERIAL
- Chair PRODUCT
- for PREP
- dining FEATURE
- table FEATURE

Dining Chair

Query: Wooden Dining Chair

- Wooden MATERIAL
- Dining Chair PRODUCT

Query: Wooden Chair for dining table

- Wooden MATERIAL
- Chair ?
- for PREP
- dining ?
- table FEATURE

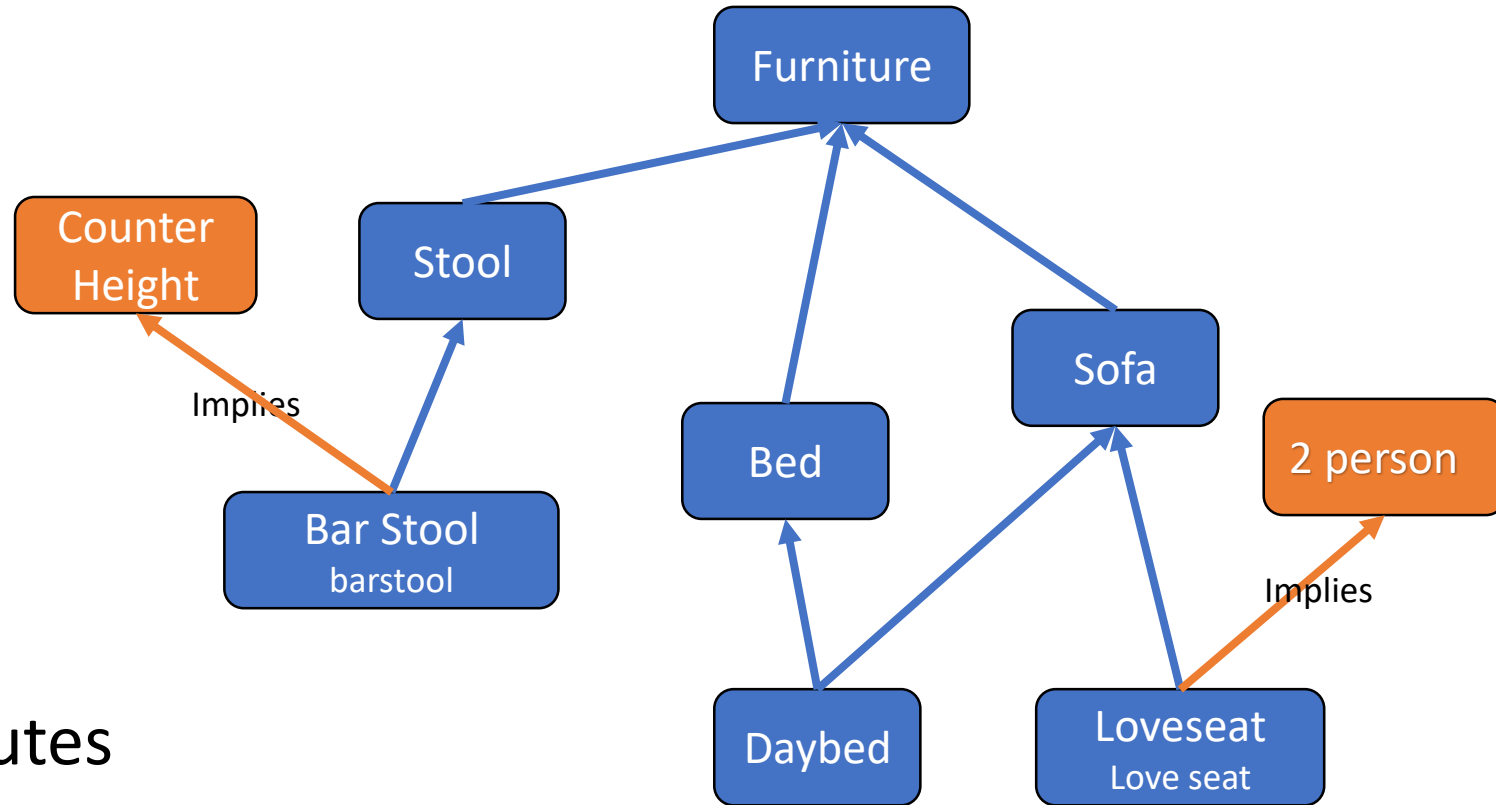
Atomicity

- Atomic does not imply a single word
 - Air Conditioner
- Have to make exceptions
 - Entities have single word variations:
Bar stool = barstool
 - Don't want some items to show up in the recall: Dress vs Wedding Dress



