

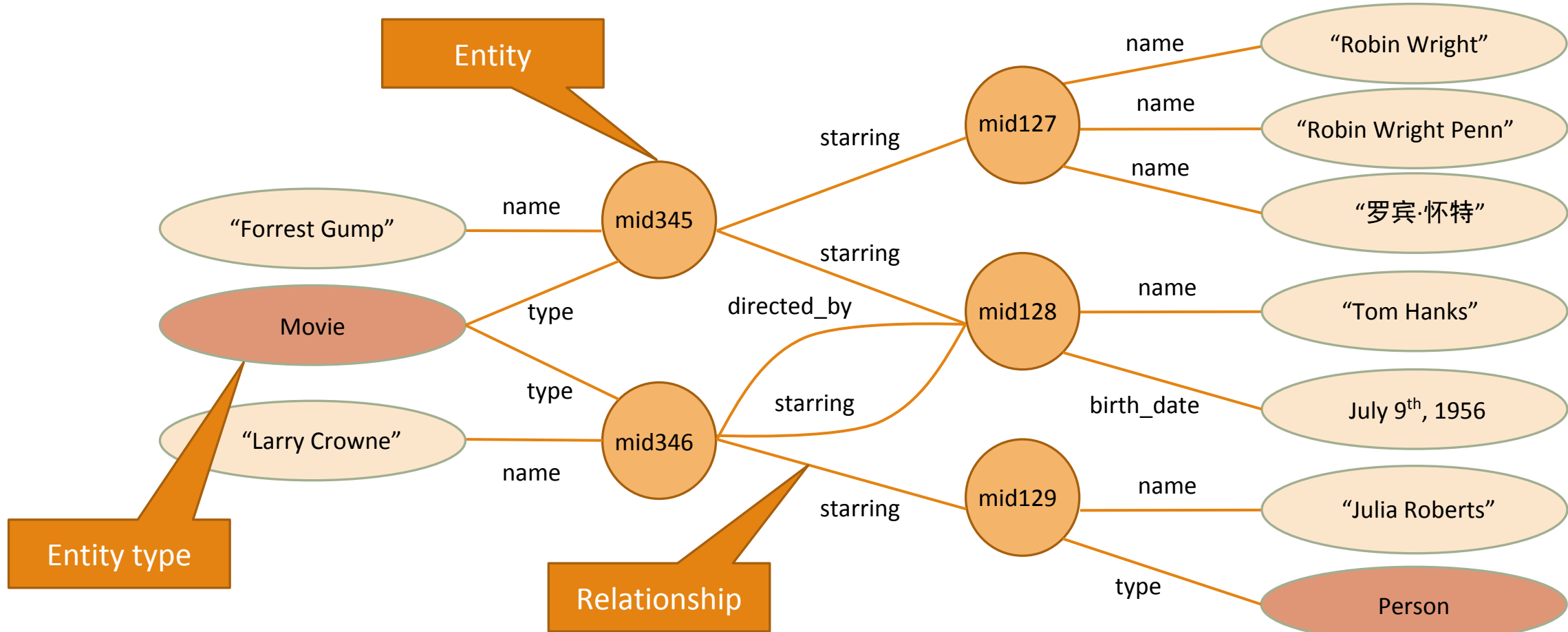
Challenges and Innovations in Building a Product Knowledge Graph

XIN LUNA DONG, AMAZON

JANUARY, 2018


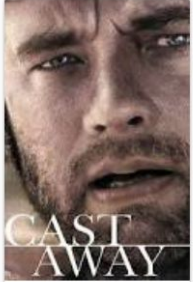







Product Graph vs. Knowledge Graph

Knowledge Graph Example for 2 Movies



Knowledge Graph in Search

Tom Hanks › Movies

 <p>Forrest Gump 1994</p>	 <p>Cast Away 2000</p>	 <p>Saving Private Ryan 1998</p>	 <p>Captain Phillips 2013</p>	 <p>Big 1988</p>	 <p>Sully 2016</p>	 <p>Toy Story 1995</p>	 <p>Catch Me If You Can 2002</p>	 <p>Apollo 13 1995</p>
--	---	--	--	---	---	---	---	---

List of Tom Hanks performances - Wikipedia

https://en.wikipedia.org/wiki/List_of_Tom_Hanks_performances ▼

Jump to **Film** - The Simpsons Movie, 2007, Yes, Himself, Cameo Voice role. Mamma Mia! 2008, Yes, —, Executive producer. **City of Ember**, 2008, Yes, —.

[A Hologram for the King \(film\)](#) · [Big \(film\)](#) · [Larry Crowne](#) · [He Knows You're Alone](#)

Tom Hanks (@tomhanks) · Twitter

<https://twitter.com/tomhanks>

And don't miss this songstress at the famous Cafe Carlyle. Through Saturday nite! Hanx @RitaWilson
<pic.twitter.com/J70XJbf...>

12 hours ago · [Twitter](#)


Beware! Crass self-serving Social Media Post! This book goes on sale tomorrow! Hanx
<pic.twitter.com/V2EqPKL...>

16 hours ago · [Twitter](#)

Lost (g)love. Looking for a mate. Good luck. Hanx.
<pic.twitter.com/ApH7rEG...>

1 day ago · [Twitter](#)

Tom Hanks
American actor



Thomas Jeffrey Hanks is an American actor and filmmaker. He is known for his various comedic and dramatic film roles, including Splash, Big, Turner & Hooch, A League of Their Own, Sleepless in Seattle, ... [Wikipedia](#)

Born: July 9, 1956 (age 61), [Concord, CA](#)

Awards: [Academy Award for Best Actor](#), [MORE](#)

Spouse: [Rita Wilson](#) (m. 1988), [Samantha Lewes](#) (m. 1978–1987)

TV shows: [Bosom Buddies](#), [Celebrity Jeopardy!](#), [MORE](#)

Knowledge Graph in Personal Assistant

Alexa, play the music
by Michael Jackson



List of officially released compilations and

[\[92\]](#)[\[93\]](#)[\[94\]](#)[\[95\]](#)

- *Portrait of Michael Jackson / Portrait of Jackson 5* (1973)
- *Os Grandes Sucessos, Vol. 2* (1980)
- *Motown Superstar Series, Vol. 7* (1980)
- *Superstar* (1980)
- *Michael Jackson & The Jackson 5* (1983)
- *Ain't No Sunshine* (1984)
- *The Great Love Songs of Michael Jackson* (1984)
- *Ben / Got to Be There* (1986)
- *Looking Back to Yesterday* (1986)
- *The Original Soul of Michael Jackson* (1987)
- *Rockin' Robin* (1993)
- *Dangerous – The Remix Collection* (1993)
- *Michael Jackson Story* (1996)
- *Master Series* (1997)
- *Ghosts – Deluxe Collector Box Set* (1997)
- *Got to Be There / Forever, Michael* (1999)
- *Bad / Thriller* (2000)
- *Forever, Michael / Music & Me / Ben* (2000)
- *Classic – The Universal Masters Collection* (2001)

