

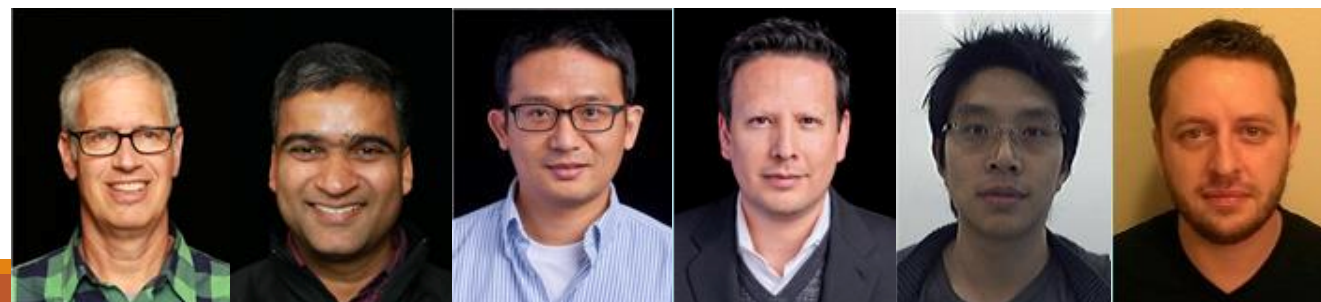
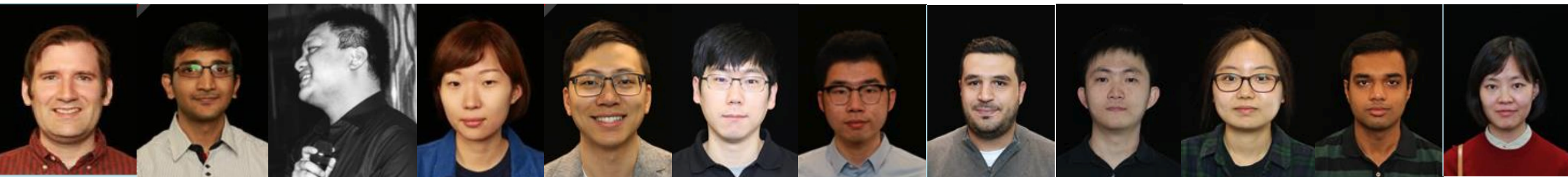
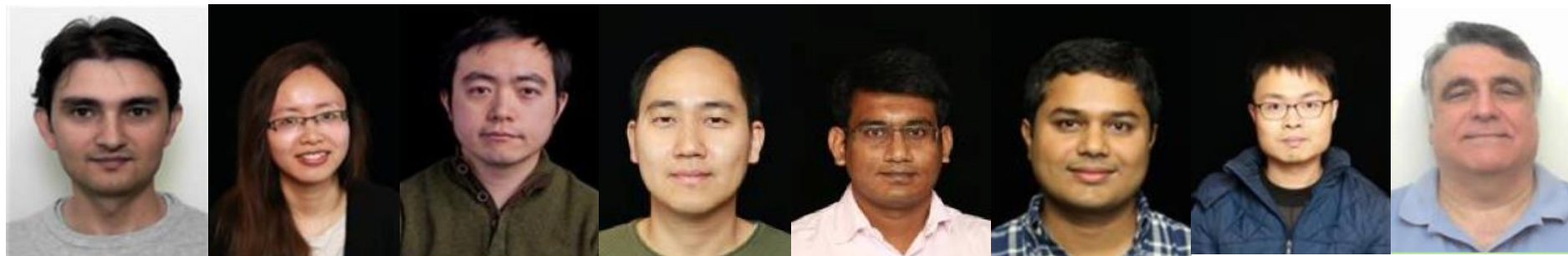
Building a Broad Knowledge Graph for Products

XIN LUNA DONG, AMAZON

ICDE, 4/2019

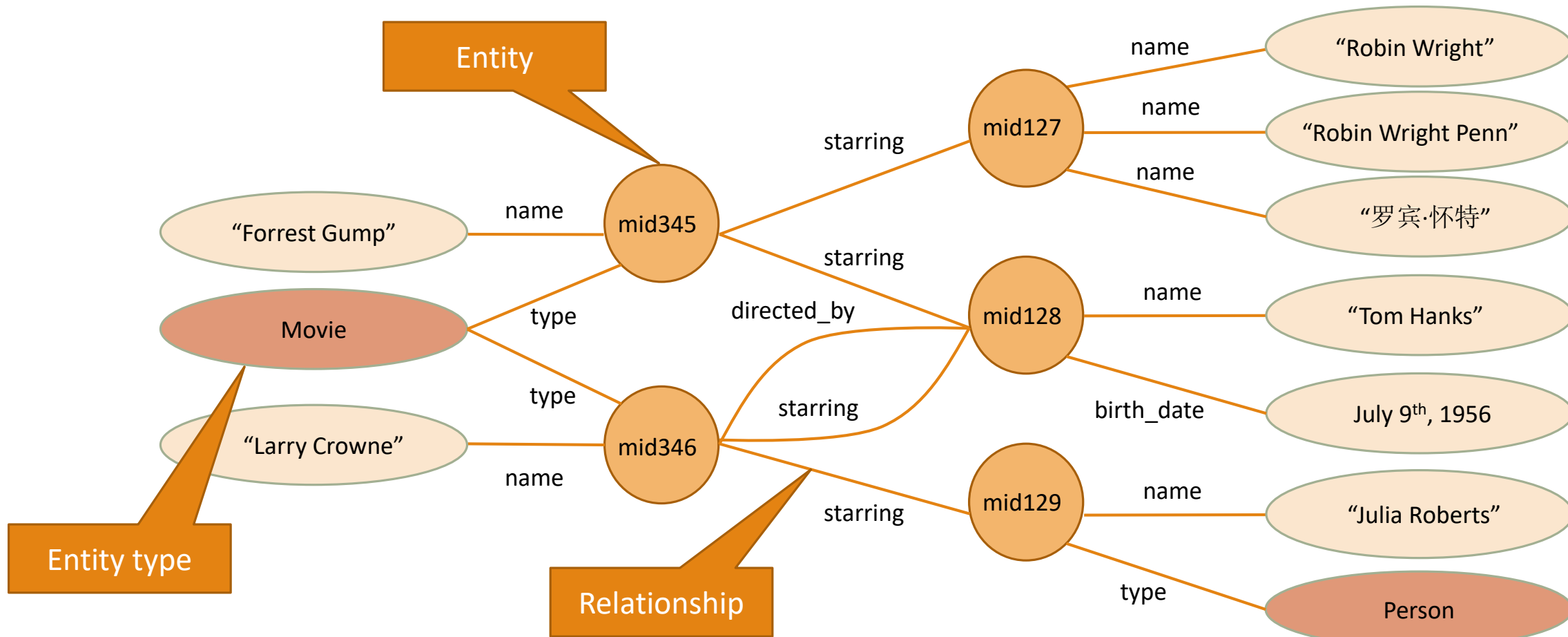


Acknowledgement




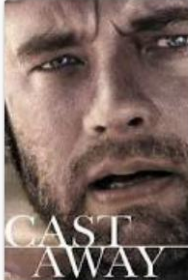







Product Graph vs. Knowledge Graph

Knowledge Graph Example for 2 Movies



Knowledge Graph in Search

Tom Hanks > Movies

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

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 Taylor Swift in the past year



Taylor Swift > Songs

Love Story
Fearless · 2008



Look What You Made Me...
Reputation · 2017



Shake It Off
1989 · 2014

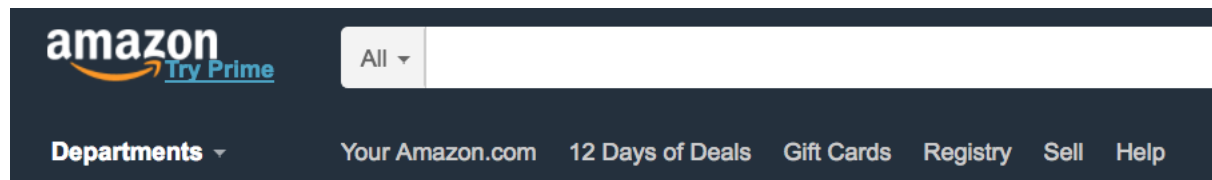


Delicate
Reputation · 2017



Product Graph

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



Customers who bought this item also bought



Cars 3 Playland with 20 Balls Playset
 ★★★★★☆ 3
 \$28.55 ✓prime



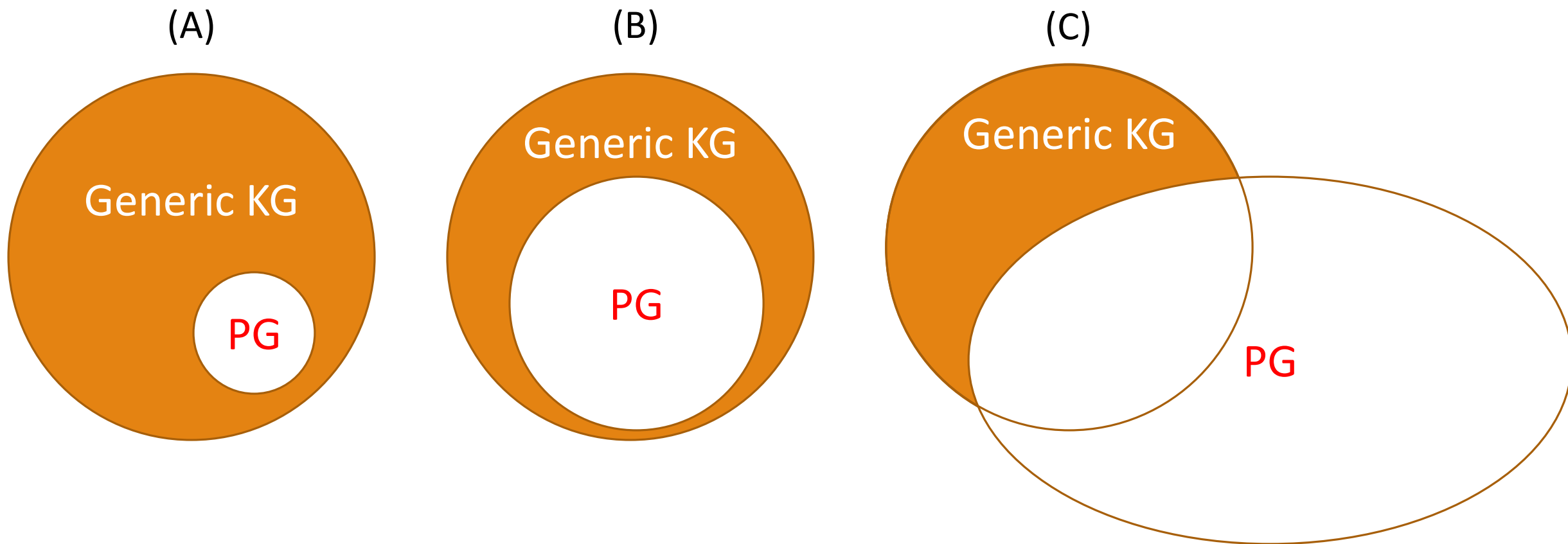
Step2 Push Around Sport Buggy
 ★★★★★☆ 18
 \$49.99 ✓prime