Product

- Is sold on the site
 - TV's are OK
 - Baby is NOT
- Atomic
- Can have Synonyms
 - TV = Television
- Can have Parents + Attributes
 - Bar stool = Stool + Counter Height



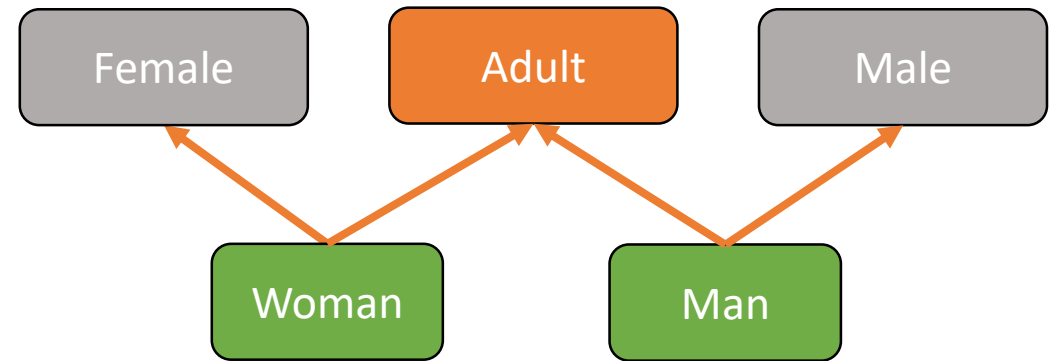
Attributes / Features

- Provides more information about a product
- Atomic
- Can have synonyms
 - waterproof = water proof
- Can have parents
 - Windows 10 = Windows



Attributes / Features

- General
 - waterproof
- Special
 - Resolution, Operating System
- Mutually Exclusive
 - Neck Type, Sleeve Type
- Primary
 - Color, Material, Flavor
- Numeric
 - 1 inch = 25.4 mm
- Negative
 - Sleeveless = No Sleeves



Brands

- Describes who makes the product
- Has synonyms
 - Samsung = Samsung Electronics
- Can have a default product
 - Kleenex = Tissues



Putting it All Together

Calvin Klein
BRAND

Black
COLOR

Dress
PRODUCT

for
PREP

Women
GENDER: FEMALE
AGE: ADULT

without
PREP

Sleeves
SLEEVE TYPE
(NEG)



Calvin Klein Black Women's Size 12 Studded Crewneck Sheath Dress



Calvin Klein Womens Crepe Sleeveless Wear to Work Dress



Calvin Klein Womens Halter Sleeveless Cocktail Dress

Finding Products for E-Commerce Categories

- Manual Reviewers
 - Not scalable
 - Have to train reviewers
 - Prone to errors
- Data Driven Approach
 - Scalable
 - Reduces errors
 - Need a small team of experts to manually review candidates