Product Graph

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

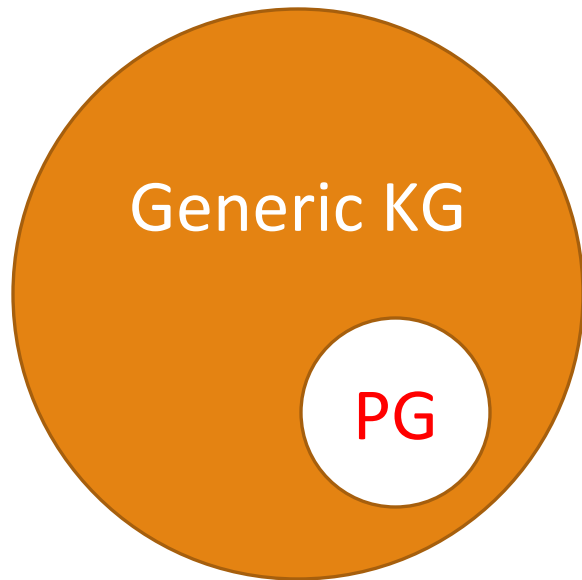


Customers who bought this item also bought

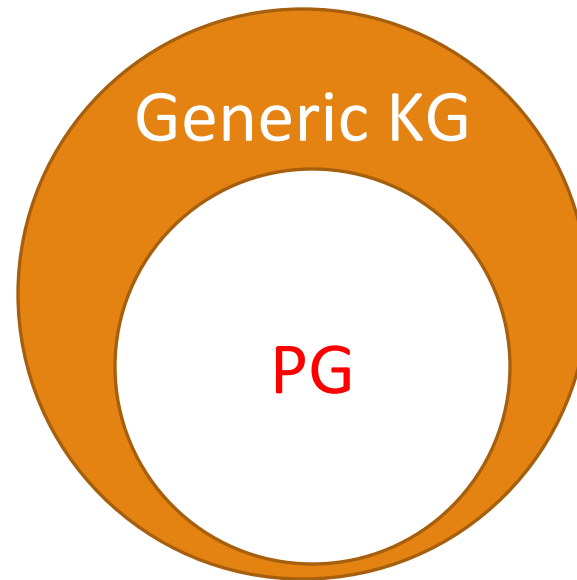


Product Graph vs. Knowledge Graph

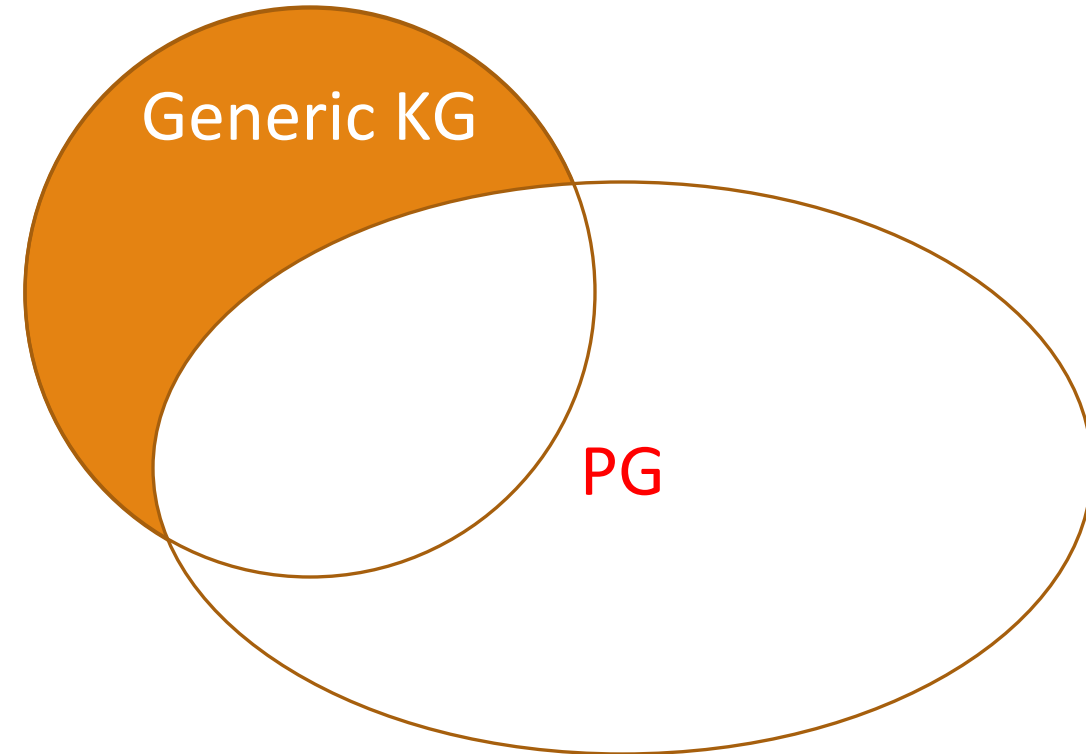
(A)



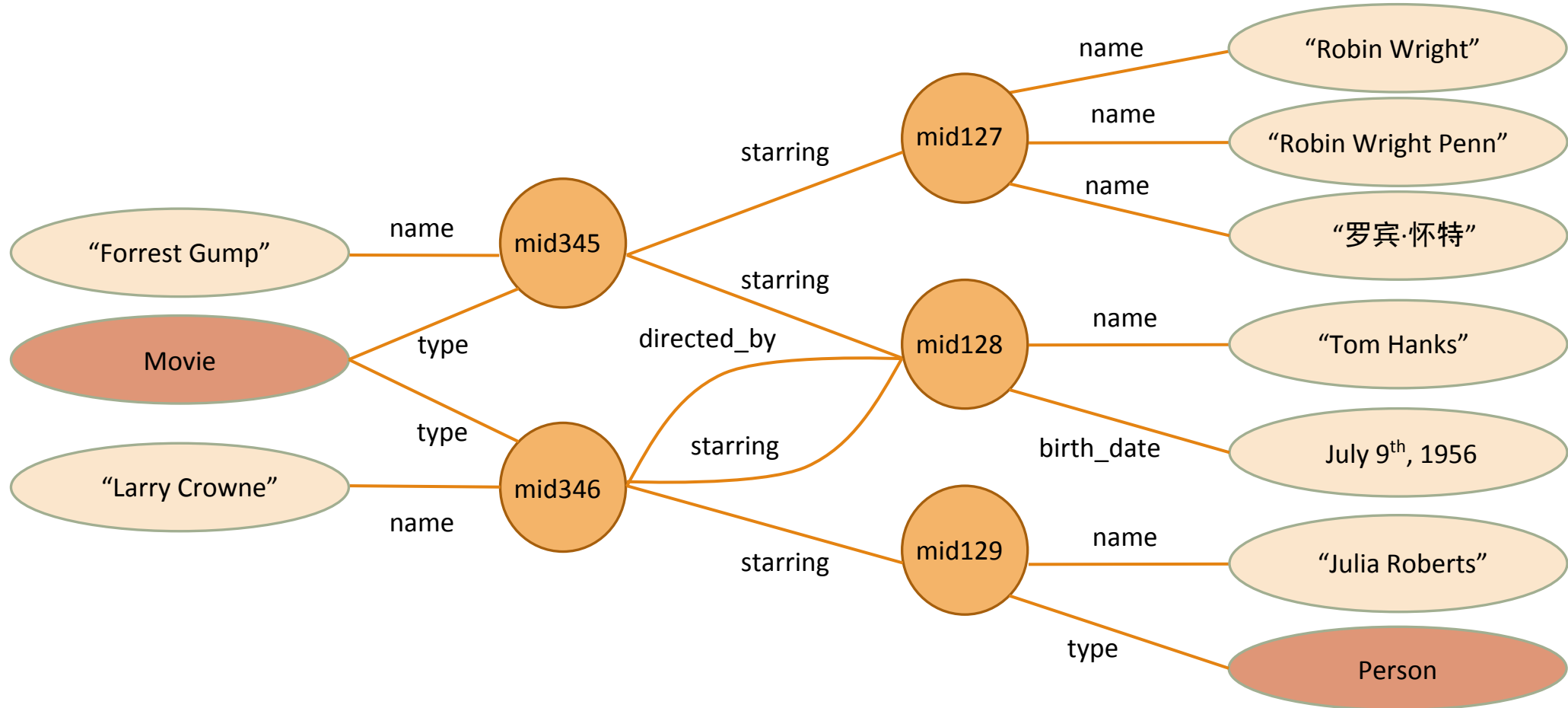
(B)



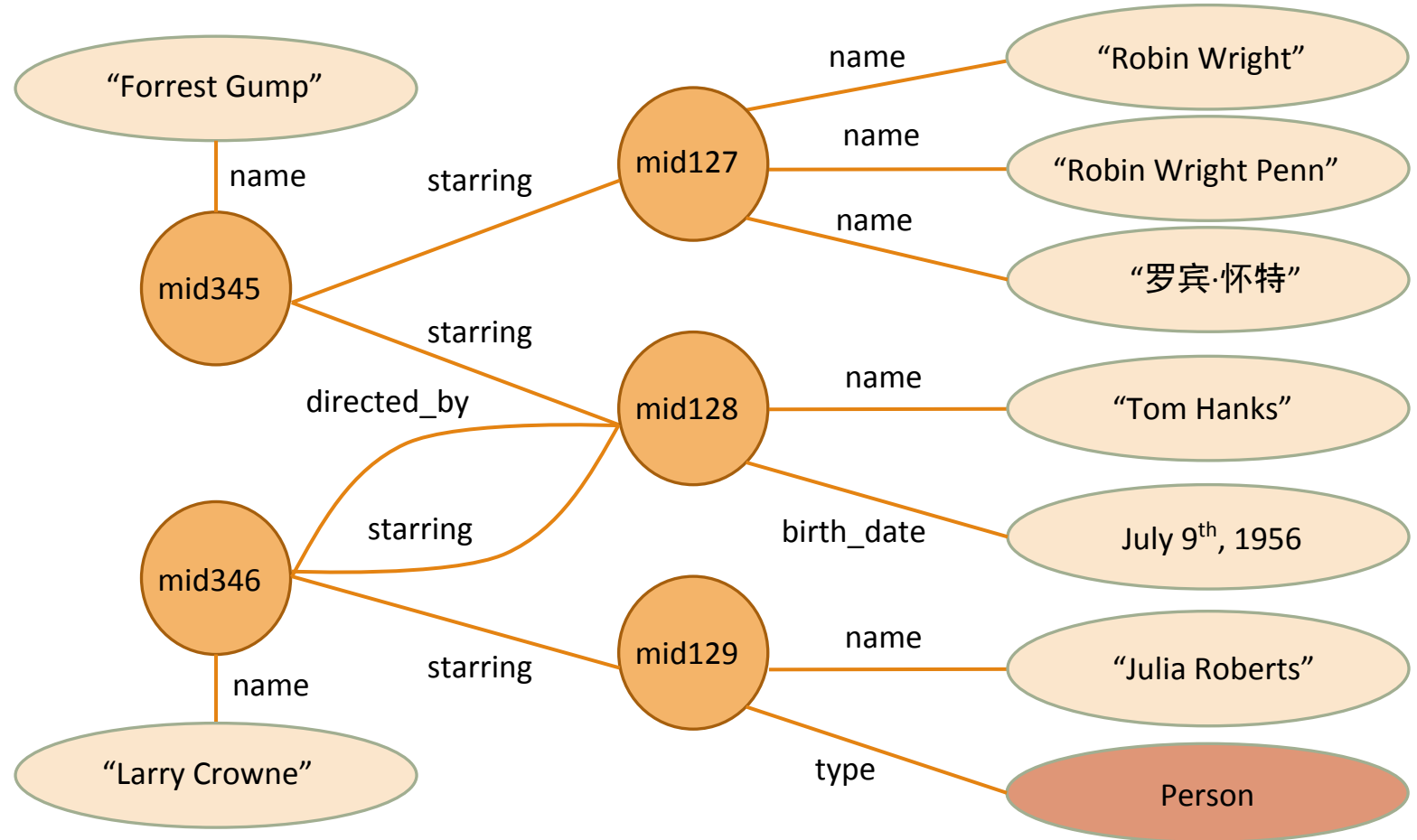
(C)