GOOD NIGHT, LIGHTNING
 > RH Disney
 ★★★★★☆ 299
 Board book
 \$7.70 ✓prime

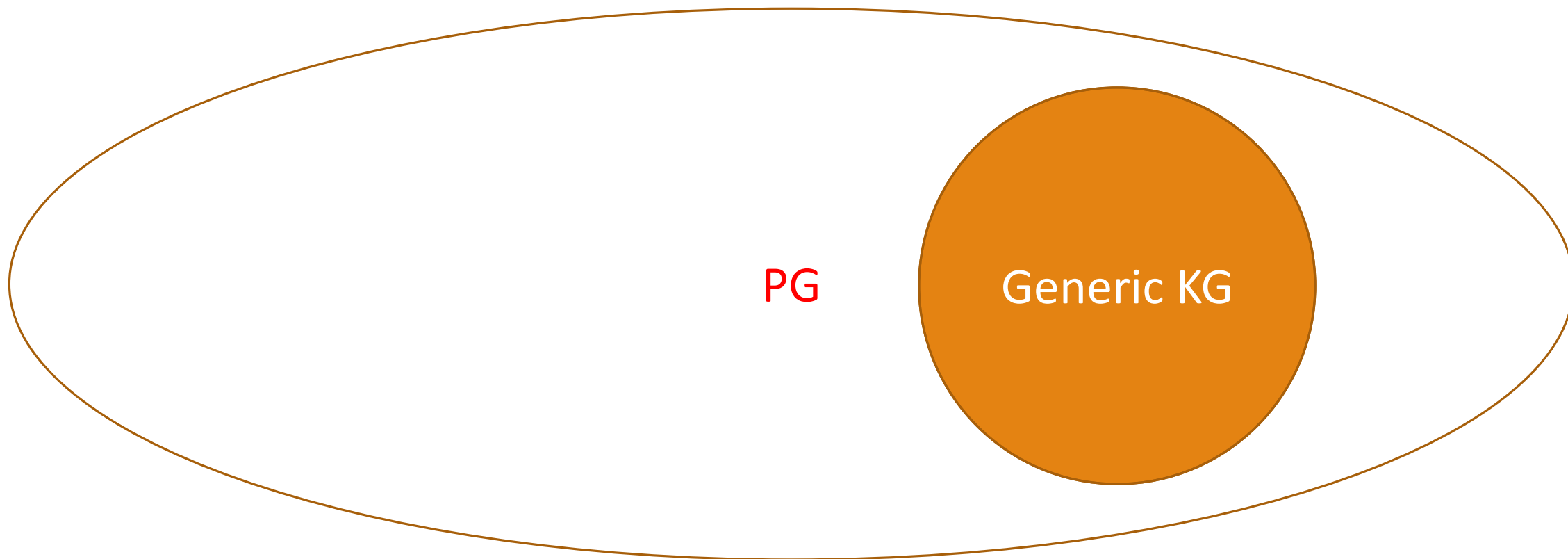


Product Graph vs. Knowledge Graph

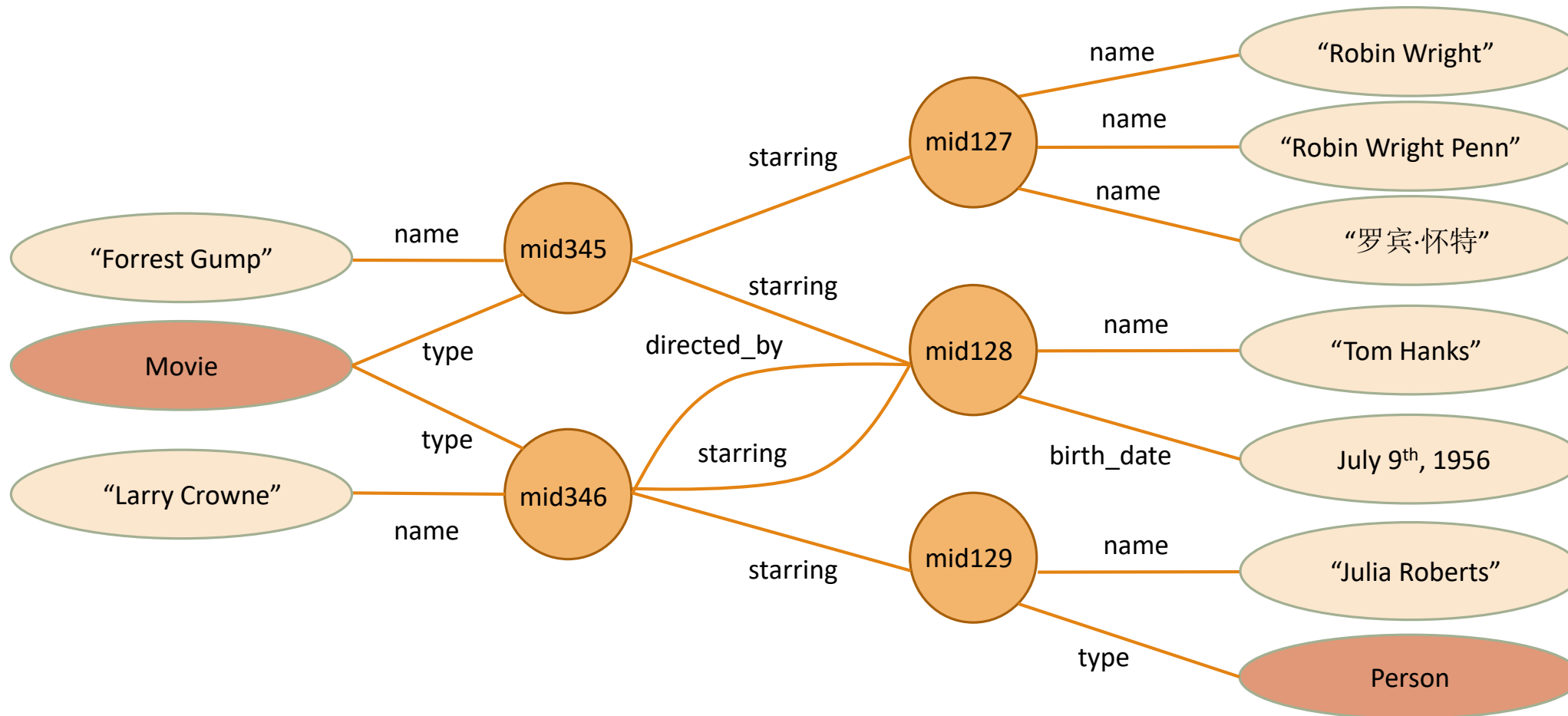


Product Graph vs. Knowledge Graph

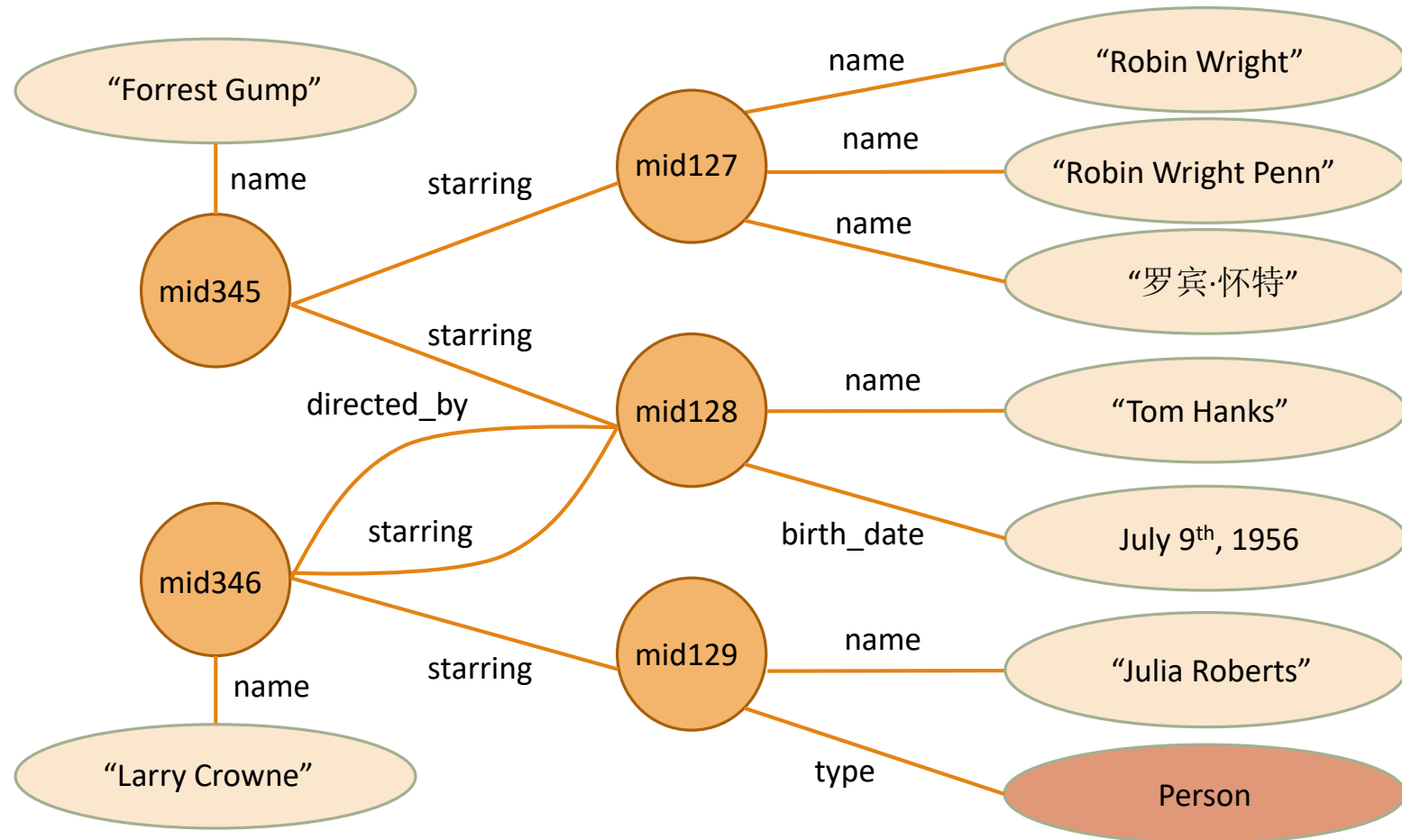
(D)