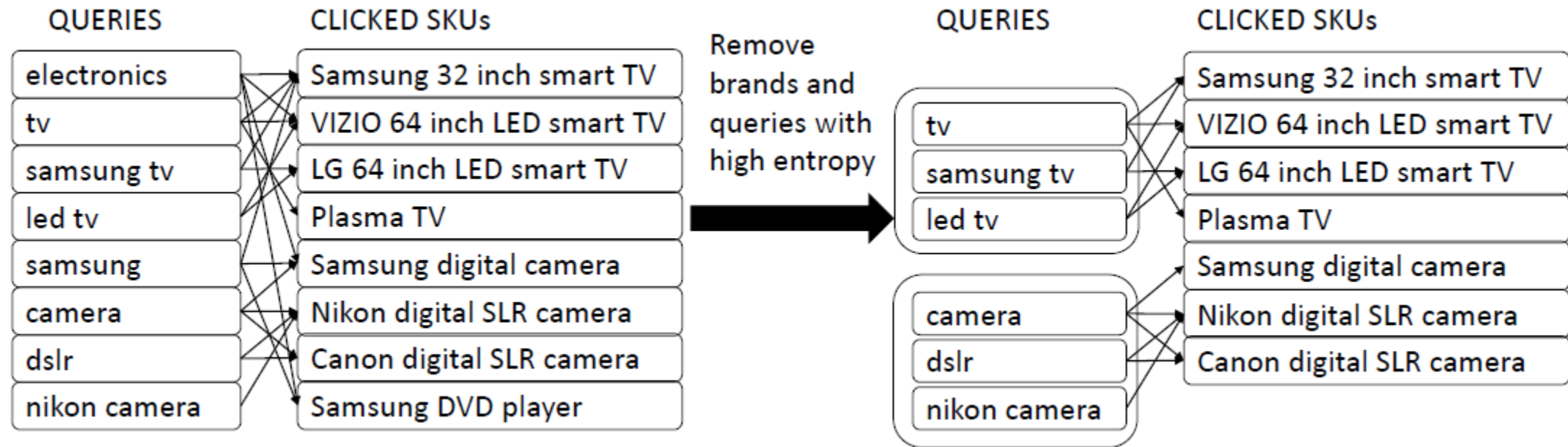


Finding Products Automatically

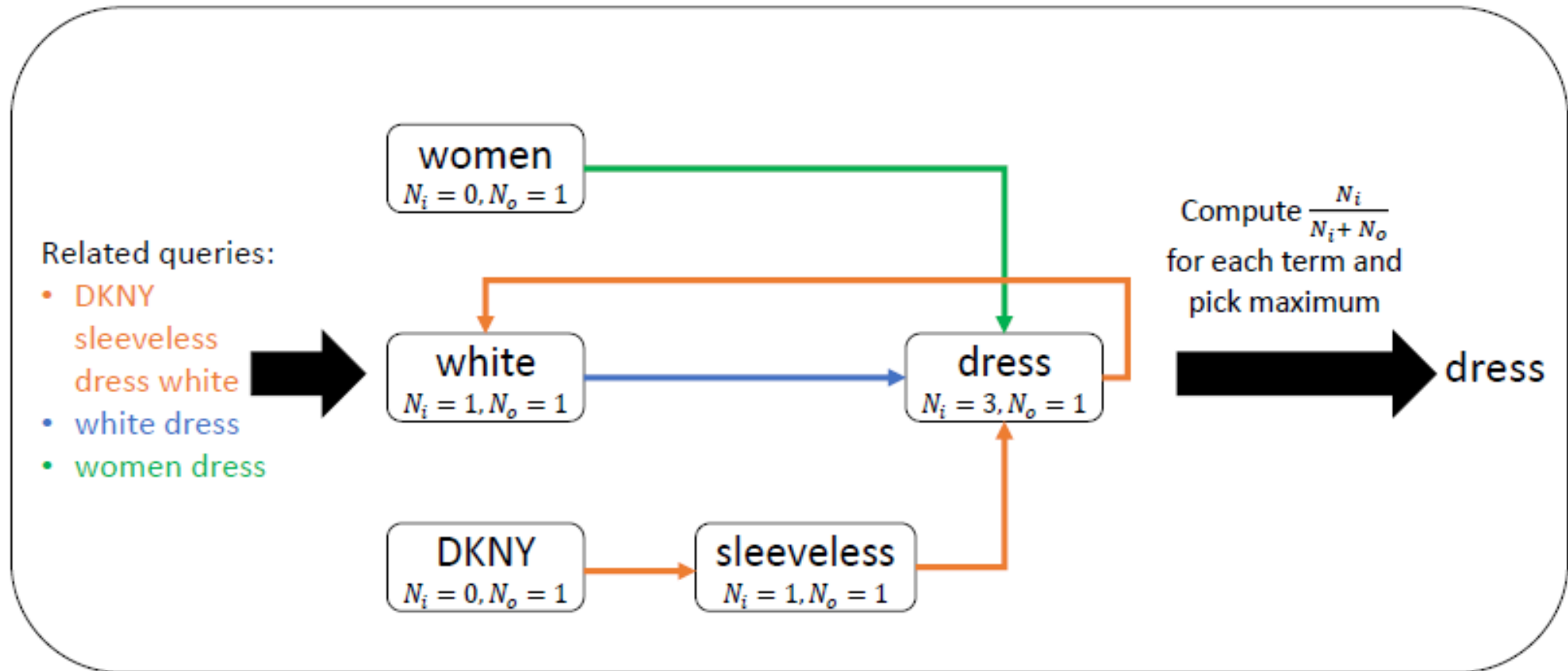
- Token Graph Method
- LSTM-CRF Tagger
- Augmented Graph Method