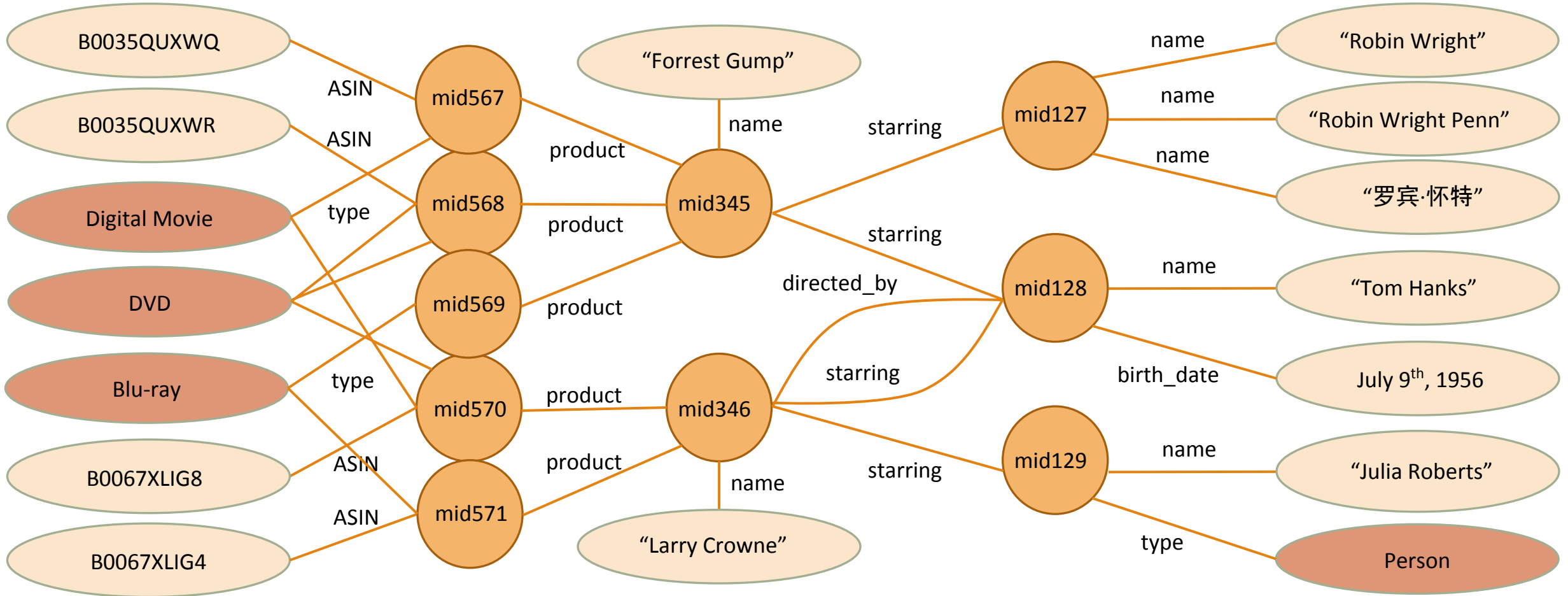
Knowledge Graph Example for 2 Movies



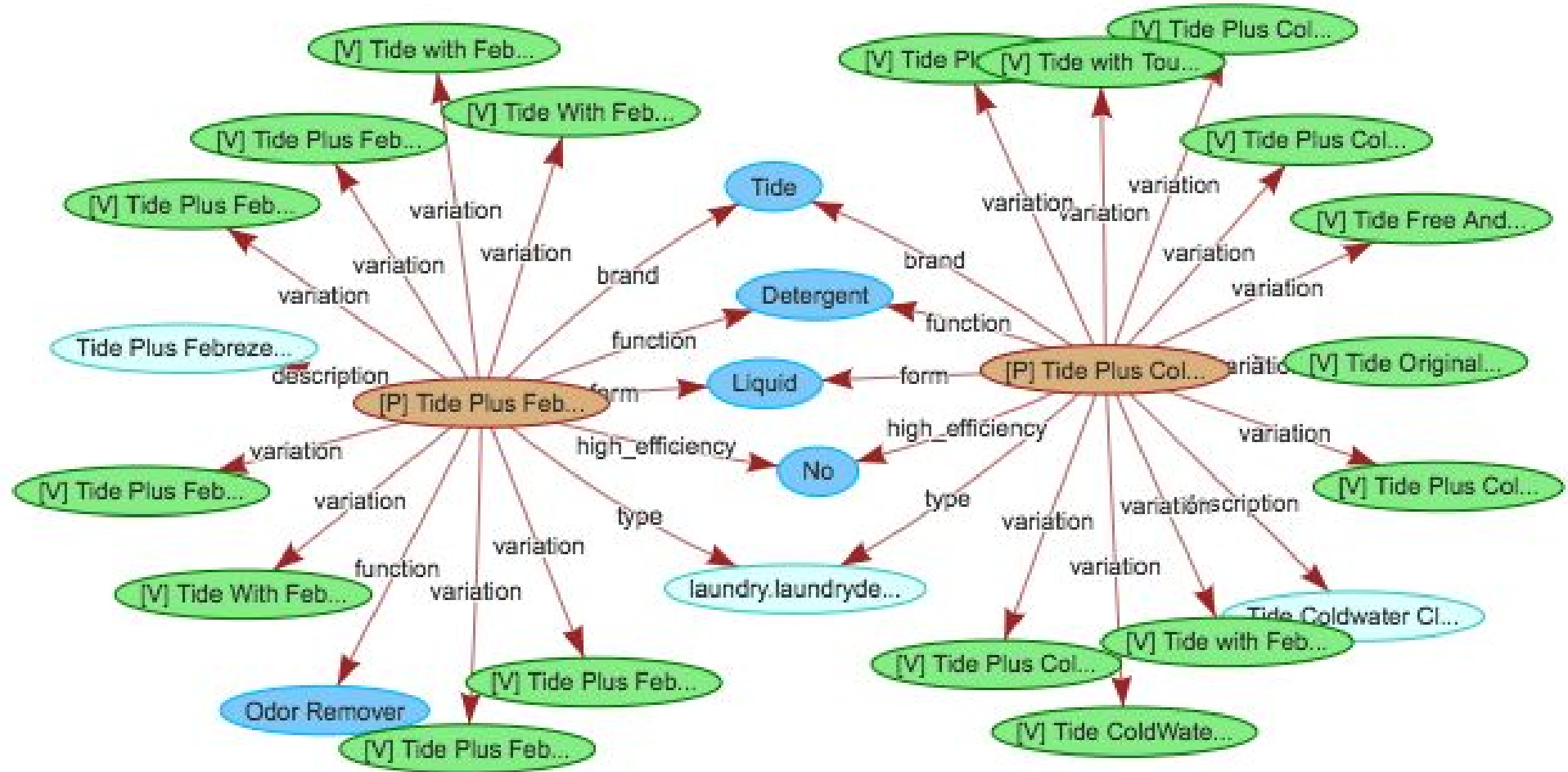
Product Graph vs. Knowledge Graph



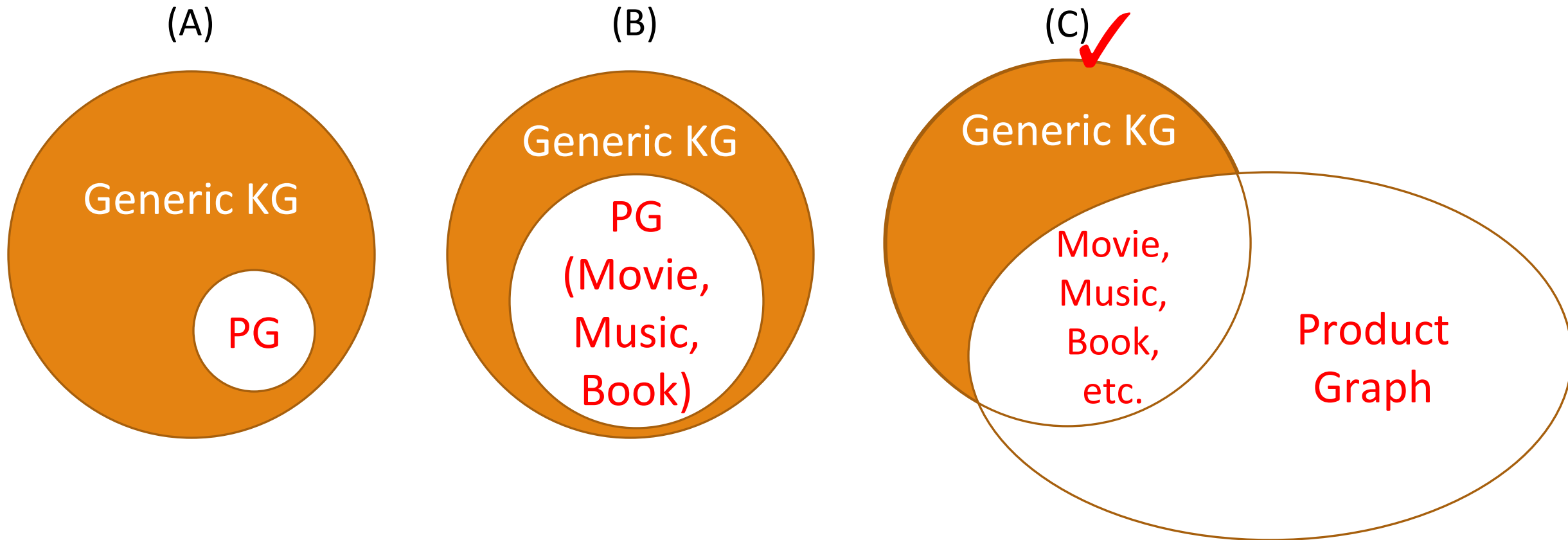
Product Graph vs. Knowledge Graph

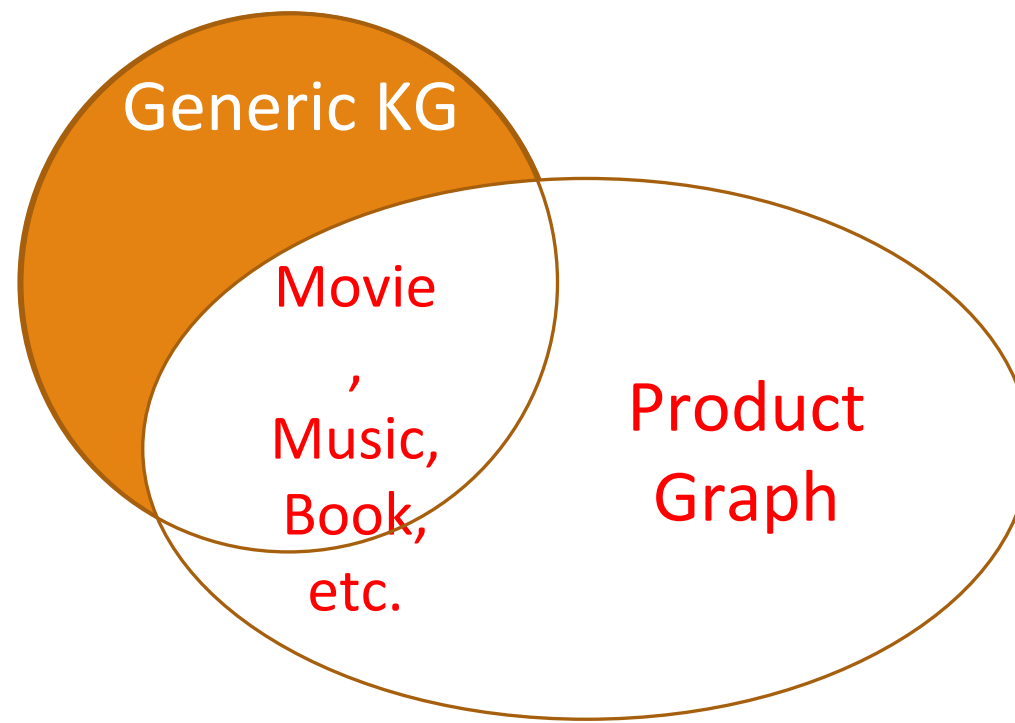


Another Example of Product Graph



Knowledge Graph vs. Product Graph





But, Is The Problem Harder?

Challenges in Building Product Graph I

- ❑ No major sources to curate product knowledge from
 - ❑ Wikipedia does not help too much