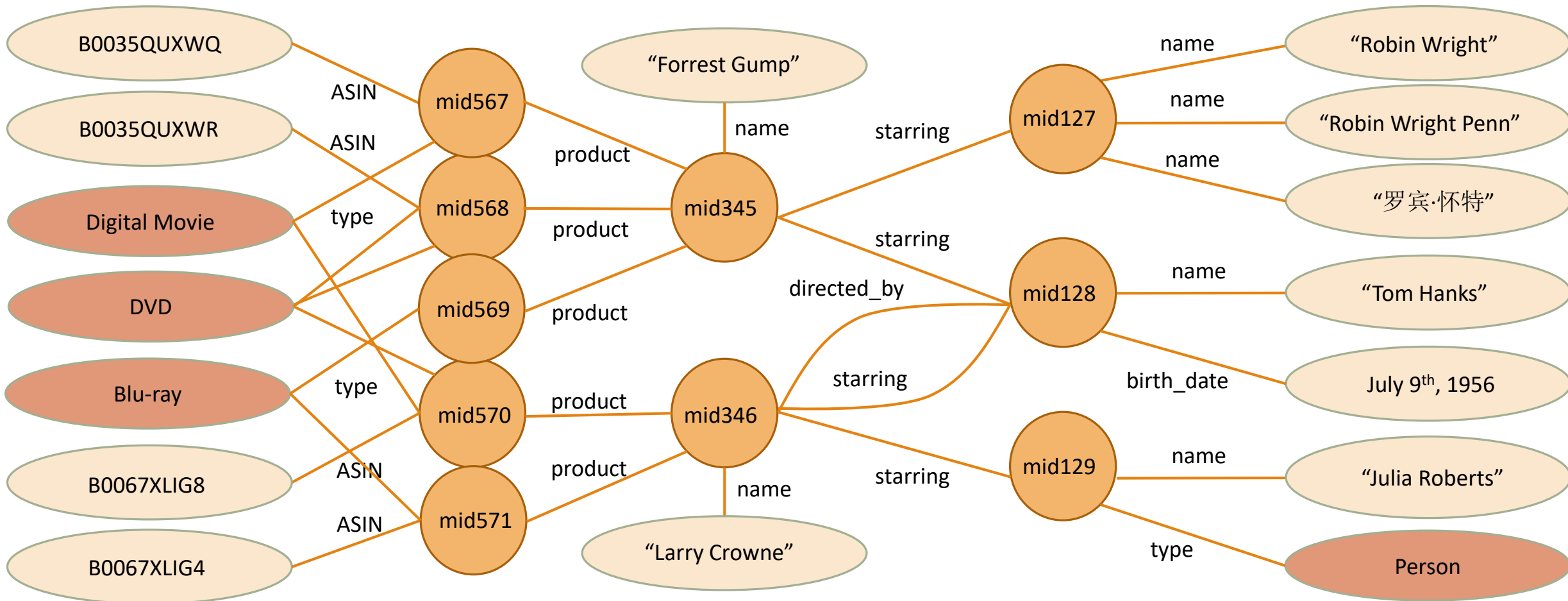
Knowledge Graph Example for 2 Movies



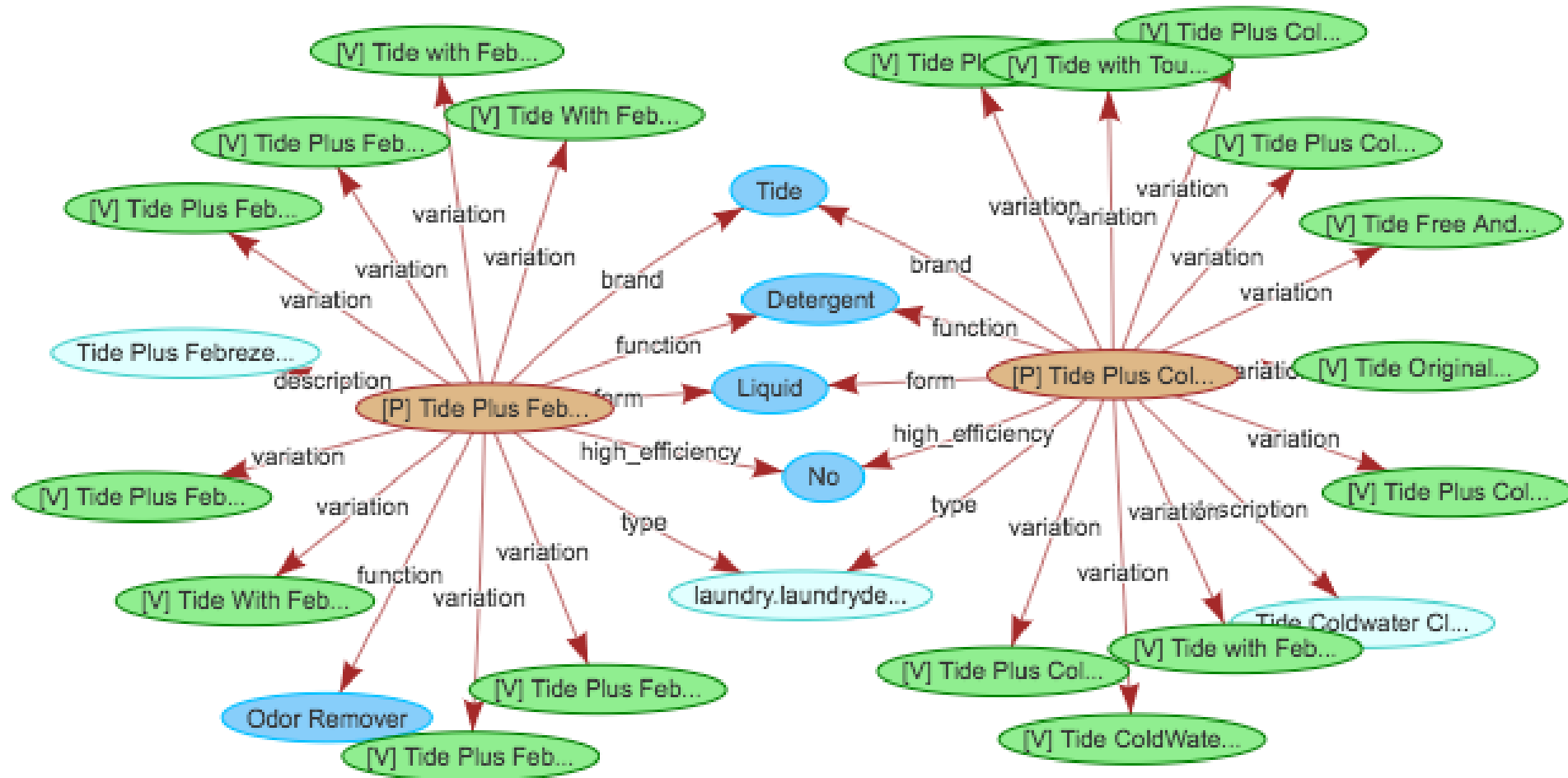
Product Graph vs. Knowledge Graph



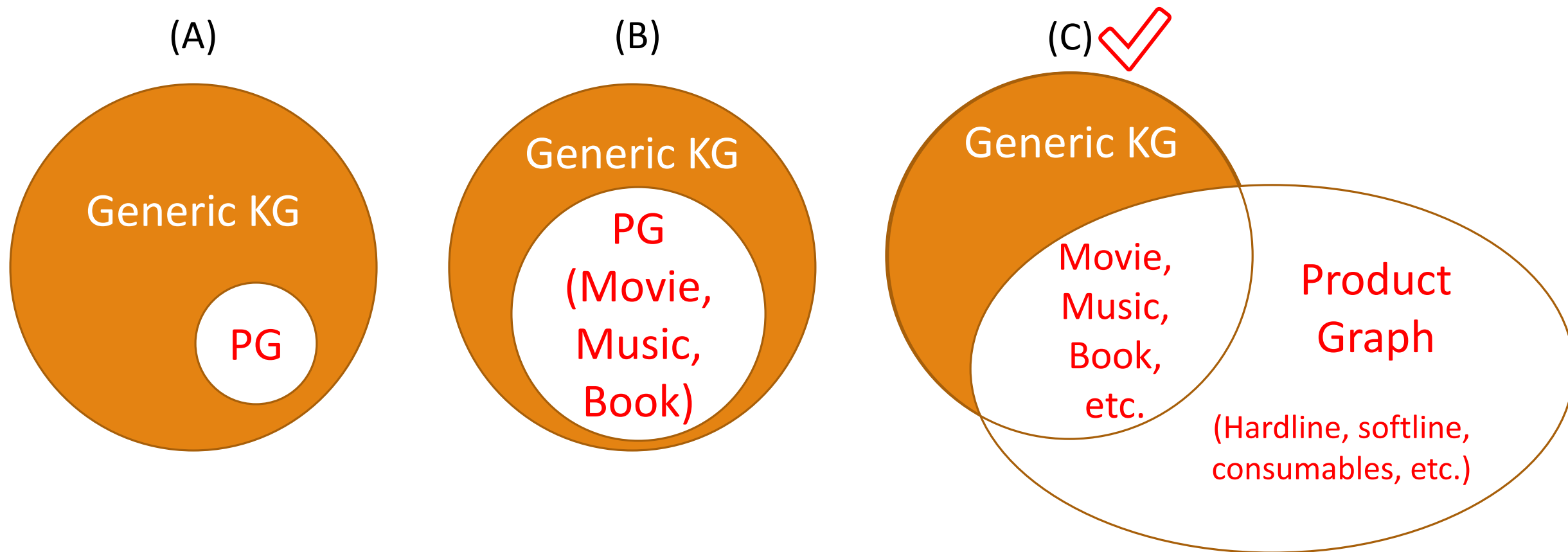
Product Graph vs. Knowledge Graph

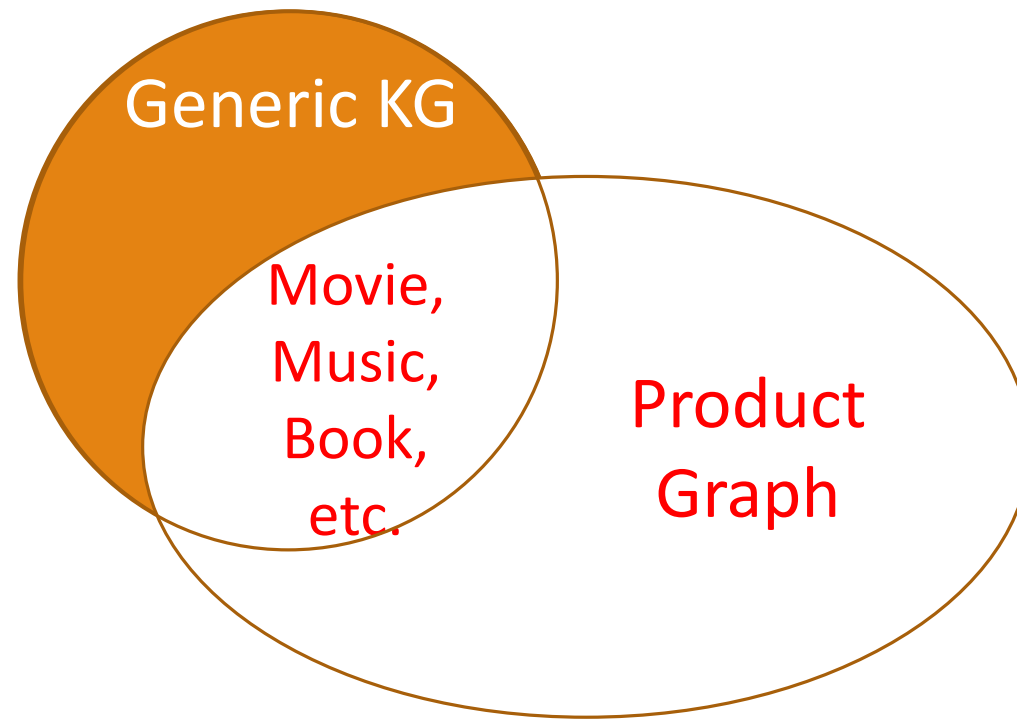


Another Example of Product Graph



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

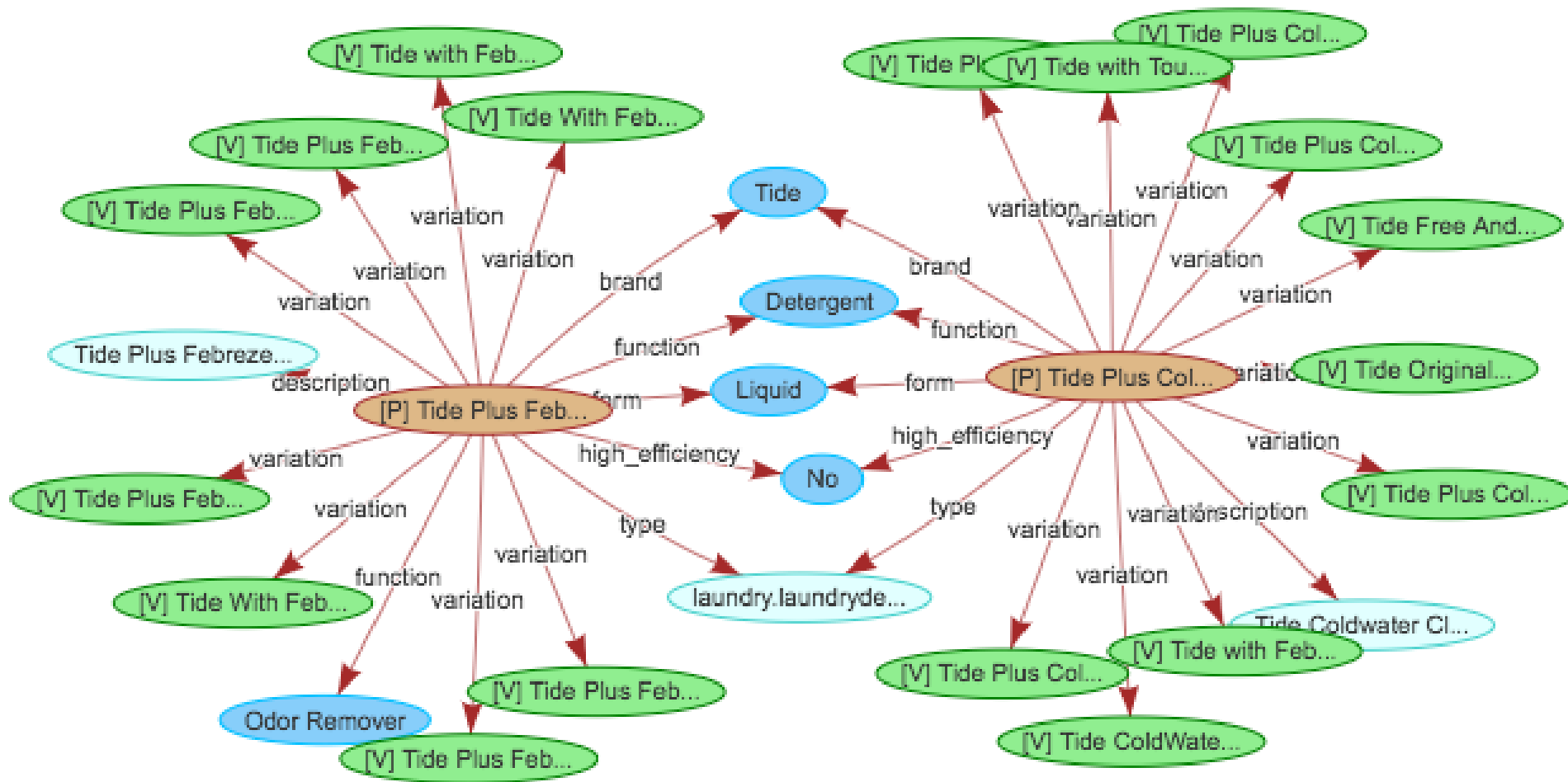
Challenges in Building Product Graph III

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

Challenges in Building Product Graph IV

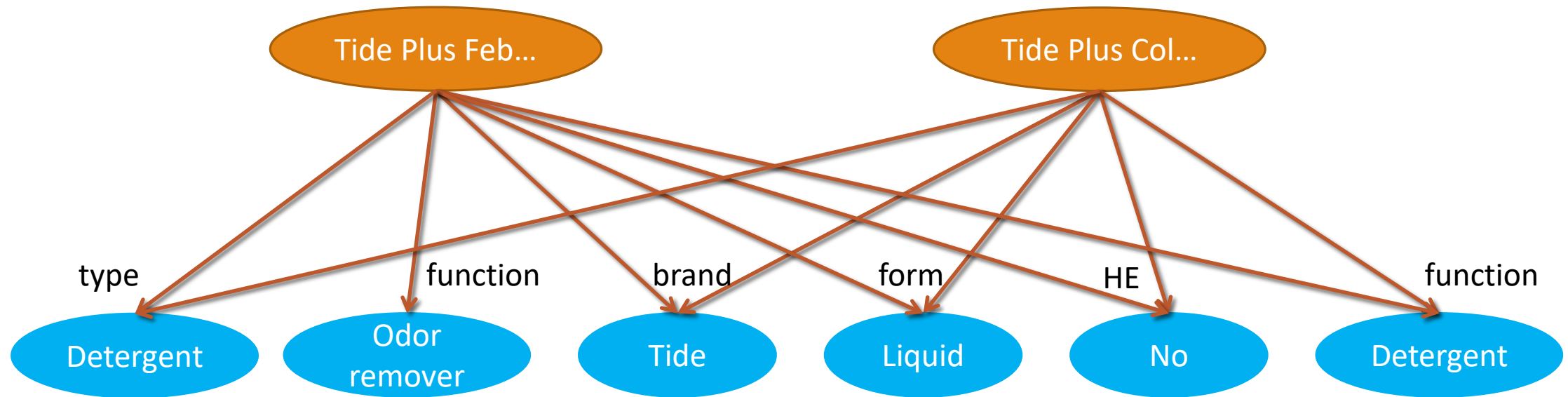
- ❑ Many entities are not named entities
 - ❑ Named Entity Recognition does not apply
 - ❑ New challenges for extraction, linking, and search

A 100-Year Project



Our Solution: Building a Broad Graph

A Broad & Shallow Version of the Same Graph



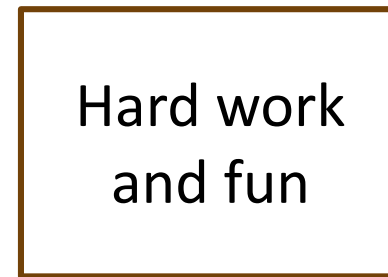
Design Principles for Broad Graph

- ❑ Start simple
 - ❑ Bi-partite graph
 - ❑ Core types and relationships
- ❑ Grow and clean the graph in a pay-as-you-go fashion
 - ❑ Ontology: user log analysis, web extraction
 - ❑ Data: product profile extraction, web extraction
 - ❑ Cleaning

Input and Output of Broad Graph

INPUT

1. A given vertical
2. Amazon Eco System
3. Semi-structured Web

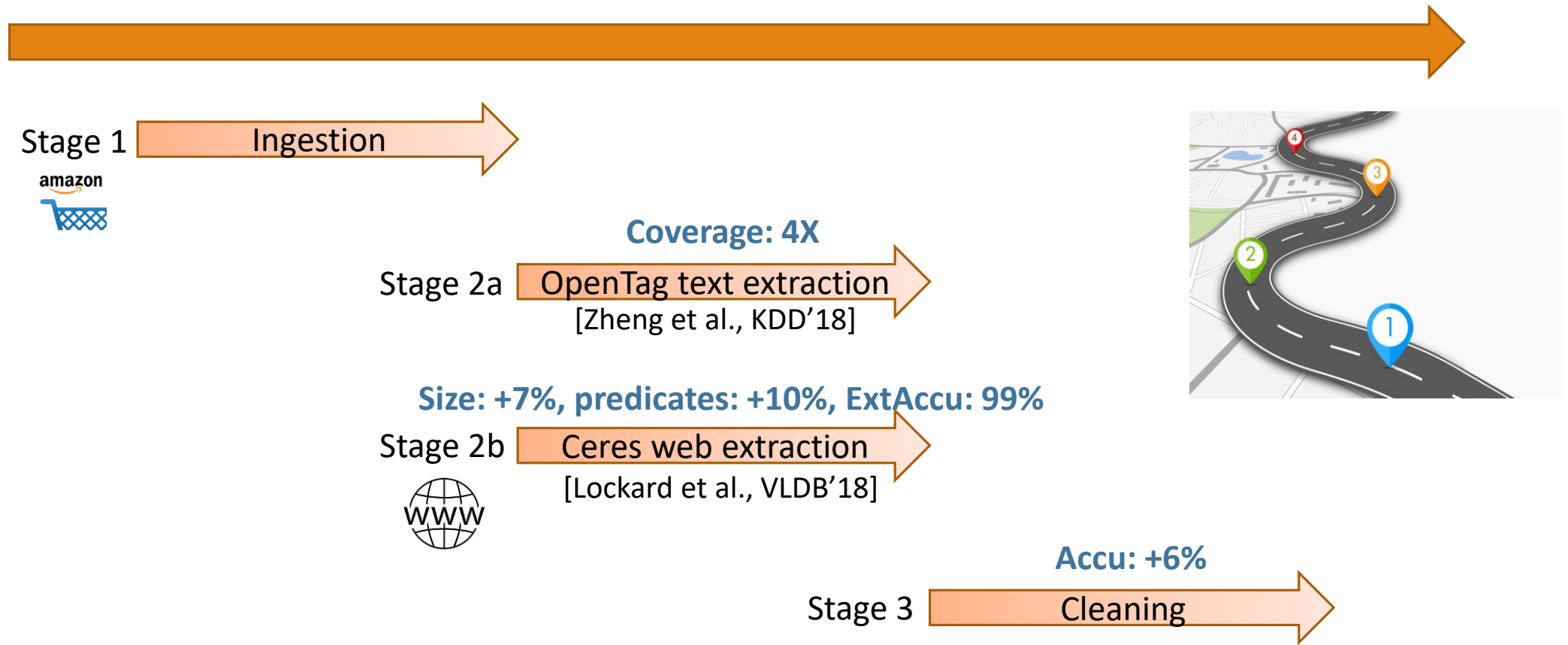


OUTPUT

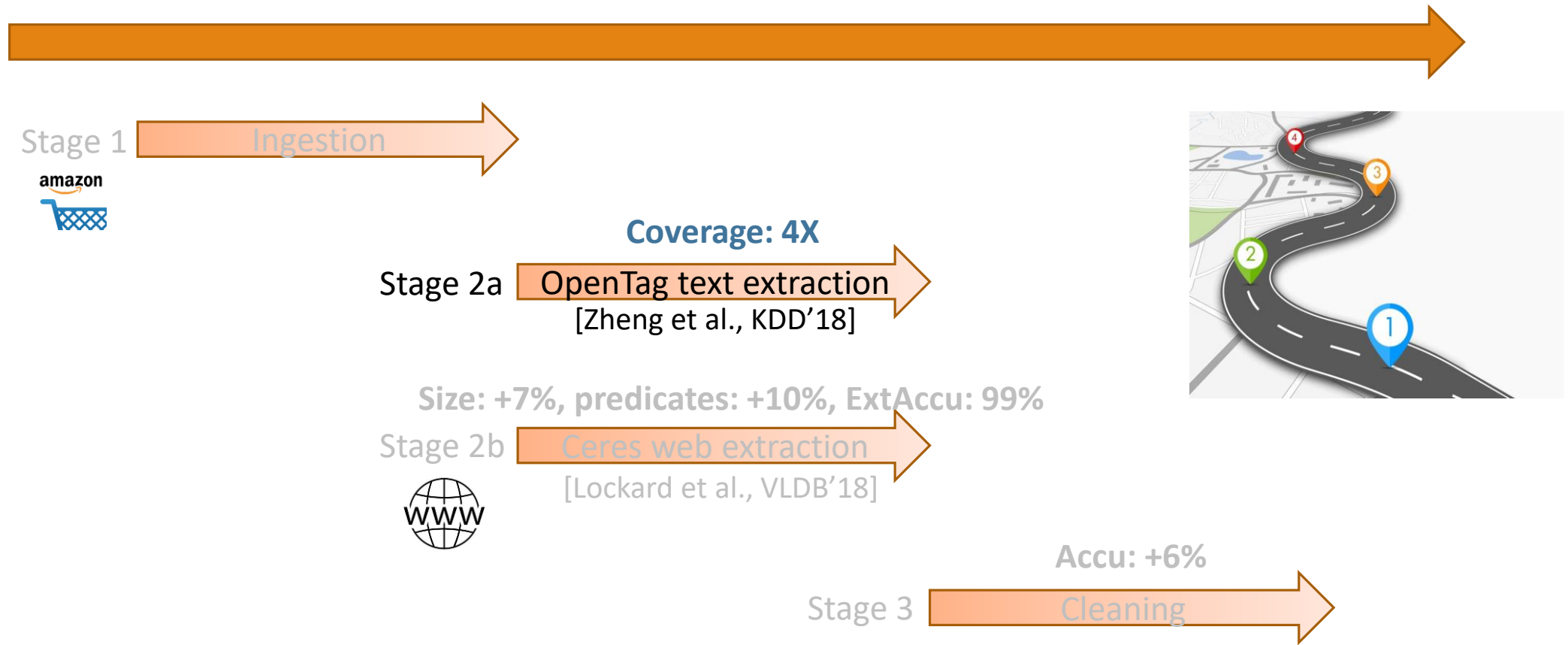
1. Ontology
2. Broad Graph
3. Comprehensive quality metrics

subject to: *hard_work* \leq *fun*

Stages in Building a Broad Graph



Stages in Building a Broad Graph



Stage 2a. OpenTag Extraction from Product Profiles [KDD'18]

| 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 | | |

Stage 2a. OpenTag Extraction from Product Profiles [KDD'18]

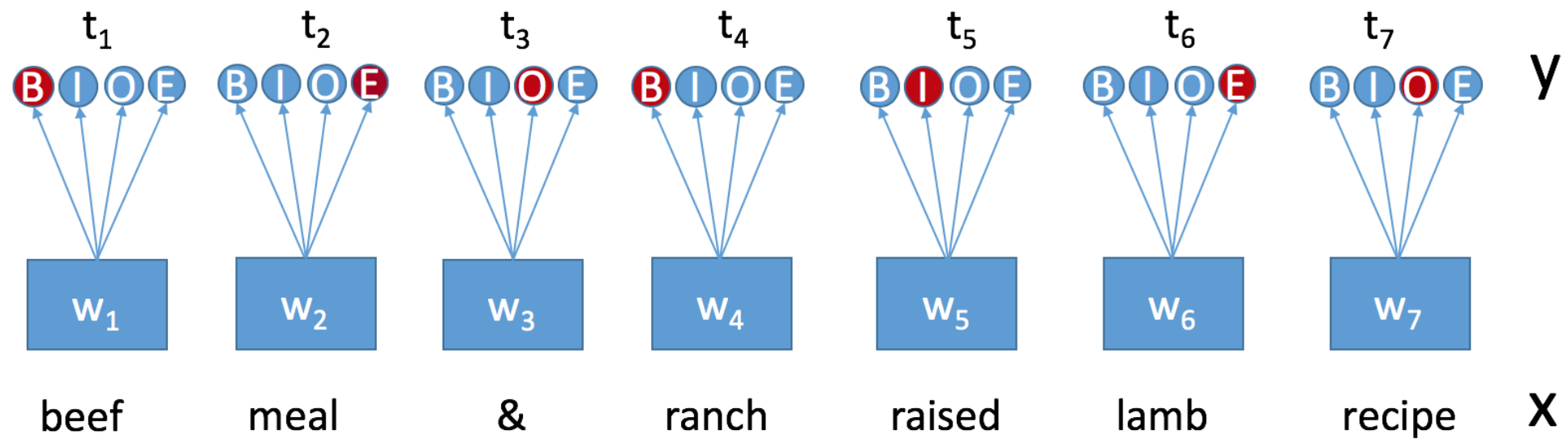
- B** Beginning of attribute value
- I** Inside of attribute value
- O** Outside of attribute value
- E** End of attribute value