Grouping Similar Queries

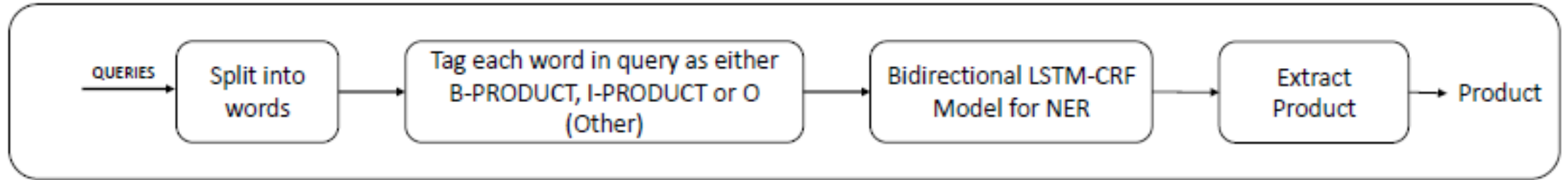


Finding Products: Token Graph Method



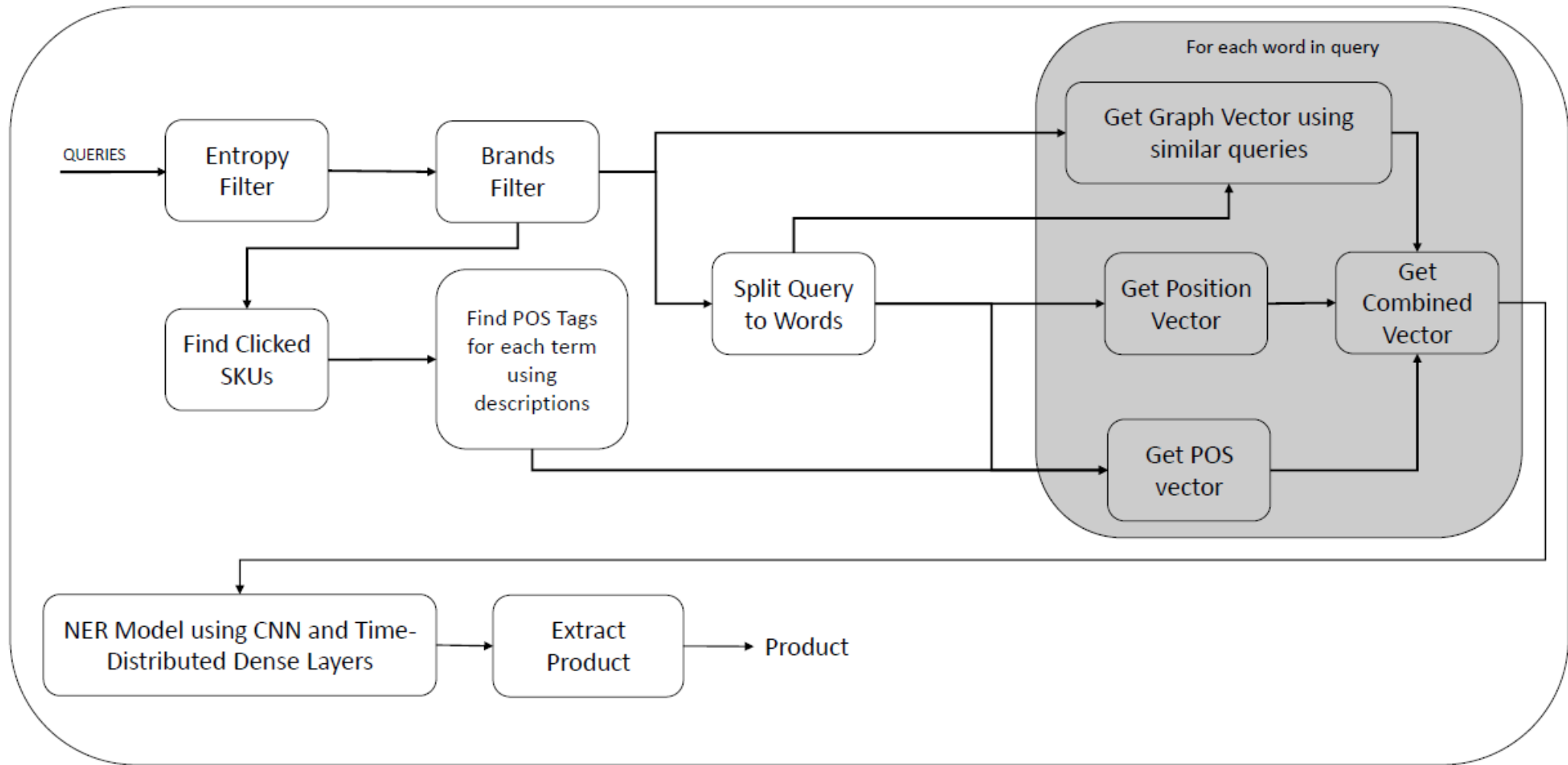
Intuition: Most queries end with product terms
Can filter out known attributes and brands