 - ❑ A lot of structured data buried in text descriptions in Catalog
 - ❑ Retailers gaming with the system so noisy data

Challenges in Building Product Graph II

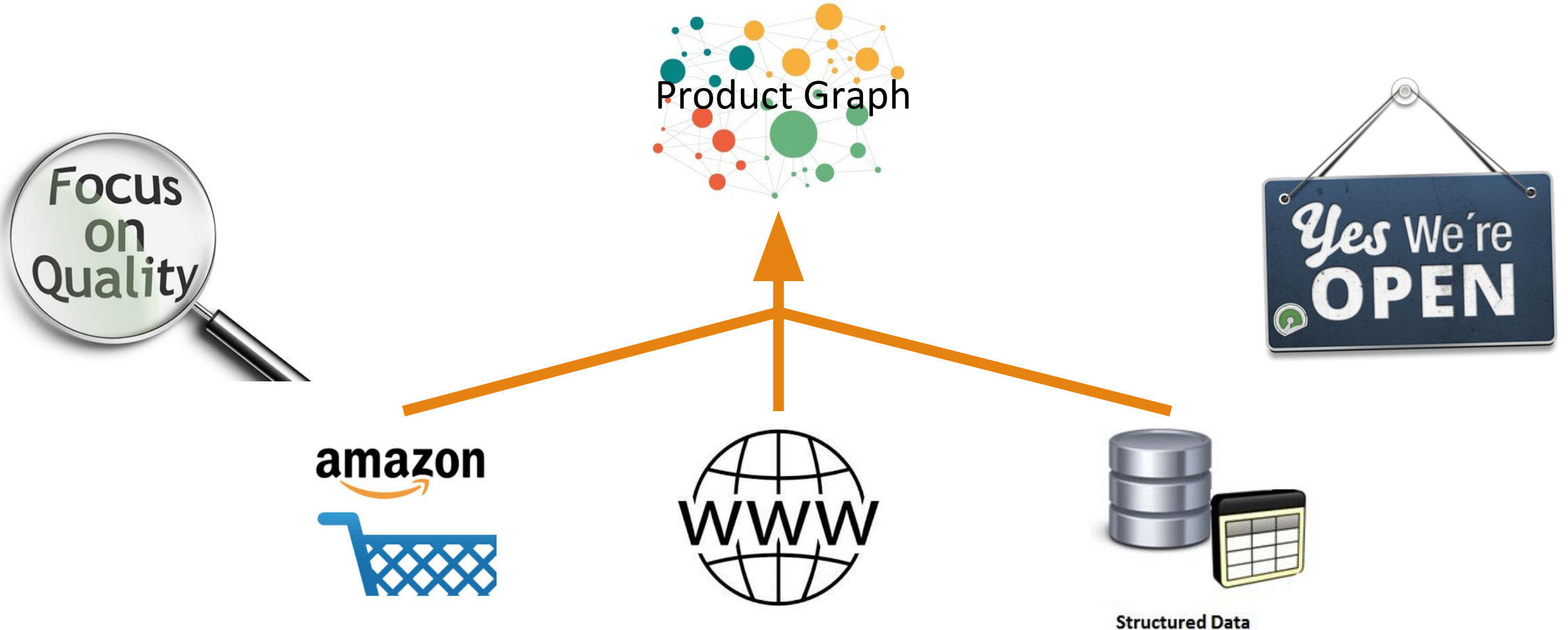
- ❑ Large number of new products everyday
 - ❑ Curation is impossible
 - ❑ Freshness is a big challenge

Challenges in Building Product Graph III

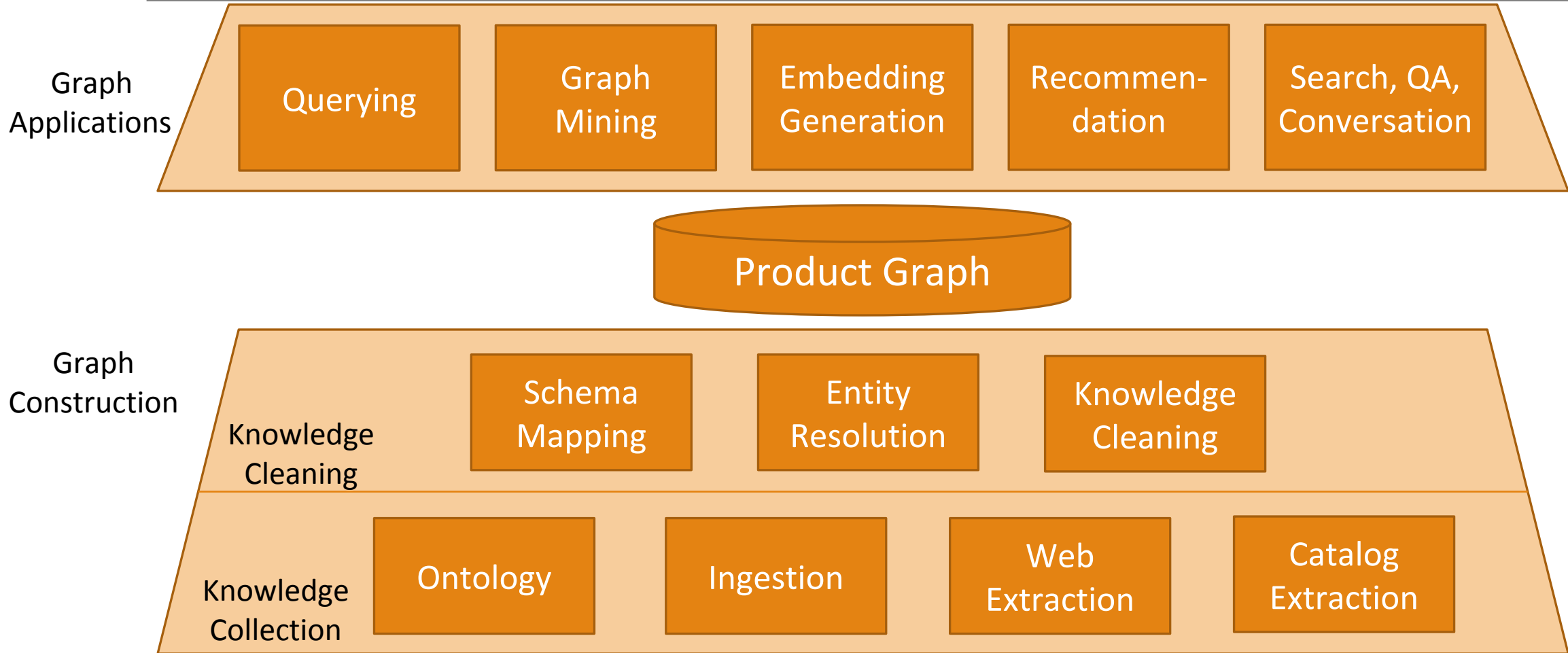
- ❑ Large number of product categories
 - ❑ A lot of work to manually define ontology
 - ❑ Hard to catch the trend of new product categories and properties

How to Build a Product Graph?