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

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

OUTPUT

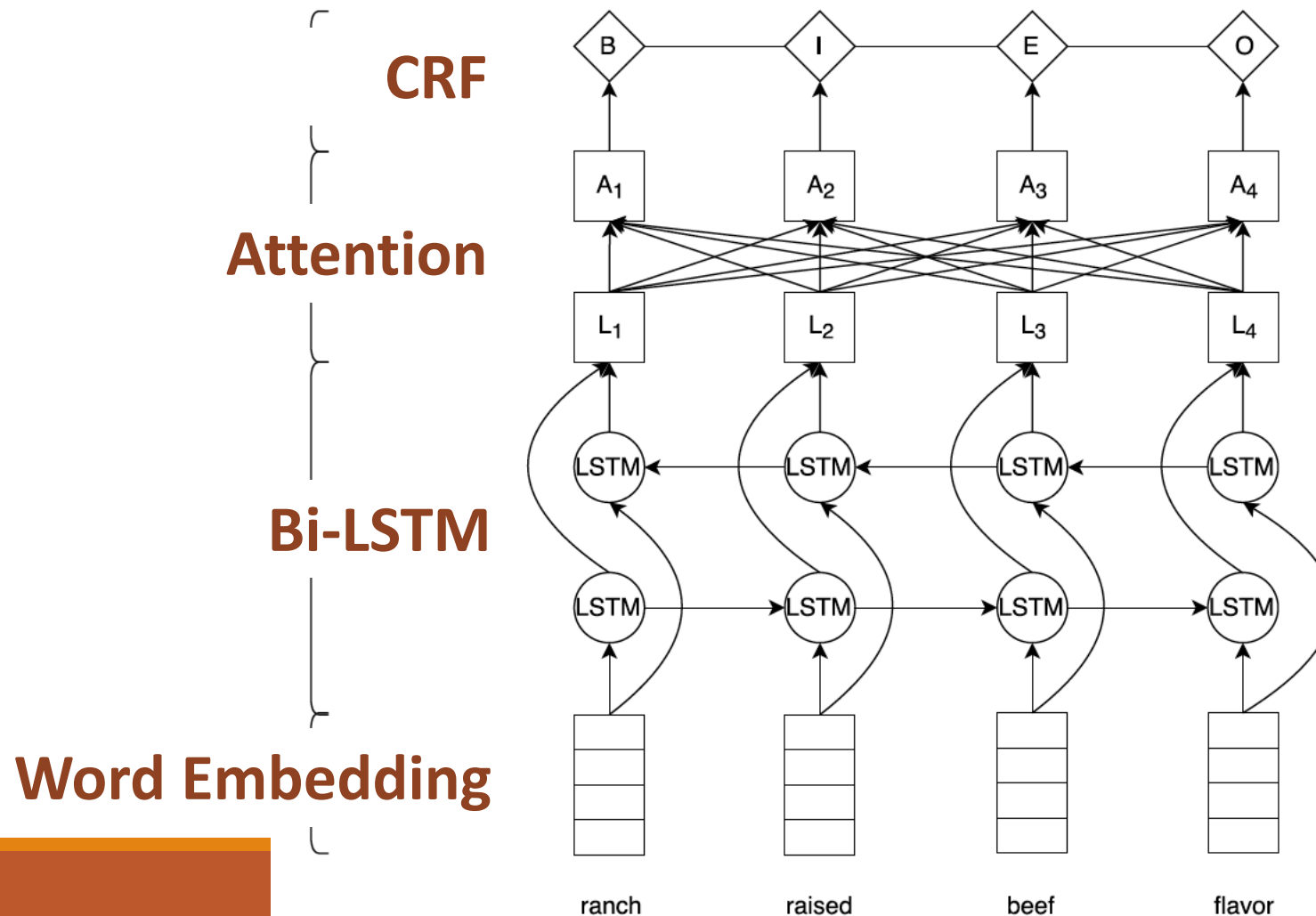
{beef meal} ← Flavor Extractions → {ranch raise lamb}



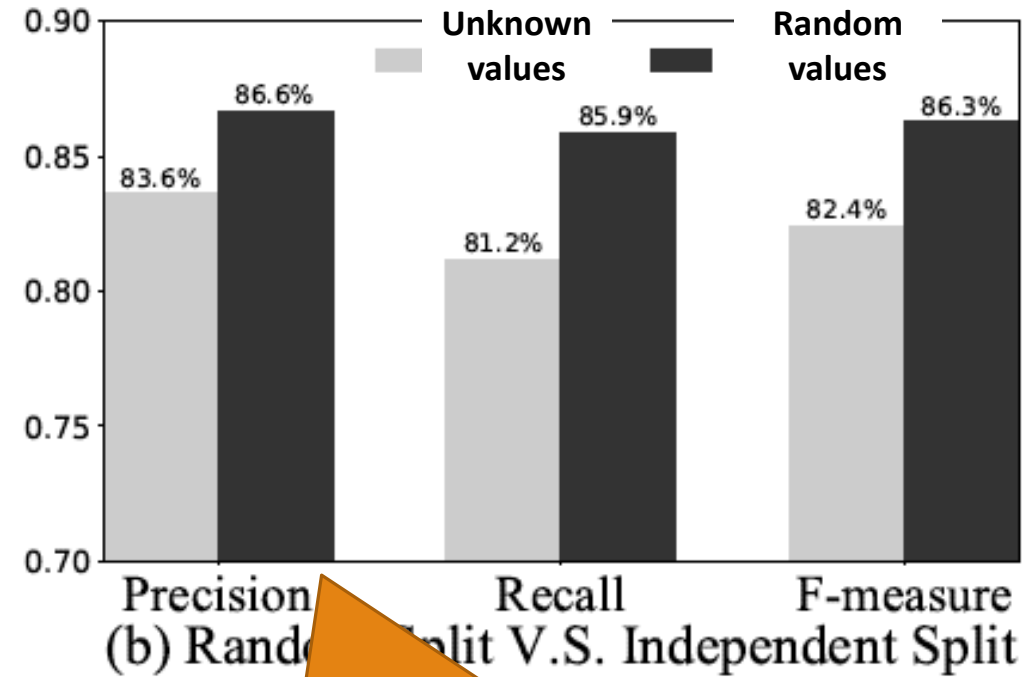
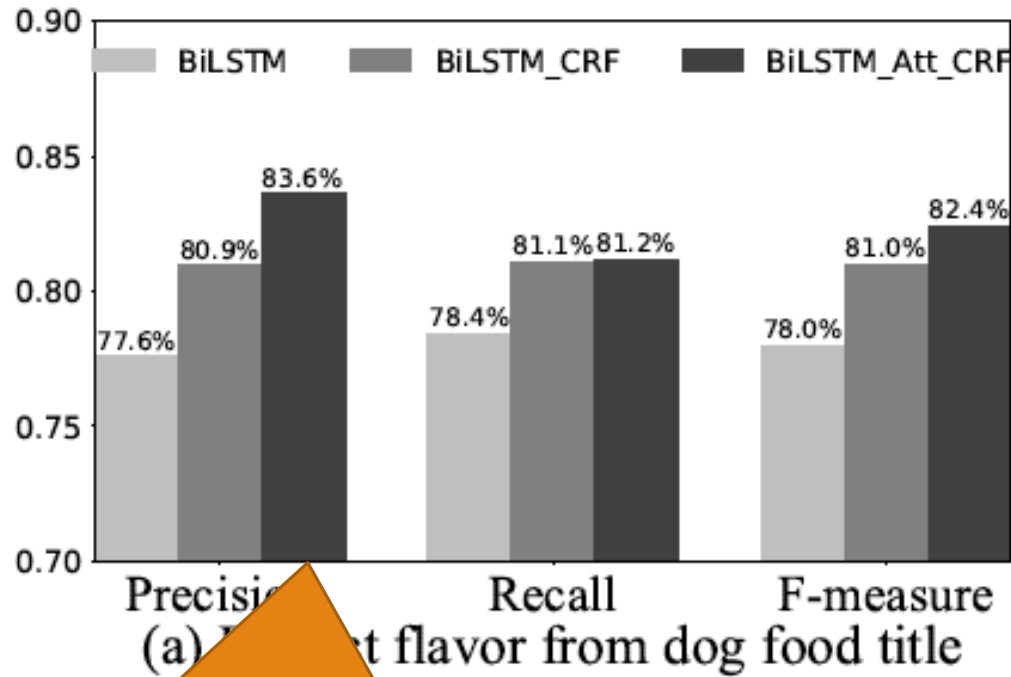
INPUT

X

Stage 2a. OpenTag Extraction from Product Profiles [KDD'18]



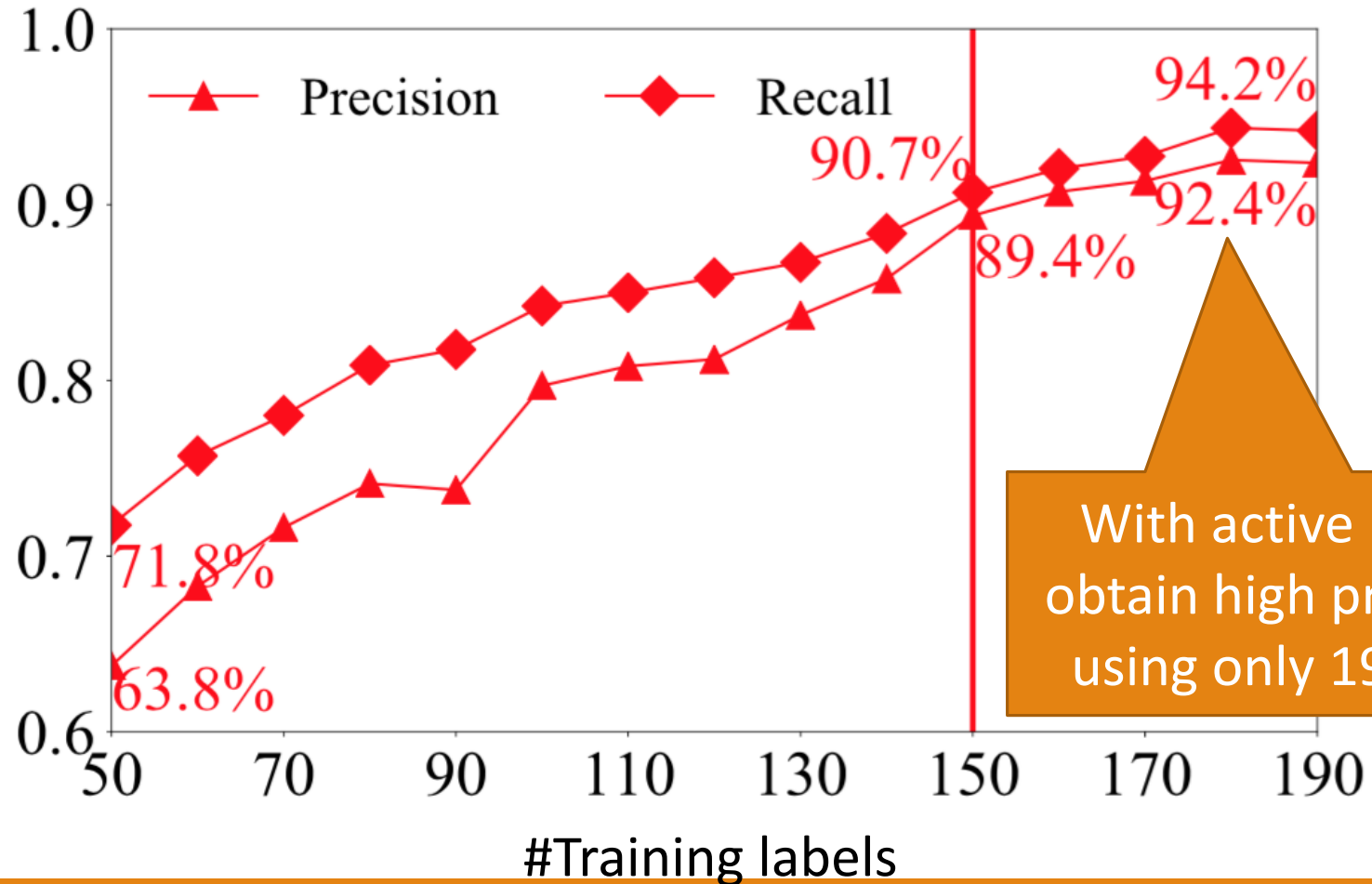
Stage 2a. OpenTag Extraction from Product Profiles [KDD'18]



BiLSTM+CRF+Attention obtains best results

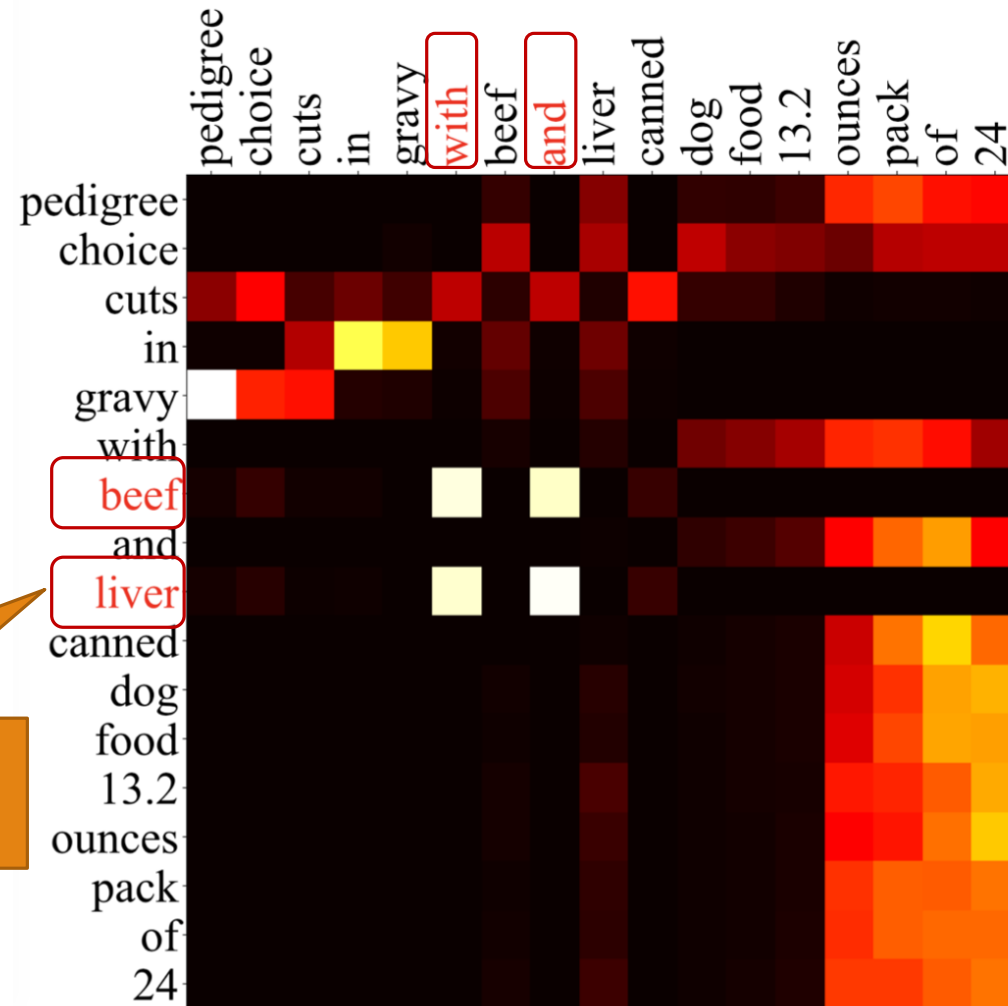
Extraction on new values is comparable to already known values

Stage 2a. OpenTag Extraction from Product Profiles [KDD'18]



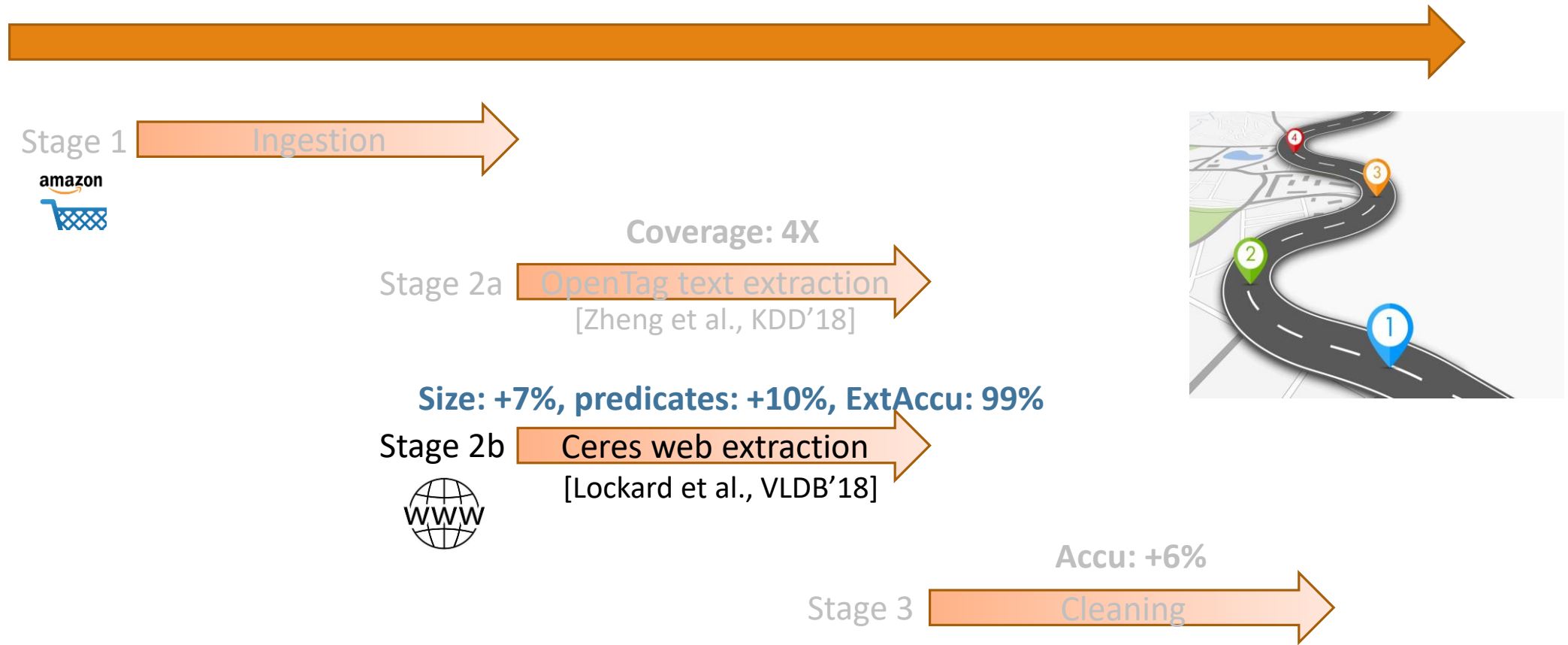
With active learning, we can obtain high precision and recall using only 190 training labels