Finding Products: LSTM-CRF Model



- Training Data: Automatically labelled queries from six categories (Electronics, Women's Clothing, Men's Clothing, Kid's Clothing, Furniture and Home)
- Testing Data: Queries from the Baby Category
- Word2Vec with Wikipedia and SKU Data from Catalog

Finding Products: Augmented Graph Method

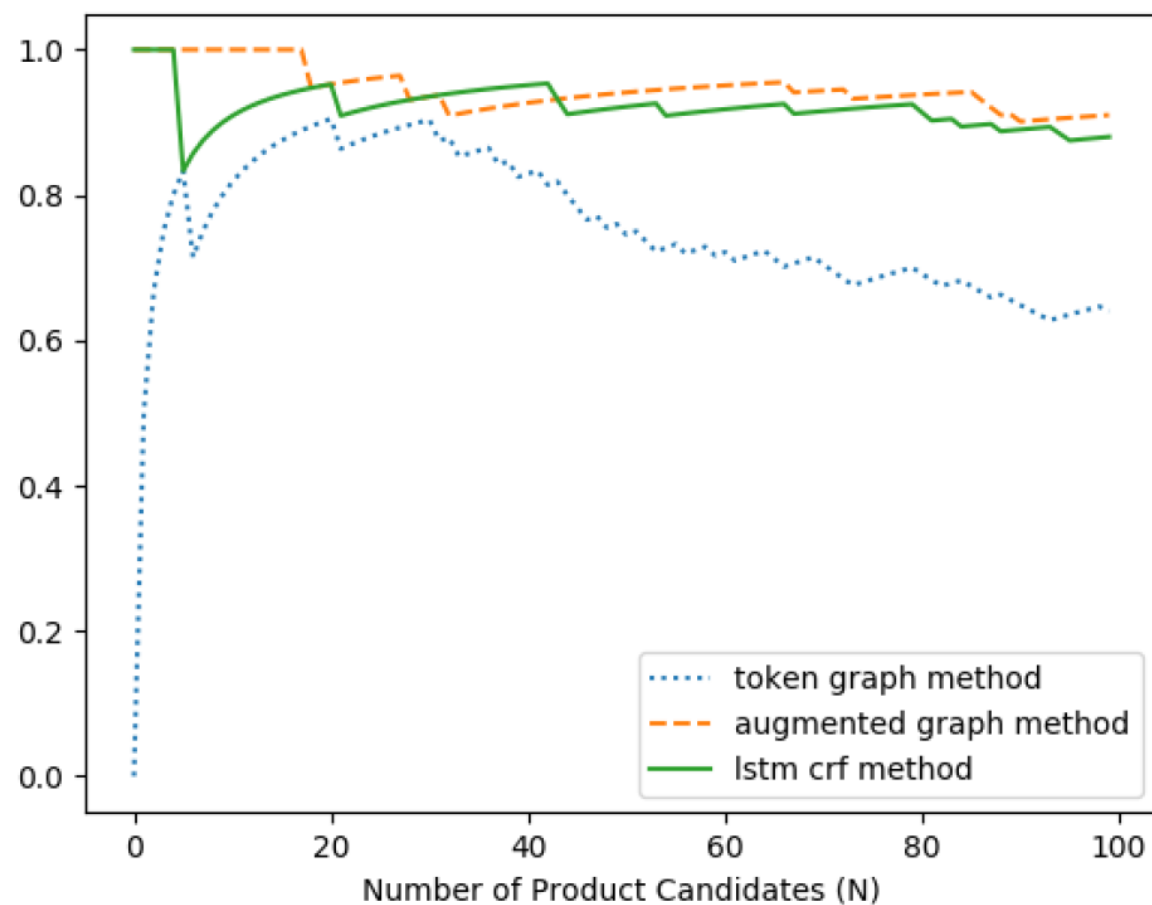
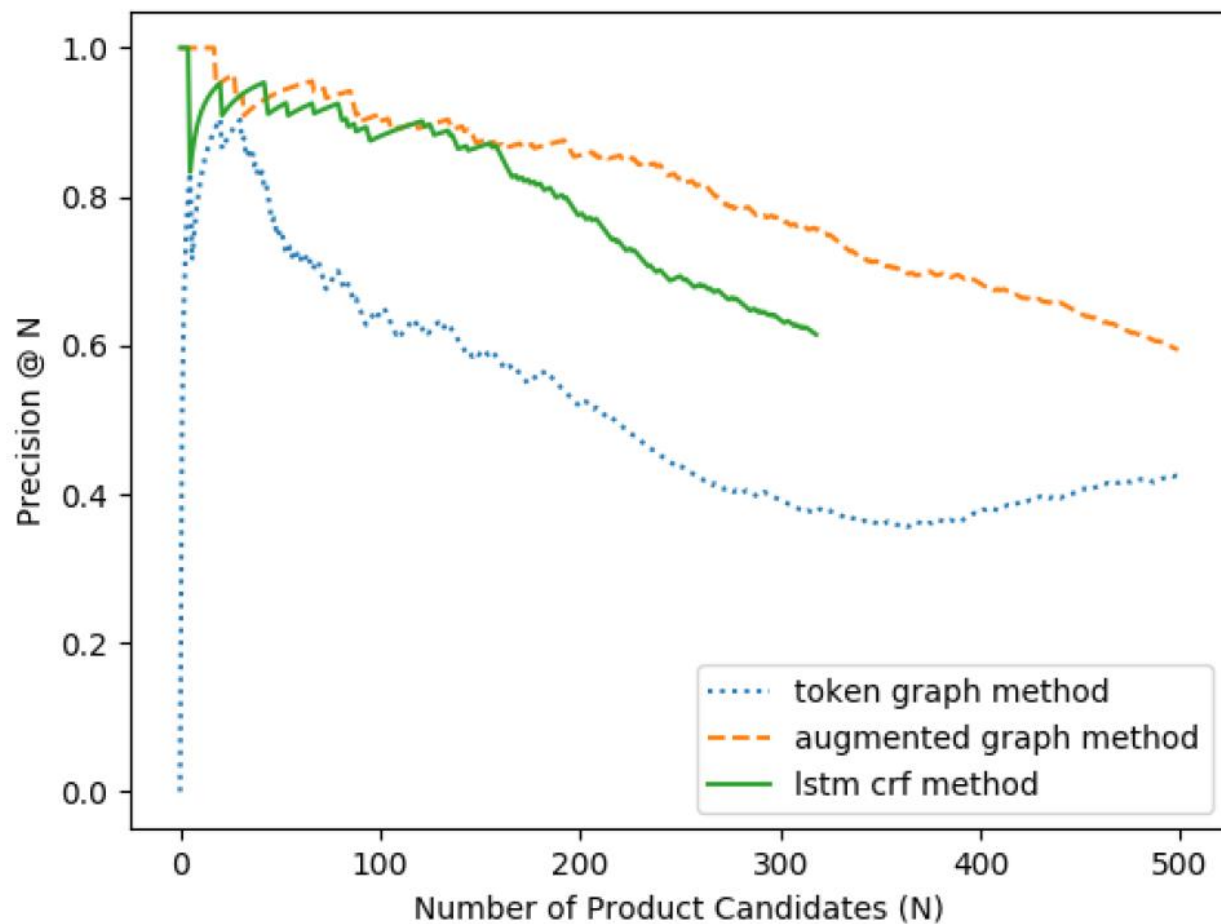


Finding Products: Augmented Graph Method

- Intuition: Add in POS tags to token graph method, remove last word assumption
- Training Data: Automatically labelled queries from six categories (Electronics, Women's Clothing, Men's Clothing, Kid's Clothing, Furniture and Home)
- Testing Data: Queries from the Baby Category
- Word2Vec with Wikipedia and SKU Data from Catalog



Results



Results (Top 10)

Num	Graph Method	Augmented Graph Model	NER Model
1	all (N)	diaper (P)	diaper (P)
2	sippycup (P)	wipe (P)	wipe (P)
3	cup (P)	formula (P)	bottle (P)
4	bib (P)	carseat (P)	bag (P)
5	playard (P)	bottle (P)	cover (P)
6	insert (P)	stroller (P)	ups (N)
7	ct (N)	bag (P)	pants (P)
8	highchair (P)	gate (P)	seat (P)
9	case (P)	cereal (P)	pad (P)
10	stroller (P)	highchair (P)	bib (P)

P denotes a *product* and N *not a product*