Where is Knowledge from?



Architecture



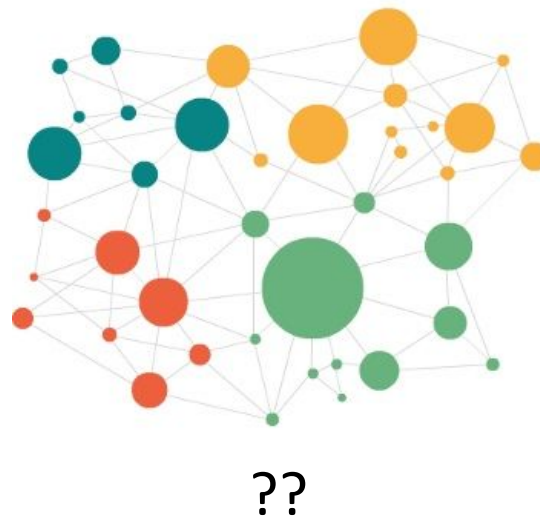
Which ML Model Works Best?



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."

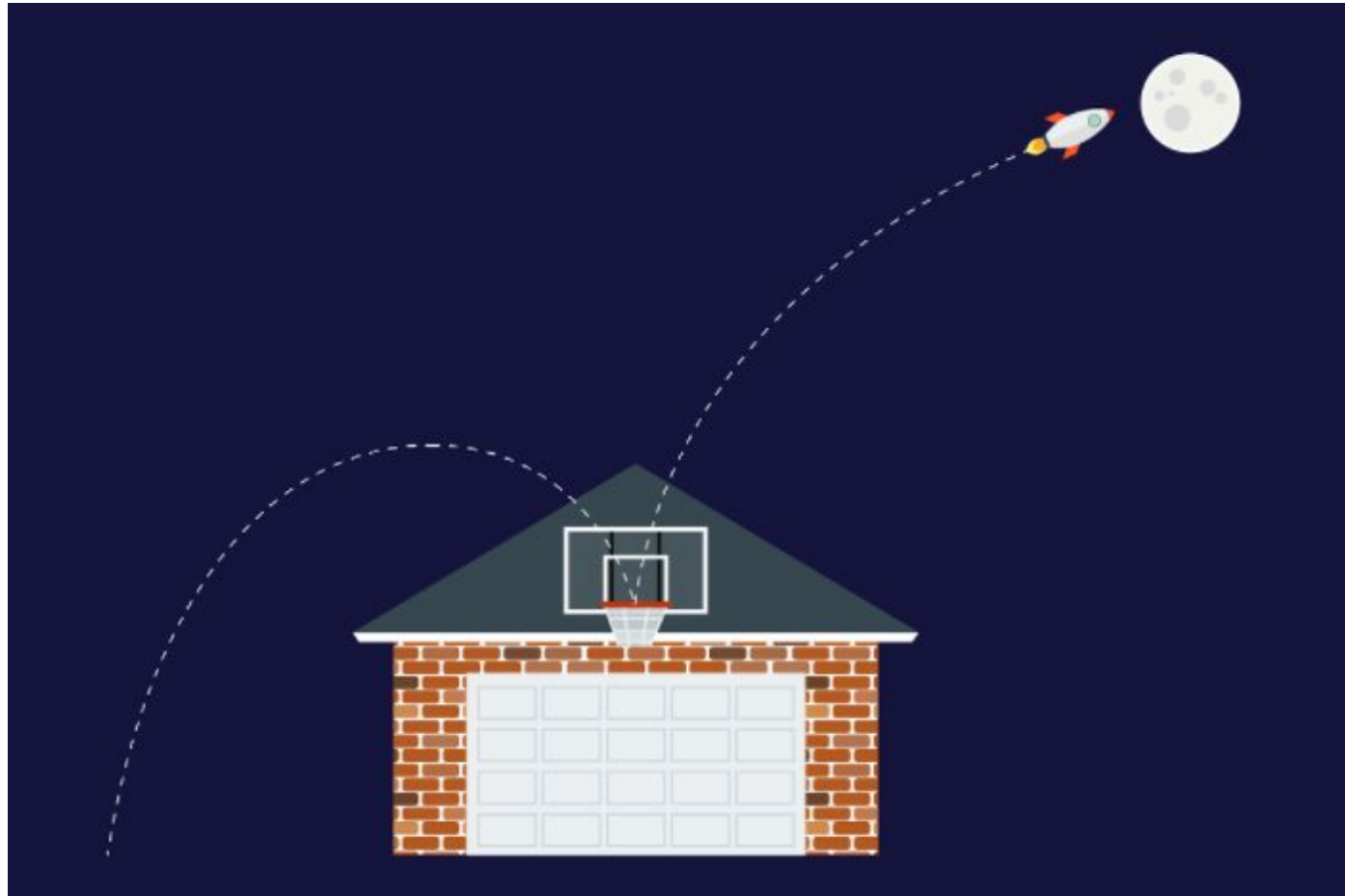
This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, 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 Veizin.



Neural network

Research Philosophy

Roofshots: Deliver incrementally and make production impacts




Moonshots: Strive to apply and invent the state-of-the-art

I. Extracting Knowledge from Semi-Structured Data on the Web

FULL CAST AND CREW | TRIVIA | USER REVIEWS

+ Top Gun (1986)

PG | 1h 50min | Action, Drama, Romance



Watch Now
From \$2.99 (SD) on Amazon Video

As students at the United States Navy's elite fighter class, one daring young pilot learns a few things from his instructor in the classroom.

Director: Tony Scott
Writers: Jim Cash, Jack Epps Jr. | 1 more credit
Stars: Tom Cruise, Tim Robbins, Kelly McGillis | 5 more

Metascore 50 From metacritic.com
Reviews 401 user | 173 critic



Aamir Khan is receiving rave reviews for Dangal.

Dangal
Cast: Aamir Khan, Sakshi Tanwar, Fatima Khurana, Sanya Malhotra
Director: Nitesh Tiwari
Rating: 4/5

卧虎藏龙 臥虎藏龍 (2000)



导演: 李安
编剧: 王蕙玲 / 詹姆斯·夏慕斯 / 蔡国荣
主演: 周润发 / 杨紫琼 / 章子怡 / 张震 / 郑佩佩 / 更多...
类型: 剧情 / 动作 / 爱情 / 武侠 / 古装
制片国家/地区: 台湾 / 香港 / 美国 / 中国大陆
语言: 汉语普通话
上映日期: 2000-10-13(中国大陆) / 2000-05-16(戛纳电影节) / 2000-07-07(台湾) / 2000-07-13(香港) / 2001-01-12(美国)
片长: 120 分钟
又名: Crouching Tiger, Hidden Dragon
IMDb链接: tt0190332



想看 看过 评价: ☆☆☆☆☆

写短评 写影评 + 提问题 分享到

推荐

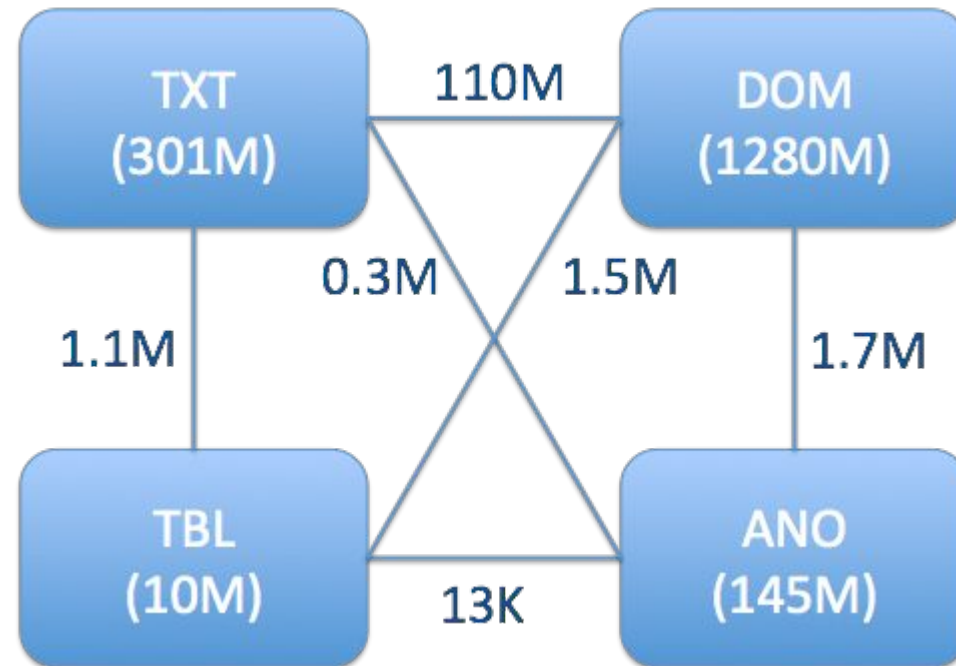
卧虎藏龙的剧情简介

一代大侠李慕白(周润发饰)有退出江湖之意,托付红颜知己俞秀莲(杨紫琼饰)将青冥剑转交给贝勒爷(郎雄饰)收藏,不料当夜遭玉娇龙(章子怡)窃取。俞秀莲暗中查访也大约知道是玉府小姐玉蛟龙所为,她想办法迫使玉蛟龙归还宝剑,免伤和气。但李慕白发现了害死师傅的碧眼狐狸(郑佩佩饰)的踪迹,她隐居于玉府并收玉蛟龙为弟子。而玉蛟龙欲以青冥剑来斩断阻碍罗小虎(张震饰)的枷锁,他们私定终身。关系变得错综复杂,俞秀莲和李慕白爱惜玉蛟龙人才难得,苦心引导,但玉蛟龙却使性任气不听劝阻..... ©豆瓣