Stage 2a. OpenTag Extraction from Product Profiles [KDD'18]



Interpretability via attention

Stages in Building a Broad Graph



Stage 2b. 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

0:50 | Trailer

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

As students at the United States Navy's elite fighter training school, one daring young pilot learns a few things from the 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 credits

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

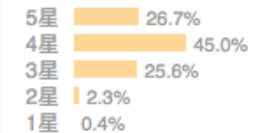
卧虎藏龙 臥虎藏龍 (2000)



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

豆瓣评分

7.9 166740人评价



好于 92% 武侠片
好于 90% 动作片

[想看](#) [看过](#) 评价: ☆☆☆☆☆

[写短评](#) [写影评](#) [+ 提问题](#) [分享到](#)

推荐

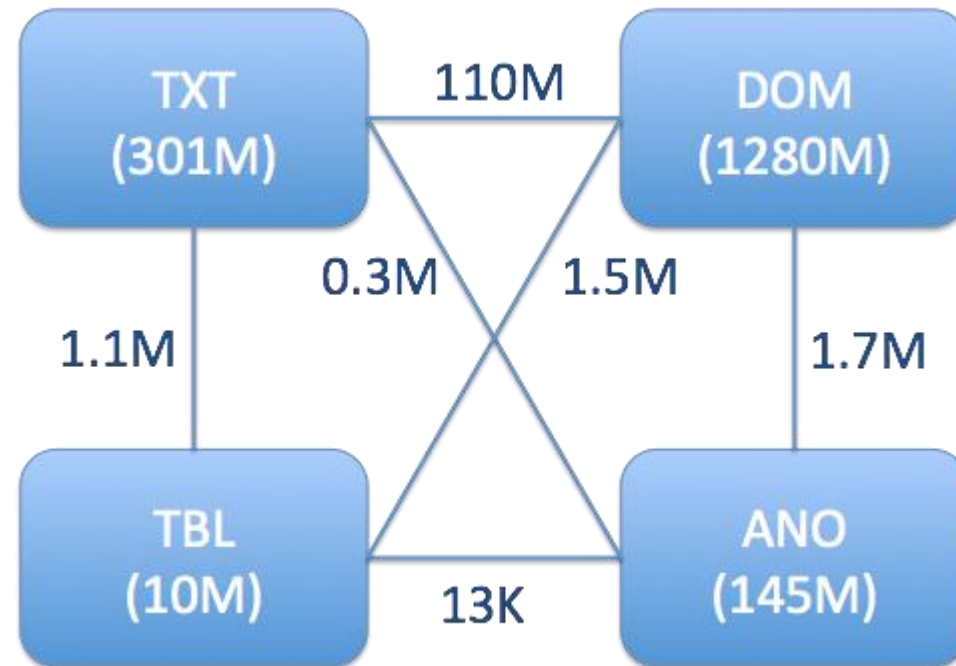
卧虎藏龙的剧情简介 ·····

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,她隐匿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

Stage 2b. Extracting Knowledge from Semi-Structured Data on the Web

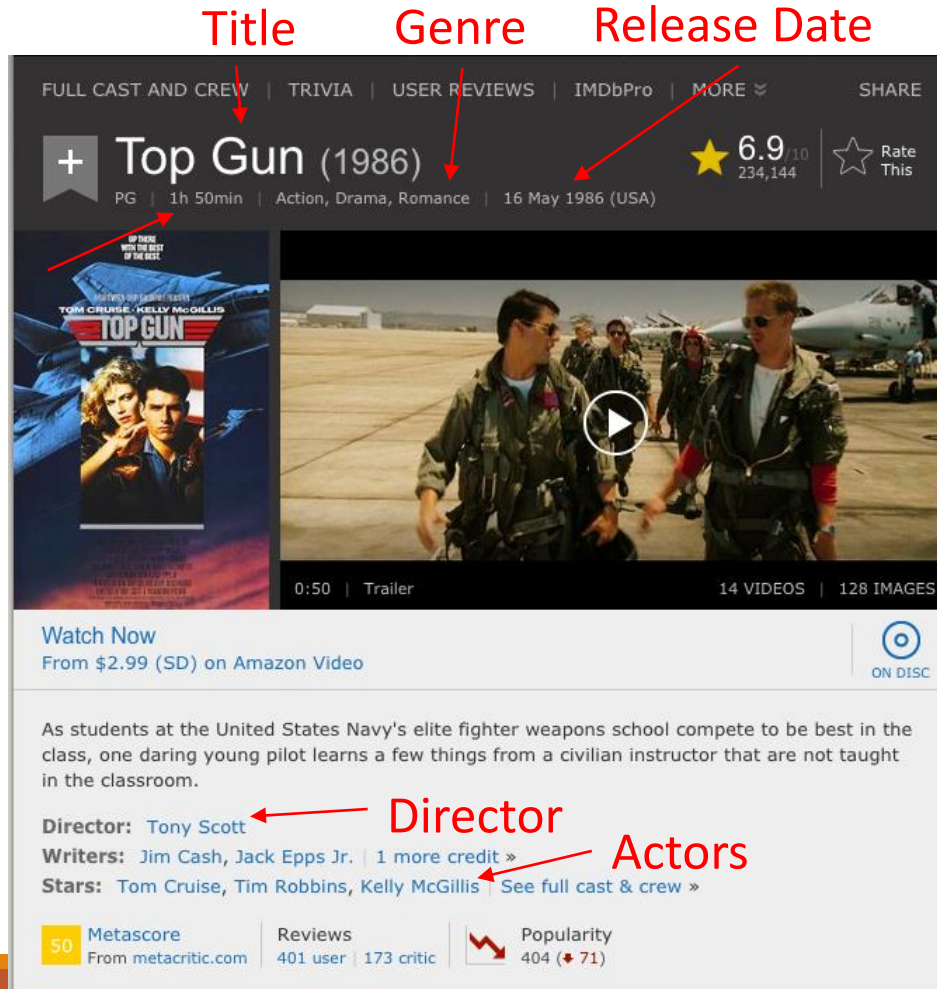
Knowledge Vault @ Google showed big potential from DOM-tree extraction [Dong et al., KDD'14][Dong et al., VLDB'14]

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



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

Stage 2b. Extracting Knowledge from Semi-Structured Data on the Web



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

Additional visible data: Rating 6.9/10 (234,144 votes), PG rating, 14 VIDEOS, 128 IMAGES, Metascore 50, 401 user reviews, 173 critic reviews, Popularity 404 (71).

Extracted relationships


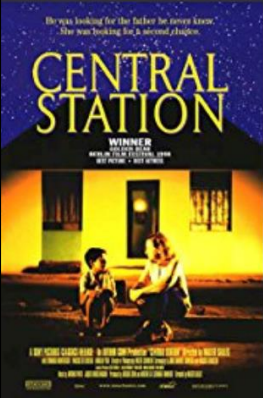
- (Top Gun, type.object.name, "Top Gun")
- (Top Gun, film.film.genre, Action)
- (Top Gun, film.film.directed_by, Tony Scott)
- (Top Gun, film.film.starring, Tom Cruise)
- (Top Gun, film.film.runtime, "1h 50min")
- (Top Gun, film.film.release_Date_s, "16 May 1986")

Stage 2b. Extracting Knowledge from Semi-Structured Data on the Web


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

Central Station (1998) ★ 8.0/10 31,520 ☆ Rate This

Central do Brasil (original title)
R | 1h 53min | Drama | 20 November 1998 (USA)



1:54 | Trailer | 1 VIDEO | 22 IMAGES

 On Disc at Amazon

An emotive journey of a former school teacher, who writes letters for illiterate people, and a young boy, whose mother has just died, as they search for the father he never knew.


Director: Walter Salles
Writers: Marcos Bernstein, João Emanuel Carneiro | 1 more credit »
Stars: Fernanda Montenegro, Vinícius de Oliveira, Marília Pêra | See full cast & crew »

80 Metascore From metacritic.com

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

Star Wars: The Last Jedi (2017) ★ 7.3/10 404,499 ☆ Rate This


Star Wars: Episode VIII - The Last Jedi (original title)
PG-13 | 2h 32min | Action, Adventure, Fantasy | 15 December 2017 (USA)




Rey develops her newly discovered abilities with the guidance of Luke Skywalker, who is unsettled by the strength of her powers. Meanwhile, the Resistance prepares for battle with the First Order.

Director: Rian Johnson
Writers: Rian Johnson, George Lucas (based on characters created by)
Stars: Daisy Ridley, John Boyega, Mark Hamill | See full cast & crew »

85 Metascore From metacritic.com | 5,463 user | 645 critic | Popularity 84 (↑ 3)

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

 ON DISC


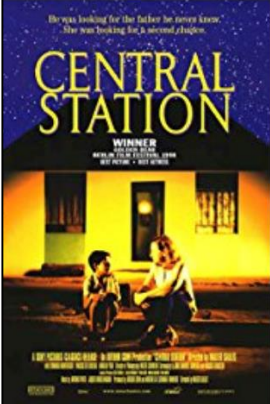
Same pred may corr. to diff DOM tree nodes

Stage 2b. Extracting Knowledge from Semi-Structured Data on the Web


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

Central Station (1998) ★ 8.0 /10 ☆ Rate This
31,520

Central do Brasil (*original title*)
R | 1h 53min | Drama | 20 November 1998 (USA)



1:54 | Trailer | 1 VIDEO | 22 IMAGES

 On Disc at Amazon



An emotive journey of a former school teacher, who writes letters for illiterate people, and a young boy, whose mother has just died, as they search for the father he never knew.

Director: [Walter Salles](#)
Writers: [Marcos Bernstein](#), [João Emanuel Carneiro](#) | 1 more credit »
Stars: [Fernanda Montenegro](#), [Vinicius de Oliveira](#), [Marília Pêra](#) | See full cast & crew »


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

The Fog of War: Eleven Lessons from the Life of Robert S. McNamara (2003) ★ 8.2 /10 ☆ Rate This
20,953

PG-13 | 1h 47min | Documentary, Biography, History | 5 March 2004 (USA)



2:09 | Trailer | 2 VIDEOS | 11 IMAGES

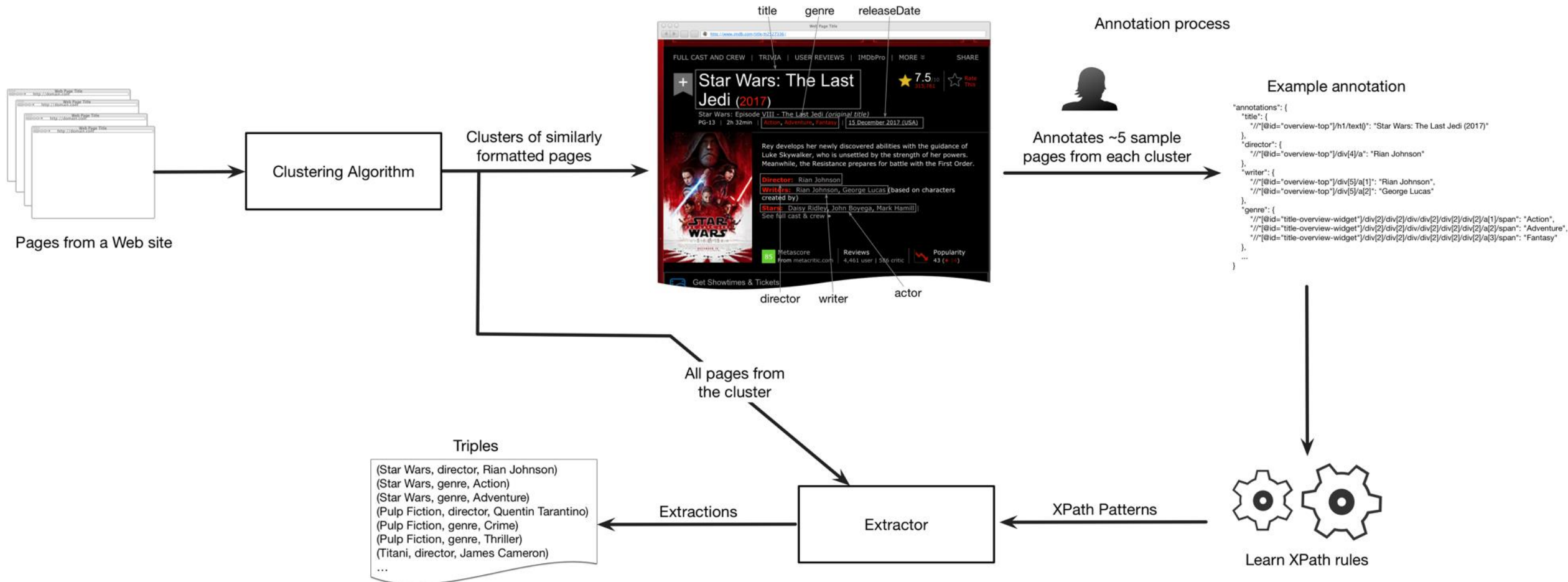
 Watch Now
From \$2.99 (SD) on Prime Video ON DISC

The story of America as seen through the eyes of the former Secretary of Defense under President [John F. Kennedy](#) and President [Lyndon Johnson](#), [Robert McNamara](#).

Director: [Errol Morris](#)
Stars: [Robert McNamara](#), [John F. Kennedy](#), [Fidel Castro](#) | See full cast & crew »

Same DOM tree node may correspond to diff preds

Stage 2b. Extracting Knowledge from Semi-Structured Data on the Web



Stage 2b. Linking Entities Between Sources

- ❑ Extraction experiments on <http://swde.codeplex.com/>
- ❑ Near perfect precision and recall
- ❑ Annotations on 5 pages per site

| Vertical | Predicate | Vertex++ | | |
|-----------|----------------|-------------|-------------|-------------|
| | | P | R | F1 |
| Movie | Title | 1.00 | 1.00 | 1.00 |
| | Director | 0.99 | 0.99 | 0.99 |
| | Genre | 0.88 | 0.87 | 0.87 |
| | MPAA Rating | 1.00 | 1.00 | 1.00 |
| | Average | 0.97 | 0.97 | 0.97 |
| NBAPlayer | Name | 0.99 | 0.99 | 0.99 |
| | Team | 1.00 | 1.00 | 1.00 |
| | Weight | 1.00 | 1.00 | 1.00 |
| | Height | 1.00 | 1.00 | 1.00 |
| | Average | 1.00 | 1.00 | 1.00 |

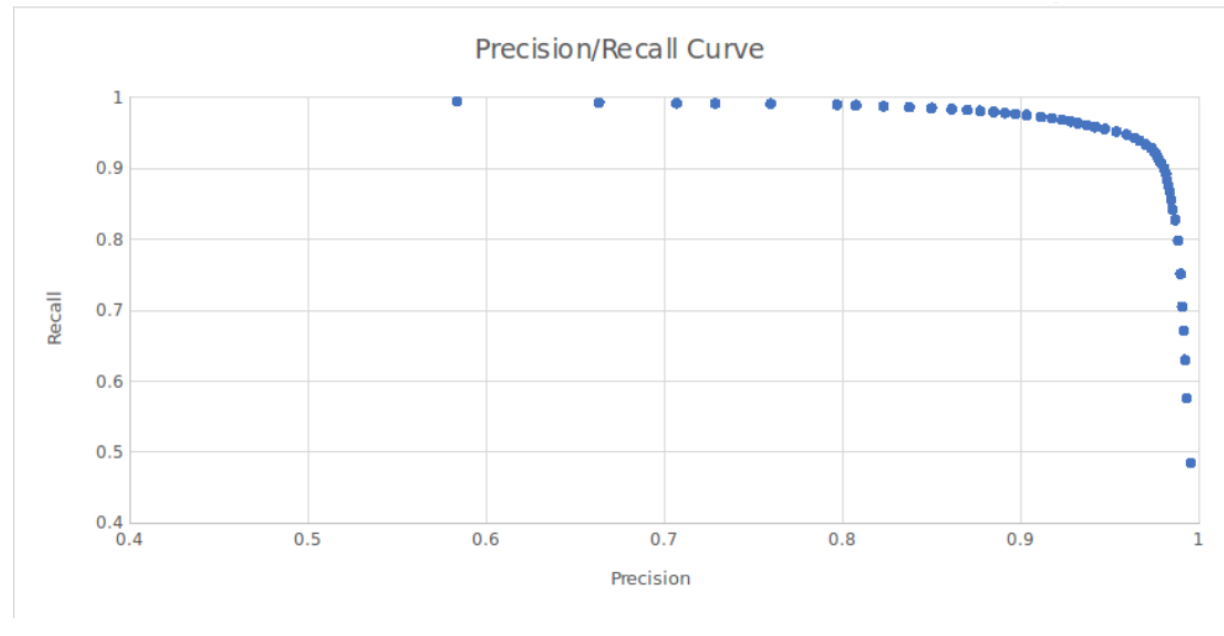
| Vertical | Predicate | Vertex++ | | |
|----------------|------------------|-------------|-------------|-------------|
| | | P | R | F1 |
| University | Name | 1.00 | 1.00 | 1.00 |
| | Type | 1.00 | 1.00 | 1.00 |
| | Phone | 0.97 | 0.92 | 0.94 |
| | Website | 1.00 | 1.00 | 1.00 |
| | Average | 0.99 | 0.98 | 0.99 |
| Book | Title | 0.99 | 0.99 | 0.99 |
| | Author | 0.97 | 0.96 | 0.96 |
| | Publisher | 0.85 | 0.85 | 0.85 |
| | Publication Date | 0.90 | 0.90 | 0.90 |
| | ISBN-13 | 0.94 | 0.94 | 0.94 |
| Average | 0.93 | 0.93 | 0.93 | |