I. Extracting Knowledge from Semi-Structured Data on the Web

Knowledge Vault @ Google showed big potential from DOM-tree extraction [Dong et al., KDD'14][Dong et al., VLDB'14]

Accu	Accu (conf $\geq .7$)
0.36	0.52

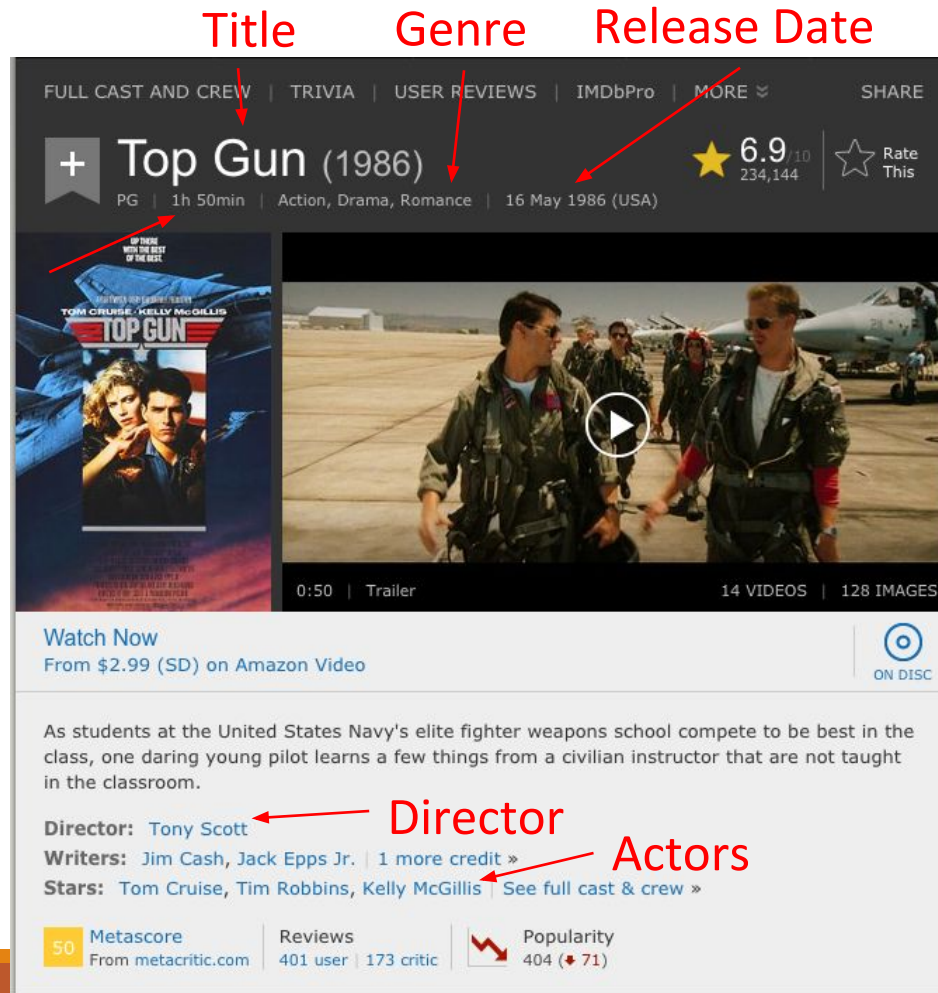


Accu	Accu (conf $\geq .7$)
0.43	0.63
0.09	0.62

I. Extracting Knowledge from Web—Annotation-Based DOM Extraction

Annotations on the IMDb page for **Top Gun (1986)**:

- Title**: Top Gun (1986)
- Genre**: Action, Drama, Romance
- Release Date**: 16 May 1986 (USA)
- Runtime**: 1h 50min
- Director**: Tony Scott
- Actors**: Tom Cruise, Tim Robbins, Kelly McGillis



Annotation-based knowledge extraction

Extracted relationships

- (Top Gun, type.object.name, "Top Gun")
- (Top Gun, film.film.genre, Action)
- (Top Gun, film.film.directed_by, Tony Scott)
- (Top Gun, film.film.starring, Tom Cruise)
- (Top Gun, film.film.runtime, "1h 50min")
- (Top Gun, film.film.release_Date_s, "16 May 1986")

I. Extracting Knowledge from Web—Annotation-Based DOM Extraction



Annotation-based
knowledge extraction



Alexa, When did Padme Amidala die?
What model is R2D2?
Who is Luke Skywalker's master?
Where is Boba Fett from?
Who is Darth Vader's apprentice?

I. Extracting Knowledge from Web—Distantly Supervised DOM Extraction

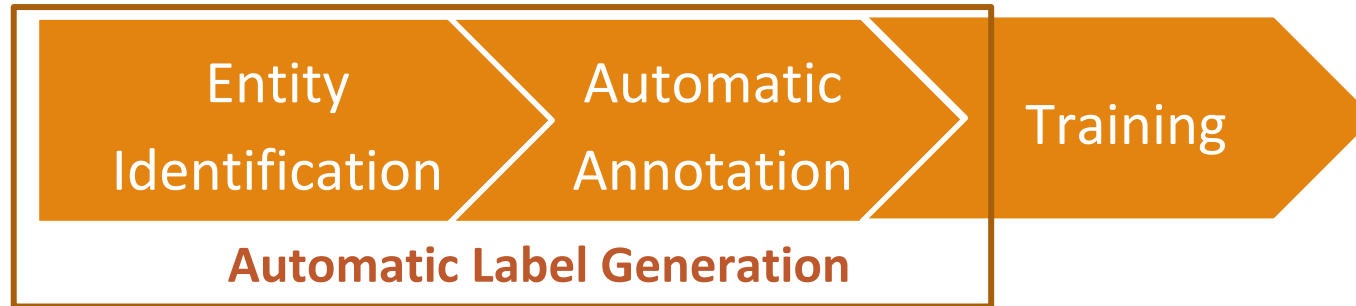


Annotation-based
knowledge extraction



Distantly supervised
web extraction

I. Extracting Knowledge from Web—Distantly Supervised DOM Extraction



Movie entity

Genre Release Date

Runtime

Director Actors

FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE

Top Gun (1986) ★ 6.9 ¹⁰ / 234,144

PG | 1h 50min | Action, Drama, Romance | 16 May 1986 (USA)

Watch Now
From \$2.99 (SD) on Amazon Video

As students at the United States Navy's elite fighter weapons school compete to be best in the class, one daring young pilot learns a few things from a civilian instructor that are not taught in the classroom.

Director: Tony Scott
Writers: Jim Cash, Jack Epps Jr. | 1 more credit »
Stars: Tom Cruise, Tim Robbins, Kelly McGillis | See full cast & crew »

FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE

Top Gun (1986) ★ 6.9 ¹⁰ / 234,144

PG | 1h 50min | Action, Drama, Romance | 16 May 1986 (USA)

Watch Now
From \$2.99 (SD) on Amazon Video

As students at the United States Navy's elite fighter weapons school compete to be best in the class, one daring young pilot learns a few things from a civilian instructor that are not taught in the classroom.

Director: Tony Scott
Writers: Jim Cash, Jack Epps Jr. | 1 more credit »
Stars: Tom Cruise, Tim Robbins, Kelly McGillis | See full cast & crew »

Extracted triples

- (Top Gun, type.object.name, "Top Gun")
- (Top Gun, film.film.genre, Action)
- (Top Gun, film.film.directed_by, Tony Scott)
- (Top Gun, film.film.starring, Tom Cruise)
- (Top Gun, film.film.runtime, "1h 50min")
- (Top Gun, film.film.release_Date_s, "16 May 1986")

I. Extracting Knowledge from Web—Distantly Supervised DOM Extraction

□ Extraction on IMDb

1. Very high extraction precision
2. Extracting triples with new entities

Predicate	Precision	Recall
Type.object.name (“name”)	1	1
People.person.place_of_birth	1	1
Common.topic.alias	1	1
Film.actor.film	0.98	0.47
Film.director.film	0.98	0.91
Film.producer.film	0.89	0.57
Film.writer.film	0.96	0.60

Predicate	Precision	Recall
Type.object.name (“title”)	0.97*	0.97*
Tv.tv_series_episode.episode_number	1	1
Tv.tv_series_episode.season_number	1	1
Film.film.directed_by	0.99	1
Film.film.written_by	1	0.98
Film.film.genre	0.90*	1
Film.film.starring	1	0.97
Tv.tv_series_episode.series	1	1

*Ground truth is incomplete. Manual inspection suggests close to 100% accuracy.

I. Extracting Knowledge from Web—Distantly Supervised DOM Extraction

□ Extraction experiments on <http://swde.codeplex.com/> (2011)

	Title		Director(s)		Genre(s)	
Site	P	R	P	R	P	R
allmovies	1	1	1	1	0.71	0.96
amctv	1	1	0.98	0.97	0.95	0.91
boxofficemojo	1	1	1	0.98	0.67*	0.91
hollywood	1	1	0.94	1	1	0.97
iheartmovies	1	1	1	1	1	1
IMDB	1	1	1	0.98	1	1
metacritic	1	1	1	1	1	1
MSN	1	1	1	1	1	1
rottentomatoes	1	1	1	1	1	0.91
yahoo	1	1	1	0.99	0.99	0.94

I. Distantly Supervised DOM Extraction

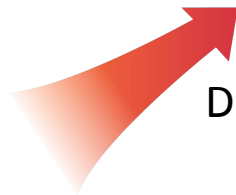
Which ML Model Works Best?

- ❑ Logistic regression: Best results (20K features on one website)
- ❑ Random forest: lower precision and recall

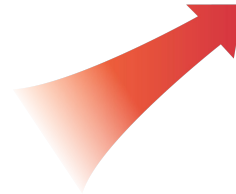
I. Extracting Knowledge from Semi-Structured Data on the Web



Annotation-based
knowledge extraction



Distantly supervised
web extraction



OpenIE
DOM extraction



Nearly-automatic
interactive extraction
on any new vertical

II. Extracting Knowledge from Product Profiles in Amazon Catalog

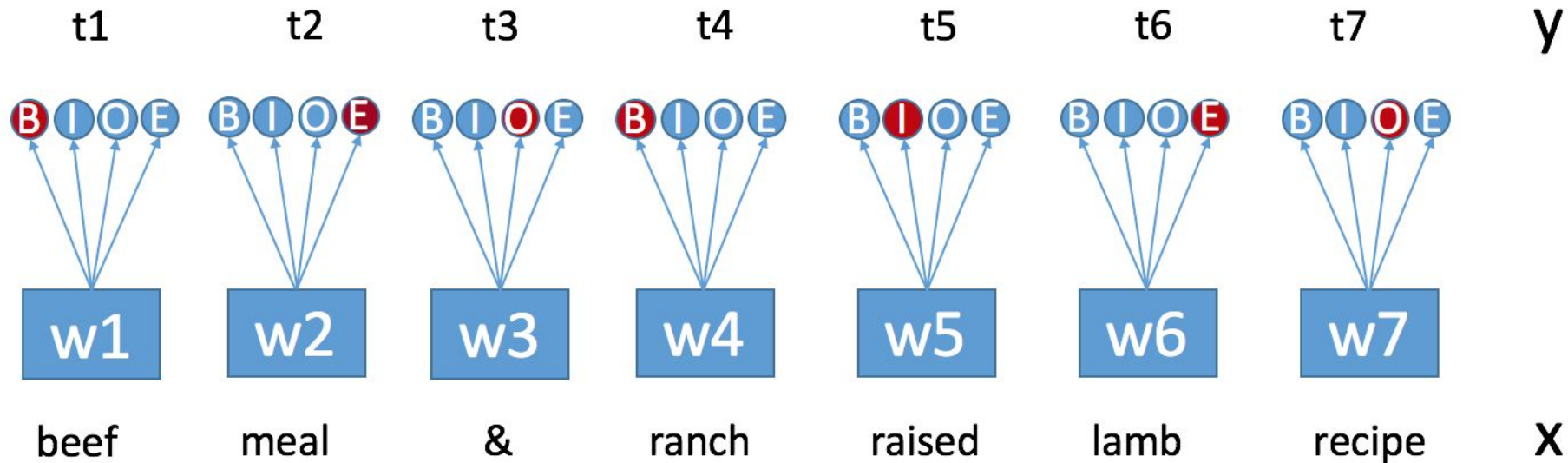
name	form	scent
Tide Detergent with Febreze Freshness		
Gain Apple Mango Tango Liquid Laundry Detergent		
Gain Joyful Expression Powder Detergent		
Tide PODS Original Scent HE Turbo Laundry Detergent Pacs 81-load Tub		
Tide PODS Free & Gentle HE Turbo Laundry Detergent Pacs 35-load Bag		

II. Open Attribute Extraction by Named Entity Recognition

- B** Beginning of entity
- I** Inside of entity
- O** Outside of entity
- E** End of entity

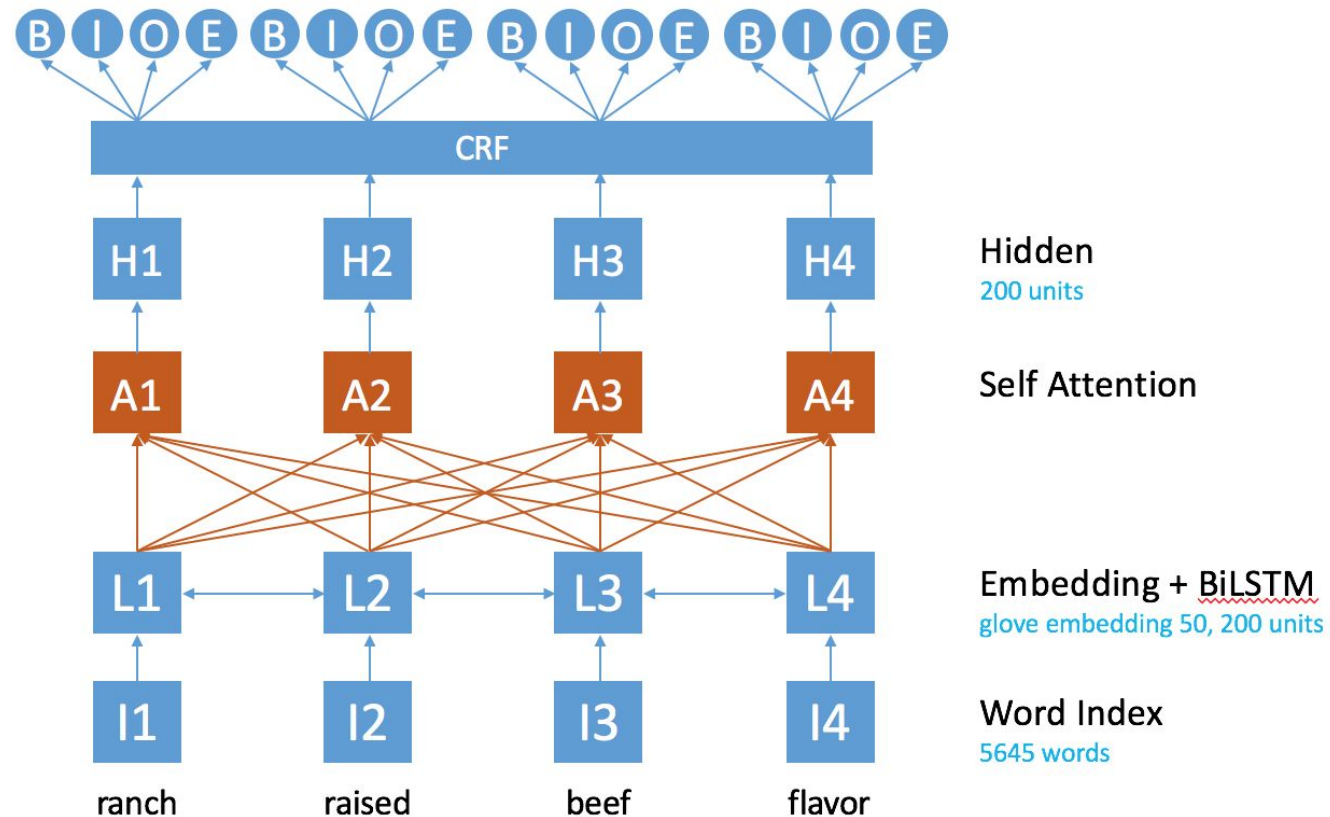
$x = \{w_1, w_2, \dots, w_n\}$ input sequence