Stage 2b. Linking Entities Between Sources

- Random forest on attribute-wise similarity
- Results between Freebase and IMDb movies

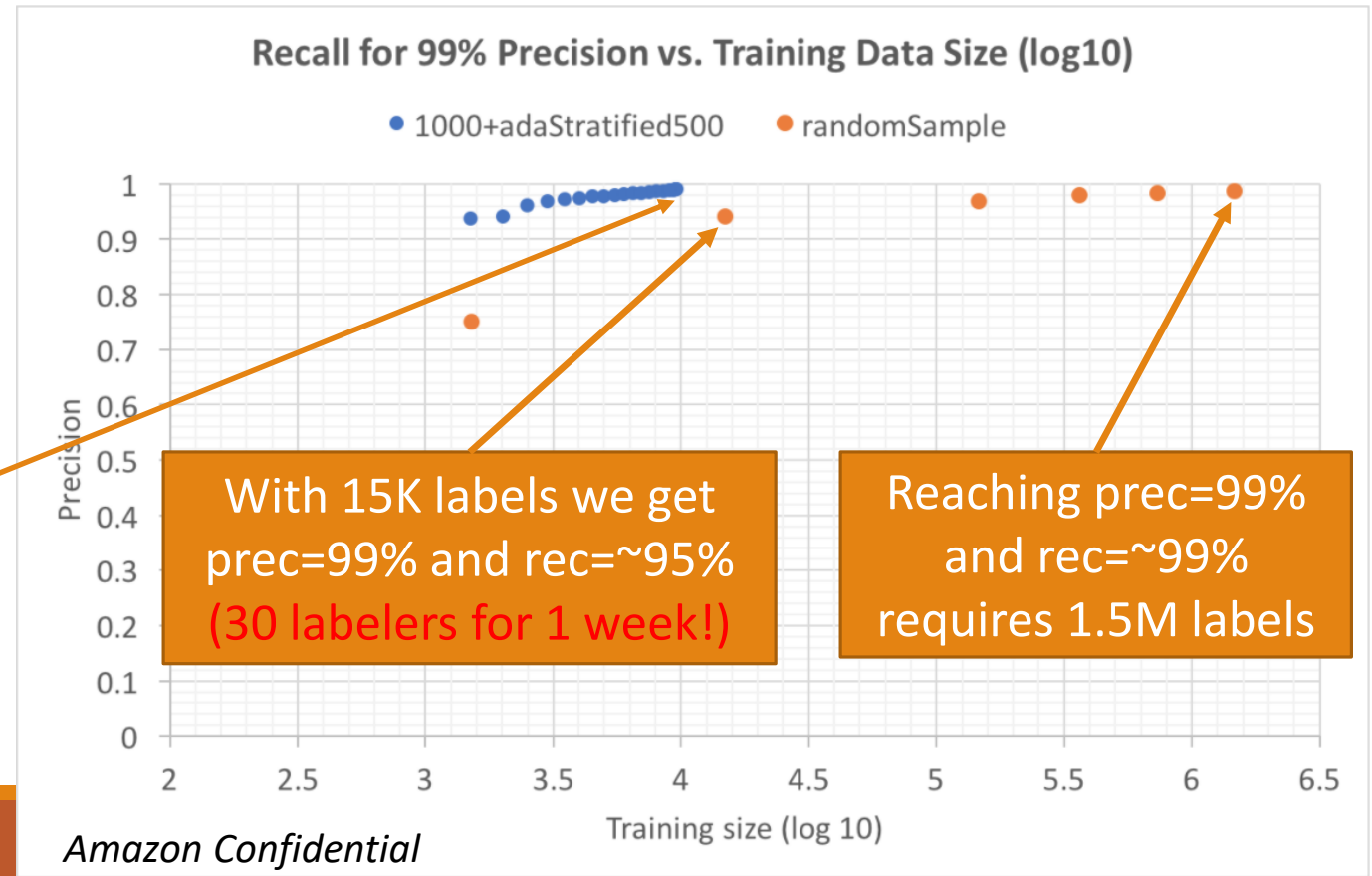
| | Precision | Recall |
|--------|-----------|--------|
| Movie | 99.0% | 98.7% |
| People | 99.3% | 99.6% |

1.5M labels

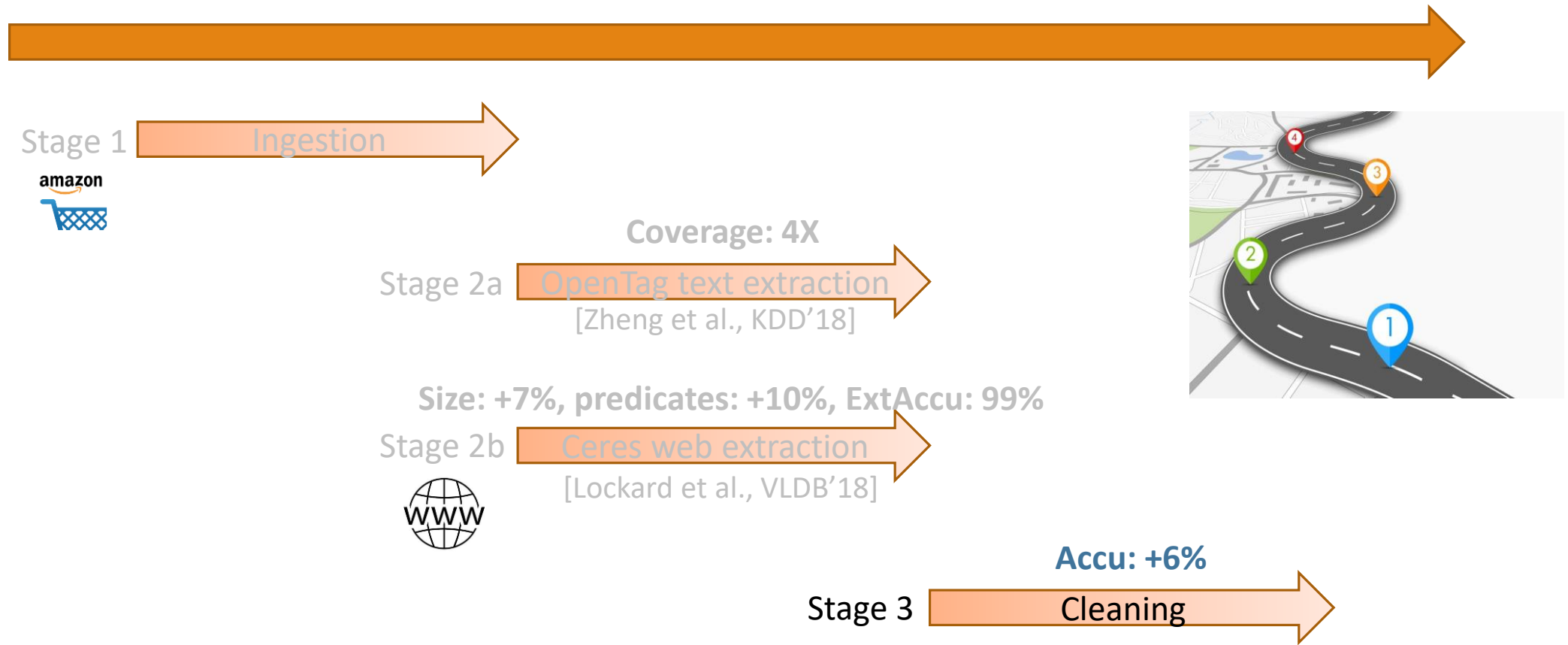


Stage 2b. Linking Entities Between Sources

- Apply active learning to minimize #labels



Stages in Building a Broad Graph



Stage 3. Supervised Knowledge Cleaning

Value noise

- Invalid brands: “1 lb”

- Lengthy descriptive brands: “The company X was founded in...”

- Unnormalized values: “Reese’s”, “reese”, “REESE’s”

Inconsistency

- E.g., `isSugarFree && sugarPerServing > 0`

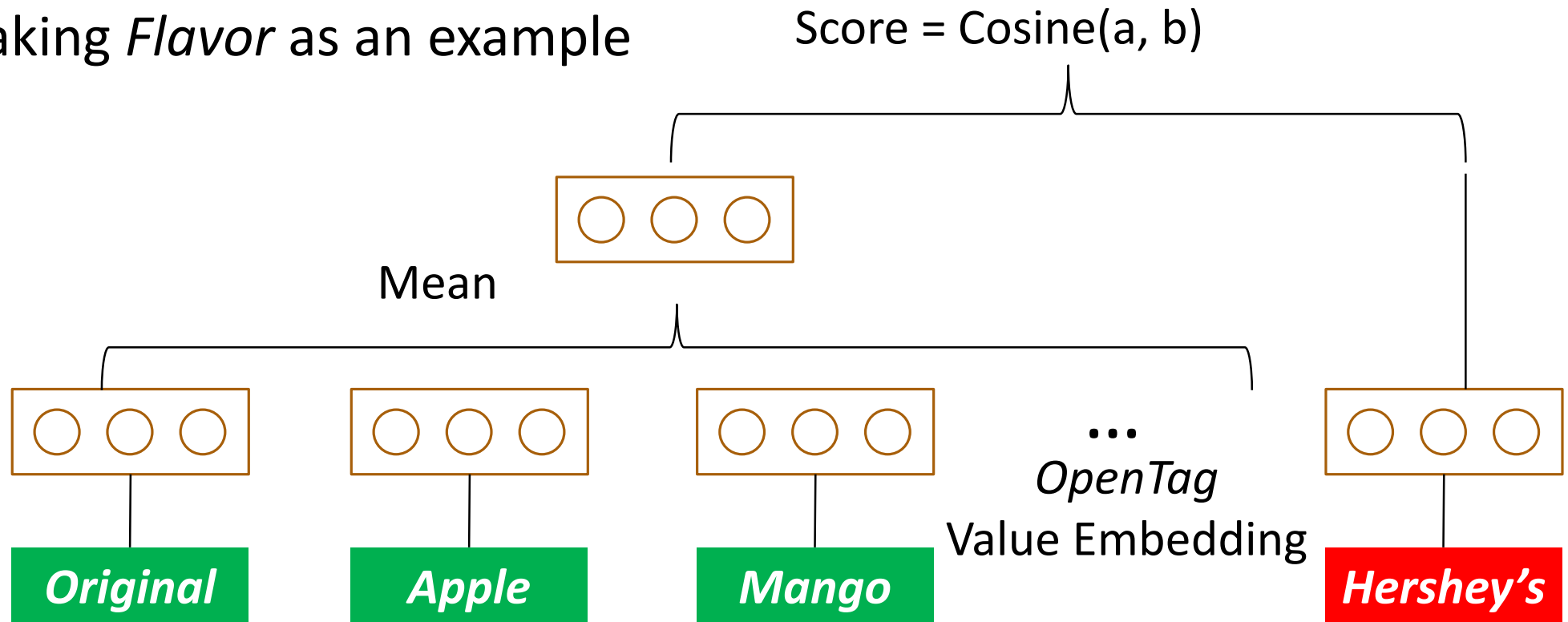
Stage 3. Supervised Knowledge Cleaning I

- A discriminative model to tell: Brand or Not a Brand?
- Features: frequency count, string lengths, special word hit, etc.
- Precision = 99.1%

Stage 3. Supervised Knowledge Cleaning II

Embedding-based filtering

Taking *Flavor* as an example



Stage 3. Supervised Knowledge Cleaning III

- Relation embedding filtering

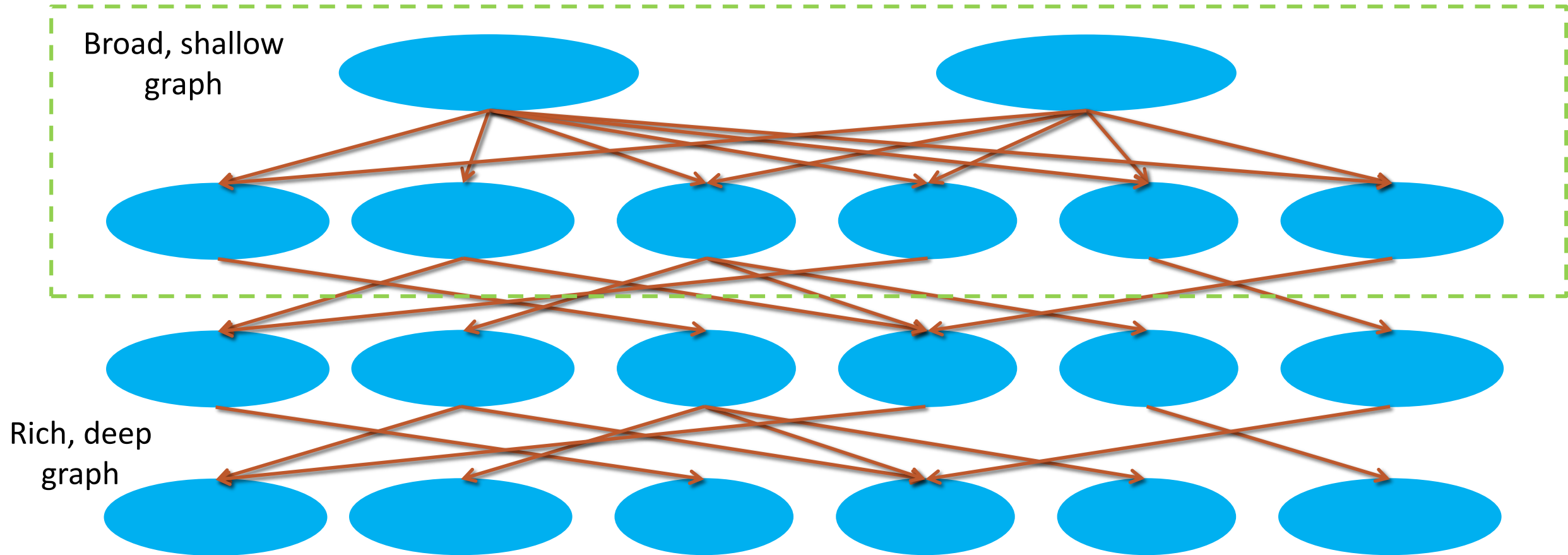
- Learn embedding from IMDb data and identify WikiData errors

| Subject | Relation | Target | Reason |
|--------------------------|-----------------|---------------------|--------------------|
| The Moises Padilla Story | writtenBy | César Ámigo Aguilar | Linkage error |
| Bajrangi Bhaijaan | writtenBy | Yo Yo Honey Singh | Wrong relationship |
| Piste noire | writtenBy | Jalil Naciri | Wrong relationship |
| Enter the Ninja | musicComposedBy | Michael Lewis | Linkage error |
| The Secret Life of Words | musicComposedBy | Hal Hartley | Cannot confirm |
| ... | ... | ... | ... |

Broad Graph Summary

- ❑ Quick development: ~15 person X month
- ❑ Easy ontology design and extension: 2W design
- ❑ Higher coverage: +8%~+68%
- ❑ Web extraction: +7% new triples, +10% new predicates, with 99% extraction accuracy
- ❑ Higher accuracy: +2%~+22%
- ❑ Live on Alexa Shopping, being deployed in Amazon Search and Browse

From Broad Graph to Rich Graph



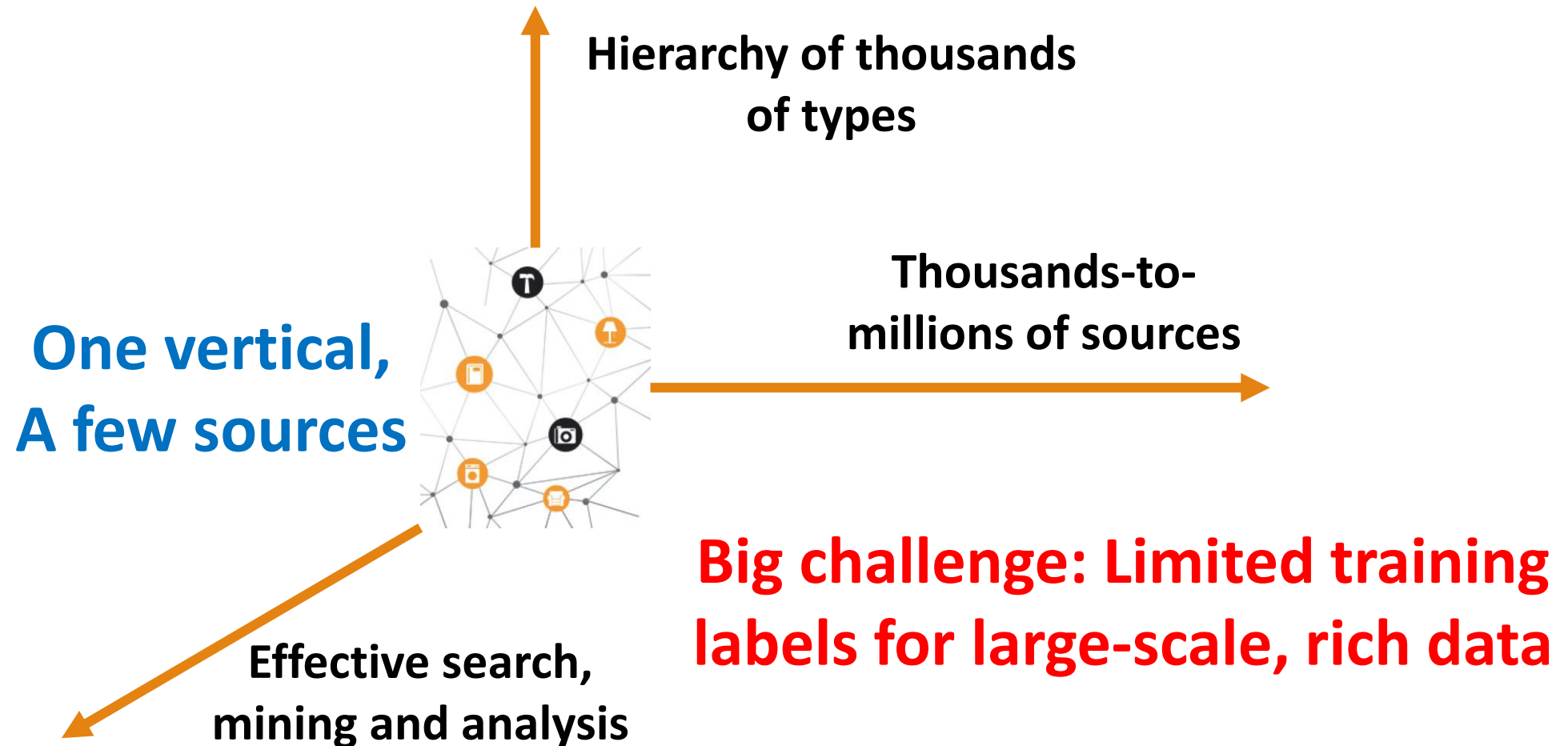
Are We There Yet?

Success Criteria

- Success criteria for Product Graph
 - #Papers?
 - Team size?
 - #Production applications?
 - \$\$?

- Becoming part of people's daily lives**

Knowledge Graph Construction Requires Data Integration & Cleaning (DI & DC)?



How to Get DI & DC to the Next Level of Success?

- ❑ Challenges: Limited training labels for large-scale, rich data
- ❑ Solution: Unsupervised learning

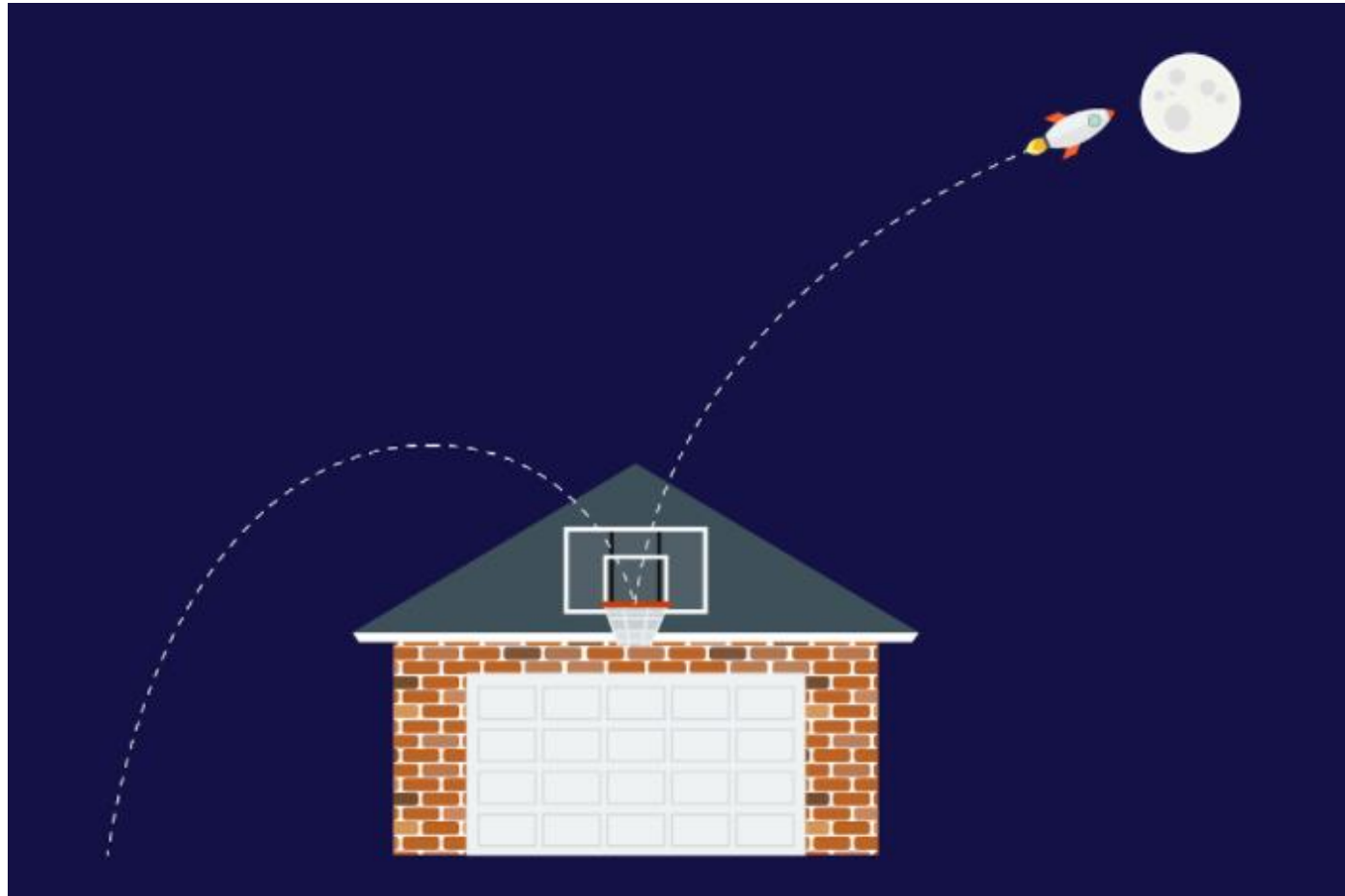


How to Get DI & DC to the Next Level of Success?

- Challenges: Limited training labels for large-scale, rich data
- Solution: Learning with limited labels ✓
 - Active learning
 - Weak learning (e.g., distance supervision, data programming)
 - Semi-supervised learning (e.g., graph-based learning)
 - Transfer learning
 - One/few-shot learning

Research Philosophy

Roofshots: Deliver incrementally and make production impacts



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

Moonshot: Open Knowledge Extraction and QA from Semi-Structured Web

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

Top Gun (1986) ★ 6.9 10 234,144 Rate This

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

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

豆瓣评分 **7.9** ★★★★★
 166740人评价

| | |
|----|-------|
| 5星 | 26.7% |
| 4星 | 45.0% |
| 3星 | 25.6% |
| 2星 | 2.3% |
| 1星 | 0.4% |

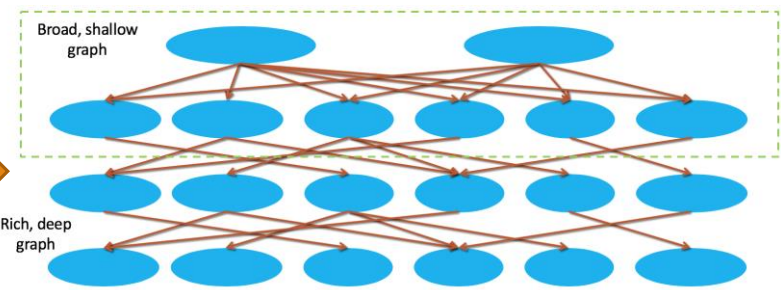
好于 92% 武侠片
 好于 90% 动作片

想看 看过 评价: ☆☆☆☆☆

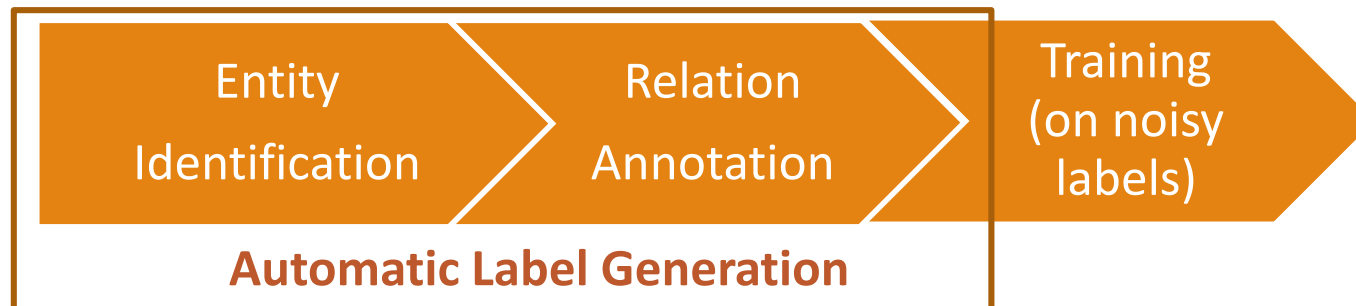
写短评 写影评 + 提问题 分享到

卧虎藏龙的剧情介绍 ·····

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, 但玉蛟龙却任性任气不听劝阻..... ©豆瓣



Ceres: Automatic ClosedIE from Semi-structured Web [VLDB'18]



Movie entity

Genre Release Date

Runtime

Director

Actors

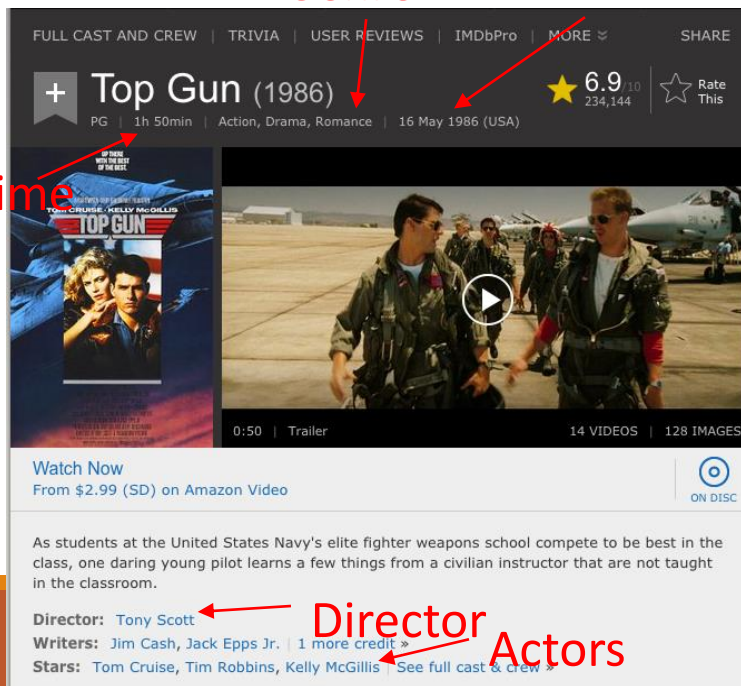





FULL CAST AND CREW | TRIVIA | USER REVIEWS | IMDbPro | MORE  | SHARE
 + **Top Gun (1986)**  **6.9** 1.0
234,144  Rate This
 PG | 1h 50min | Action, Drama, Romance | 16 May 1986 (USA)


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

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  | SHARE
 + **Top Gun (1986)**  **6.9** 1.0
234,144  Rate This
 PG | 1h 50min | Action, Drama, Romance | 16 May 1986 (USA)

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

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")

Ceres: Automatic ClosedIE from Semi-structured Web [VLDB'18]



Movie entity

Genre Release Date

Top Gun (1986)

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

6.9 ¹⁰ / 234,144

Director: Tony Scott

Writers: Jim Cash, Jack Epps Jr. | 1 more credit >

Stars: Tom Cruise, Tim Robbins, Kelly McGillis | See full cast & crew >

Runtime

Top Gun (1986)

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

6.9 ¹⁰ / 234,144

Director: Tony Scott

Writers: Jim Cash, Jack Epps Jr. | 1 more credit >

Stars: Tom Cruise, Tim Robbins, Kelly McGillis | See full cast & crew >

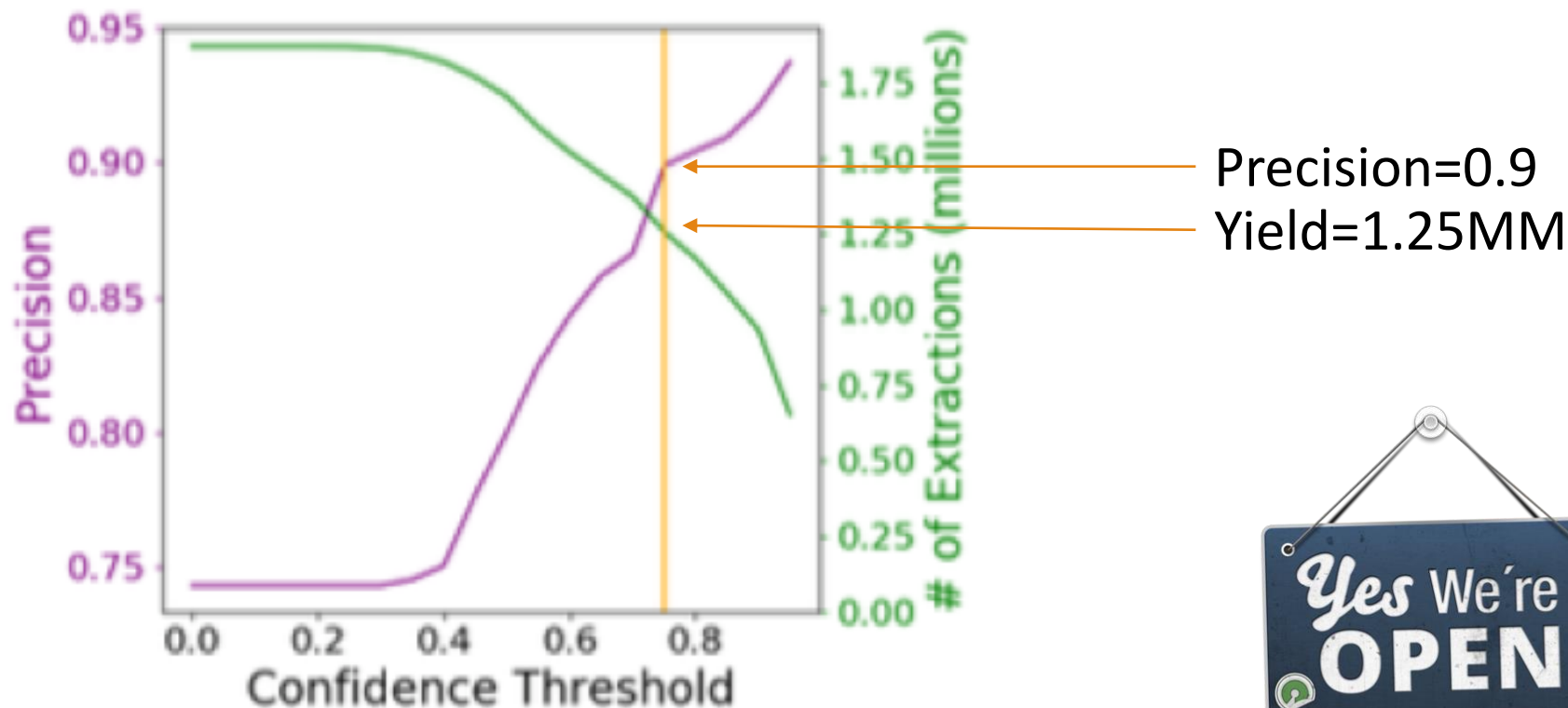
- Extracted triples
- (Top Gun, type.object.name, "Top Gun")
 - (Top Gun, film.film.genre, Action)
 - (Top Gun, film.film.directed_by, Tony Scott)
 - (Top Gun, film.film.starring, Tom Cruise)
 - (Top Gun, film.film.runtime, "1h 50min")
 - (Top Gun, film.film.release_date_s, "16 May 1986")

Ceres: Automatic ClosedIE from Semi-structured Web [VLDB'18]

| | |
|--|--|
| #Websites / #Webpages | 33 / 434K |
| Language | English and 6 other languages |
| Domains | Animated films, Documentary films, Financial performance, etc. |
| # Annotated pages | 70K (16%) |
| Annotated : Extracted #entities | 1 : 2.6 |
| Annotated : Extracted #triples | 1 : 3.0 |
| # Extractions | 1.25 M |
| Precision | 90% |

Ceres: Automatic ClosedIE from Semi-structured Web [VLDB'18]

□ Extraction on 33 movie websites with 7 languages



OpenCeres: OpenIE to Identify New Predicates [NAACL'19]