$y = \{t_1, t_2, \dots, t_n\}$ tagging decision



II. Open Attribute Extraction by NER —Which ML Model Works Best?

□ Recurrent Neural Network, CRF, Attention



II. Open Attribute Extraction by NER —Adding Active Learning

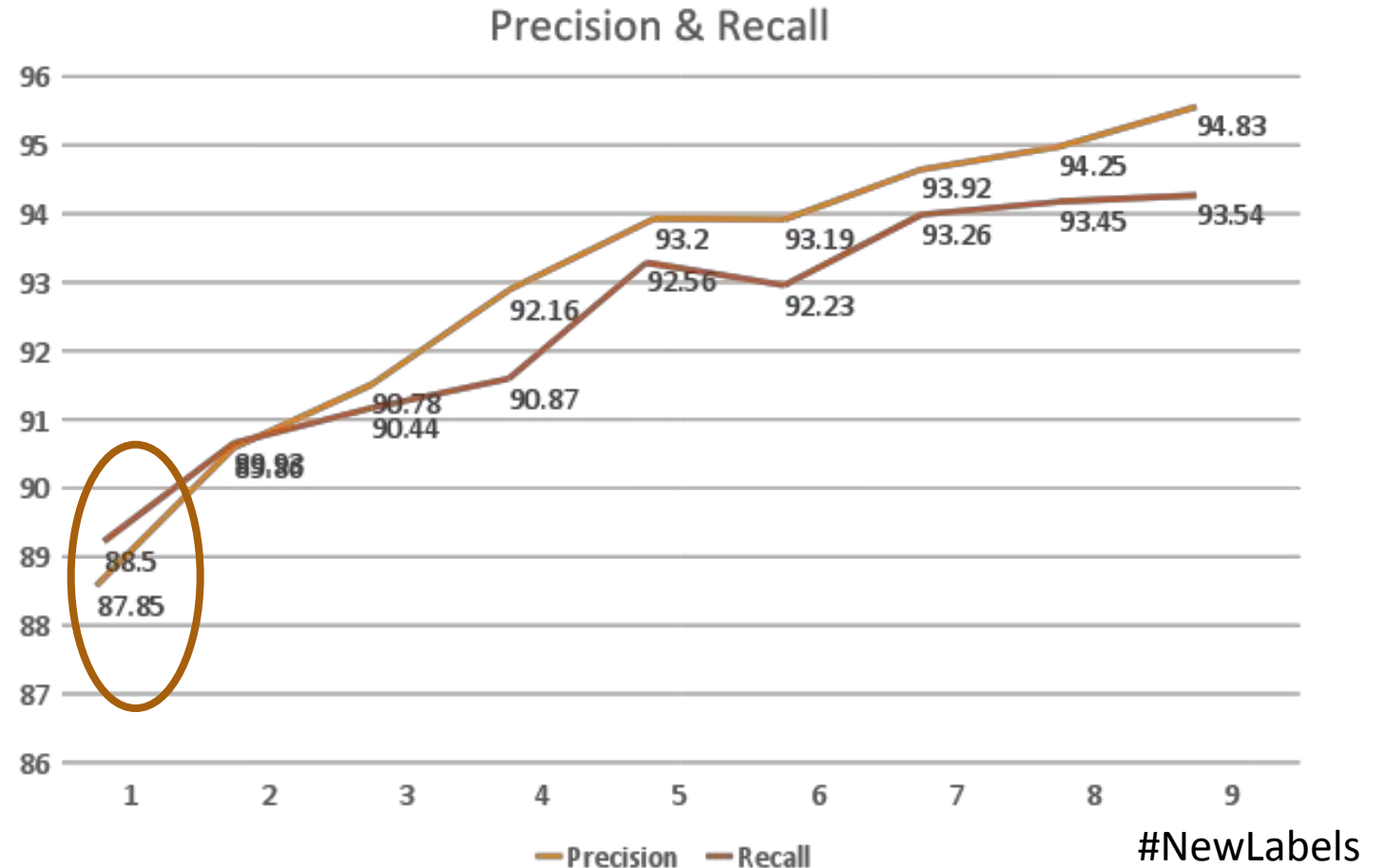
Training

500 Sentences
7927 Words
944 Flavors

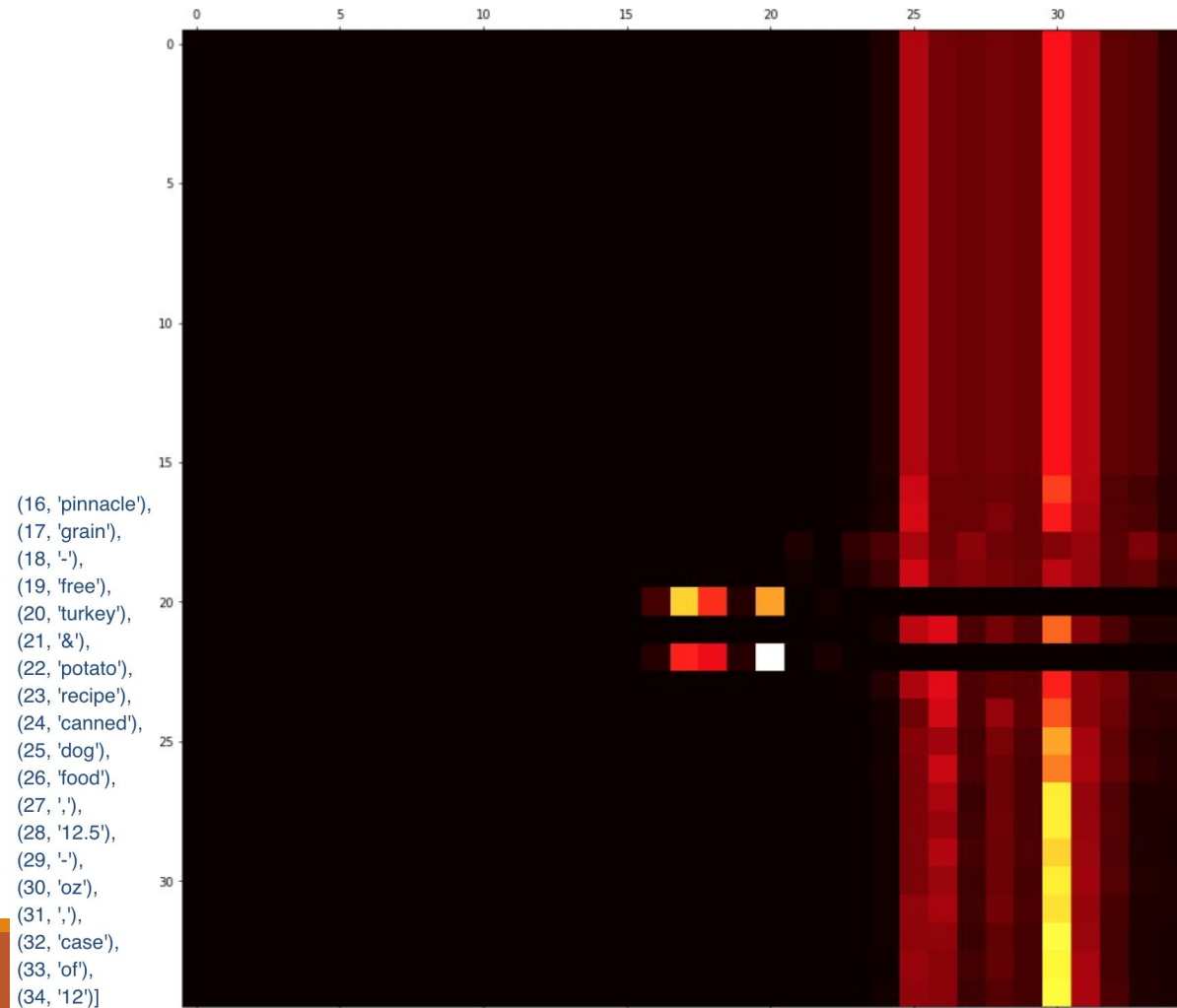
Testing

600 Sentences
7896 Words
786 Flavors

Different flavors
from Training data



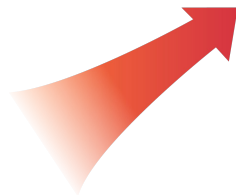
II. Open Attribute Extraction by NER —Attentions Help Find Contexts



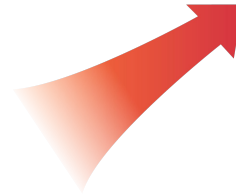
II. Extracting Knowledge from Product Profiles in Amazon Catalog



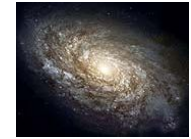
Product profile extraction



Automatically building a shallow KG



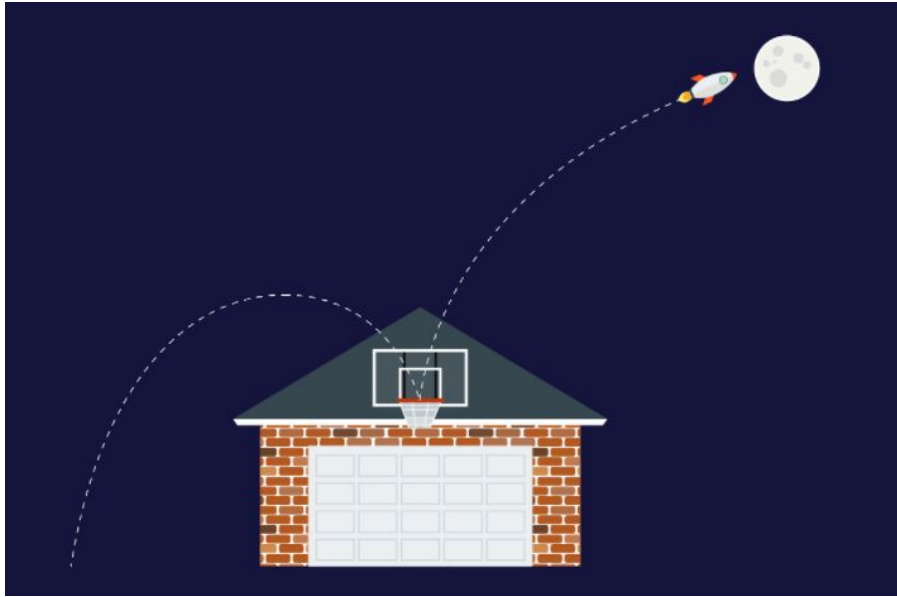
Open aspect extraction



Review extraction & sentiment analysis

Take Aways

- ❑ We aim at building an authoritative knowledge graph for all products in the world
- ❑ We shoot for roofshot and moonshot goals to realize our mission
- ❑ There are many exciting research problems that we are tackling



Thank You!