- ClosedIE: Only extracting facts corresponding to ontology
 - (“When Harry Met Sally...”, **film.film.directed_by**, “Rob Reiner”)

- OpenIE: Extract all relations expressed on the webpage
 - (“When Harry Met Sally...”, **“Director”**, “Rob Reiner”)



OpenCeres: OpenIE to Identify New Predicates [NAACL'19]

- ❑ ClosedIE: Normalize predicates by ontology
 - ❑ (“When Harry Met Sally...”, **film.film.directed_by**, “Rob Reiner”)

- ❑ OpenIE: Predicates are unnormalized strings
 - ❑ (“When Harry Met Sally...”, **“Directed By”**, “Rob Reiner”)

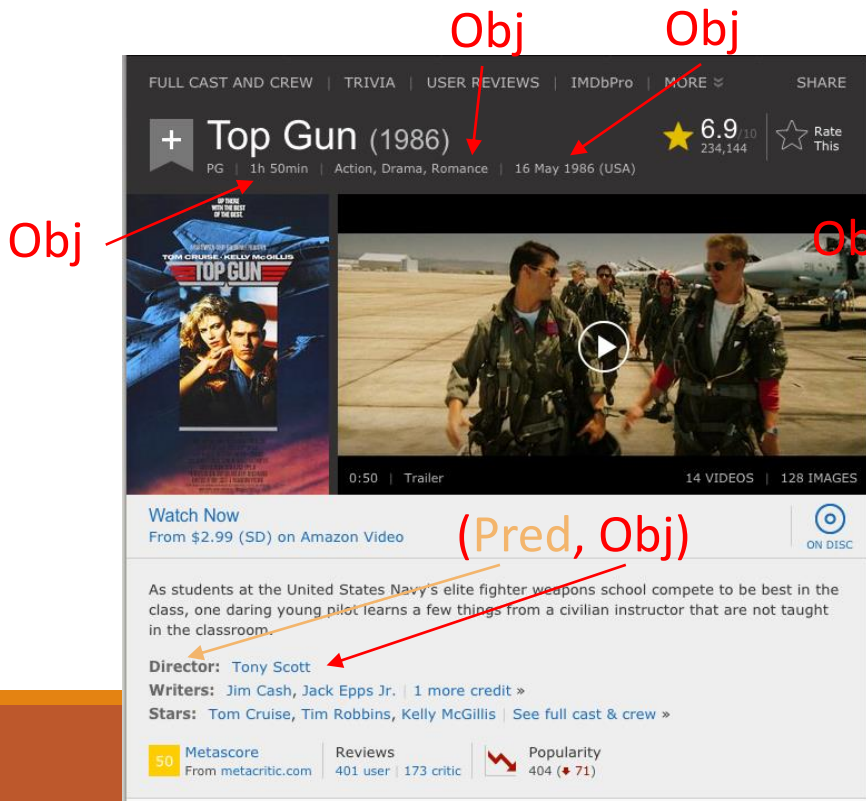


MOVIE INFO

Does sex make it impossible for men and women to b dilemma through the eleven year relationship between their own lives until they reconnect ten years later.

Rating: R
Genre: [Comedy](#), [Drama](#), [Romance](#)
Directed By: [Rob Reiner](#)
Written By: [Nora Ephron](#)
In Theaters: Jul 12, 1989 Wide
On Disc/Streaming: Oct 13, 1998
Runtime: 96 minutes

OpenCeres: OpenIE to Identify New Predicates [NAACL'19]

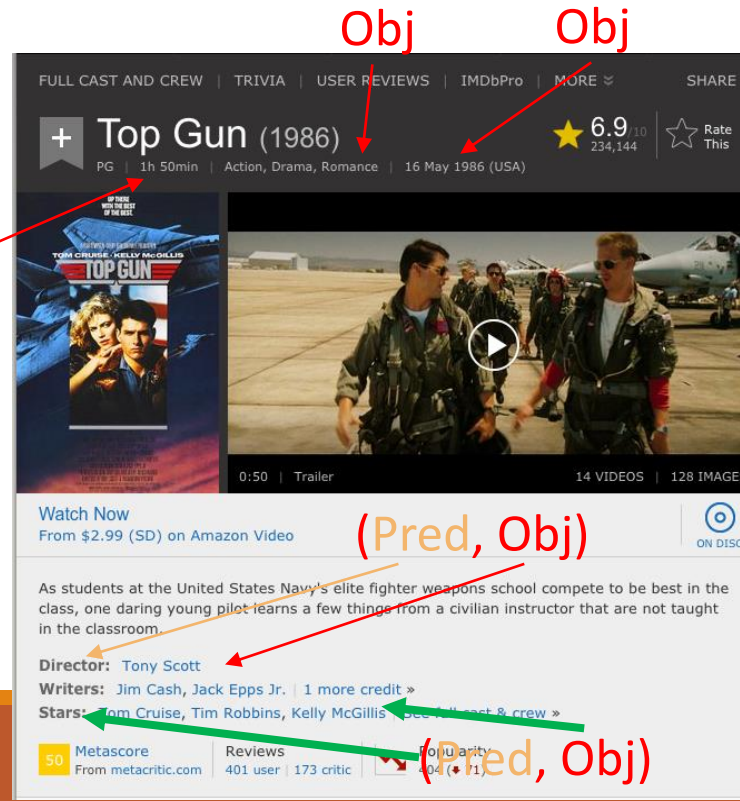



Obj Obj

Obj Obj

Obj

(Pred, Obj)



Obj Obj

Obj Obj

Obj

(Pred, Obj)

(Pred, Obj)

- Extracted triples
- (“Top Gun”, “Director”, “Tony Scott”)
 - (“Top Gun”, “Writers”, “Jim Cash”)
 - (“Top Gun”, “Writers”, “Jack Epps Jr.”)
 - (“Top Gun”, “Stars”, “Tom Cruise”)
 - (“Top Gun”, “Stars”, “Tim Robbins”)

OpenCeres: OpenIE to Identify New Predicates [NAACL'19]



Obj Obj

Obj

Obj

(Pred, Obj)

Full Cast and Crew | Trivia | User Reviews | IMDbPro | MORE | SHARE

+ **Top Gun (1986)** ★ 6.9/10
234,144 ☆ Rate This

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

0:50 | Trailer | 14 VIDEOS | 128 IMAGES

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

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 »

50 Metascore From metacritic.com | 401 user | 173 critic | Popularity 404 (↑ 71)

Obj Obj

Obj

Obj

(Pred, Obj)

(Pred, Obj)

Full Cast and Crew | Trivia | User Reviews | IMDbPro | MORE | SHARE

+ **Top Gun (1986)** ★ 6.9/10
234,144 ☆ Rate This

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

0:50 | Trailer | 14 VIDEOS | 128 IMAGES

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

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 »

50 Metascore From metacritic.com | 401 user | 173 critic | Popularity 404 (↑ 71)

- Extracted triples
- ("Top Gun", "Director", "Tony Scott")
 - ("Top Gun", "Writers", "Jim Cash")
 - ("Top Gun", "Writers", "Jack Epps Jr.")
 - ("Top Gun", "Stars", "Tom Cruise")
 - ("Top Gun", "Stars", "Tim Robbins")

OpenCeres: OpenIE to Identify New Predicates [NAACL'19]

Movie

- Seed: Director, Writer, Producer, Actor, Release Date, Genre, Alternate Title
- New: Country, Filmed In, Language, MPAA Rating, Set In, Reviewed by, Studio, Metascore, Box Office, Distributor, Tagline, Budget, Sound Mix

NBA Player

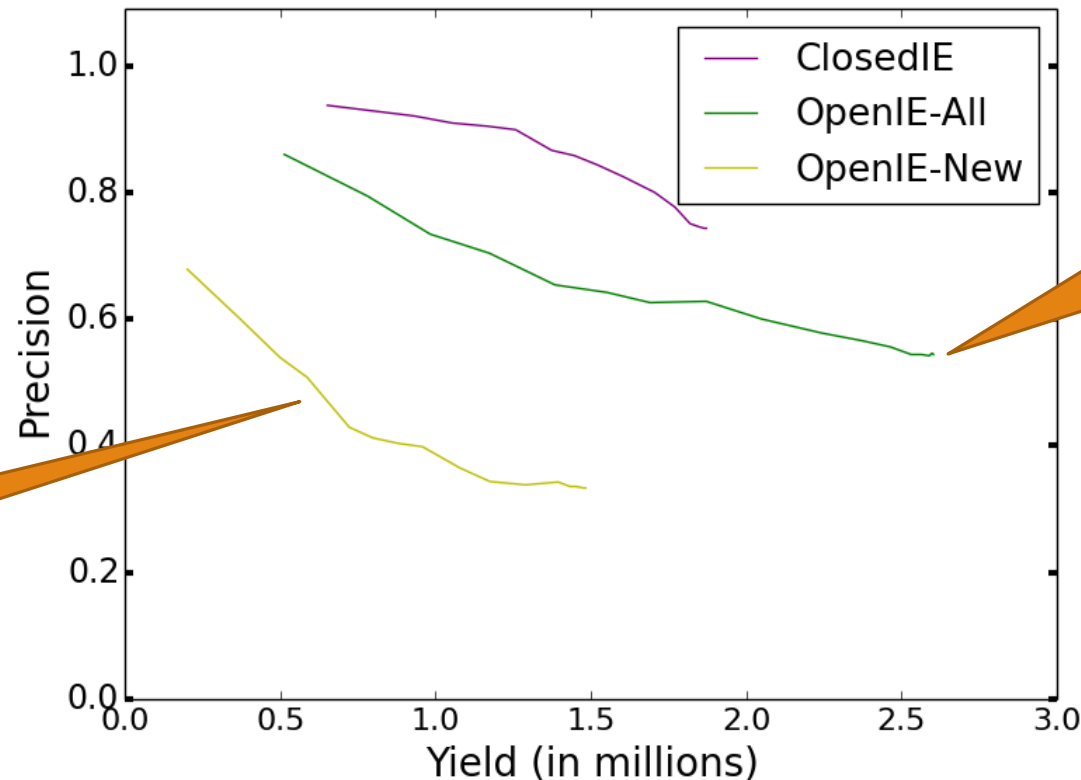
- Seed: Height, Weight, Team
- New: Birth Date, Birth Place, Salary, Age, Experience, Position, College, Year Drafted

University

- Seed: Phone Number, Web address, Type (public/private)
- New: Calendar System, Enrollment, Highest Degree, Local Area, Student Services, President

OpenCeres: OpenIE to Identify New Predicates [NAACL'19]

□ Extraction on long-tail movie websites

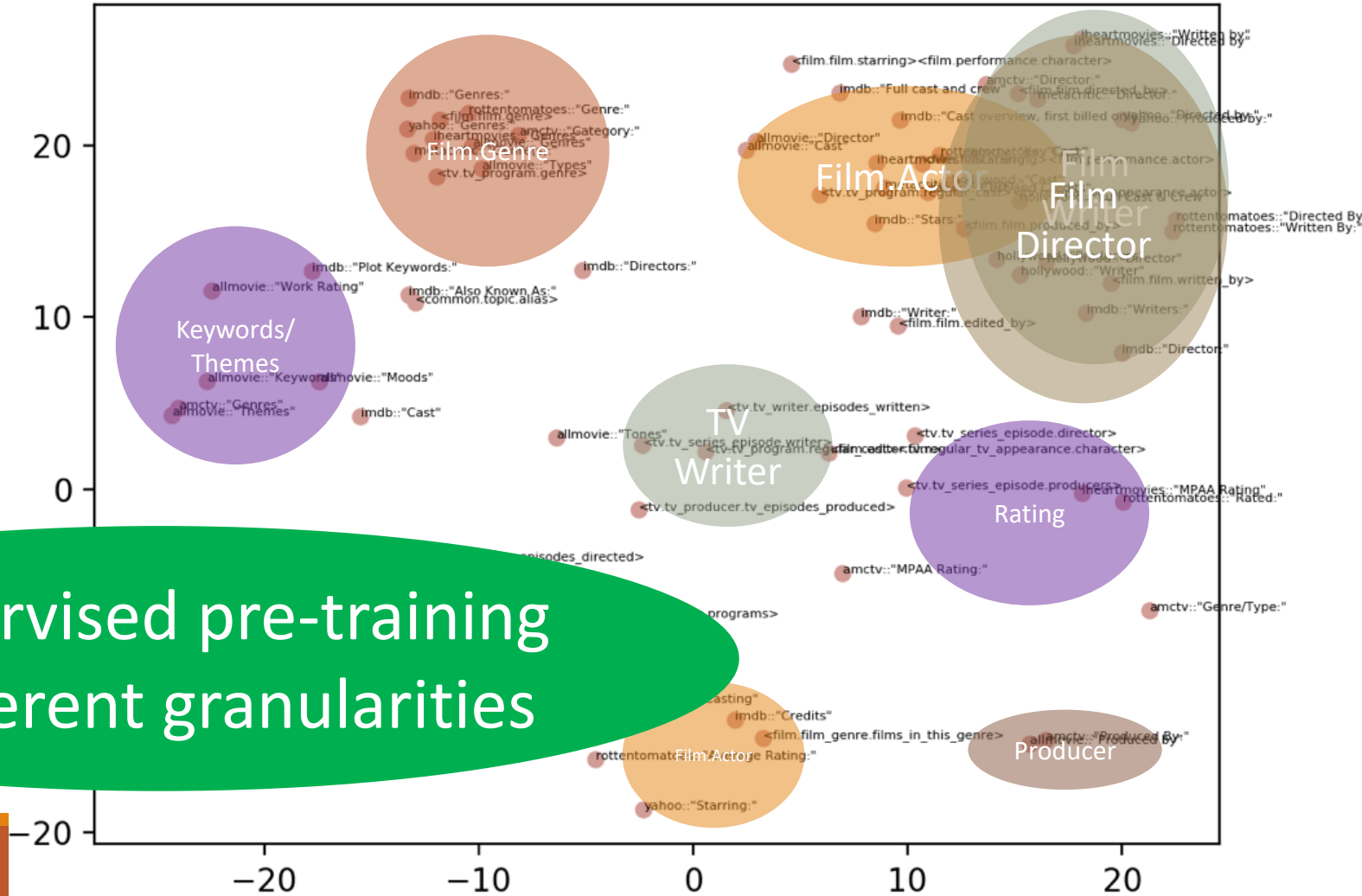


Still need prec improvement on new relations

OpenIE added significant amount of knowledge



OpenKI: Universal Schema to Cluster OpenIE Predicates [NAACL'19]



Unsupervised pre-training
for different granularities

QA Preliminary Results:

Question → Predicate

❑ Easy questions we get right

- ❑ “How many calories are there in X” -> caloriesPerServing
- ❑ “Can I eat X while on a low carbohydrates diet?” -> carbsPerServing
- ❑ “Is X nonGMO?” -> isGMOfree

❑ Hard questions we get right

- ❑ “Is X OK for people with celiac disease” -> isGlutenFree
- ❑ “Was X grown without pesticides” -> IsOrganic
- ❑ “How long can I store X” -> shelfLife
- ❑ “How many ounces of X do I get” -> Weight
- ❑ “Is X a Halal food” -> IsKosher

QA Preliminary Results: Questions → Question

is it vegetarian ?

is it vegetarian friendly

is this vegetarian friendly

is it vegetarian healthy

is vegetarian considered raw

is this contain msg

does it contain gelatine

does it contain gelatine ?

does it contain meat ?

are the ingredients all vegetarian ?

are the products any vegetarian

are any contain vegetarian ?

does it contain msg ?

does it contain gelatine ?

does it contain meat ?

were the tomatoes grown without pesticides ?

were the seed been grown irradiated ?

have the seed been chemically irradiated ?

have the seed been pesticide tested ?

have the honey been irradiated tested ?

does the seed contain sulfites ?

Is the fennel usda irradiated ?

does your fennel seeds irradiated

does these fennel seeds irradiated

are the seeds organic ?

are the seeds organic

are the tomatoes organic ?

QA Preliminary Results: Predicate → Question phrase

| | |
|----------------------|---|
| isGlutenFree | contains_wheat, gluten_free, gluten, fermented_soy, soy |
| packageSize | how_many_ounces, bulk_package, 1_lb, 10_oz, ounces, lb, bulk, oz, package |
| flavor | taste_like, taste, flavored, flavors |
| kosher | kosher_certified, kosher, halal, certified_kosher |
| fatPerServing | saturated_fat, saturated, fat |
| isVegetarian | suitable_for_vegetarians, meat, vegetarians, vegetarian |

Unsupervised
pre-training

StarFinder: Finding Entity Importance [Submission]

Top 15
Among
1MM
Artists

| Score | Name |
|-------|-------------------------|
| 8262 | Johann Sebastian Bach |
| 7039 | Wolfgang Amadeus Mozart |
| 5028 | Ludwig van Beethoven |
| 3591 | Frank Sinatra |
| 3118 | Elvis Presley |
| 3089 | Herbert von Karajan |
| 3052 | Frederic Chopin |
| 3001 | John Williams |
| 2944 | Antonio Vivaldi |
| 2746 | Manfred Eicher |
| 2638 | Norman Granz |
| 2616 | Grateful Dead |
| 2449 | Ella Fitzgerald |

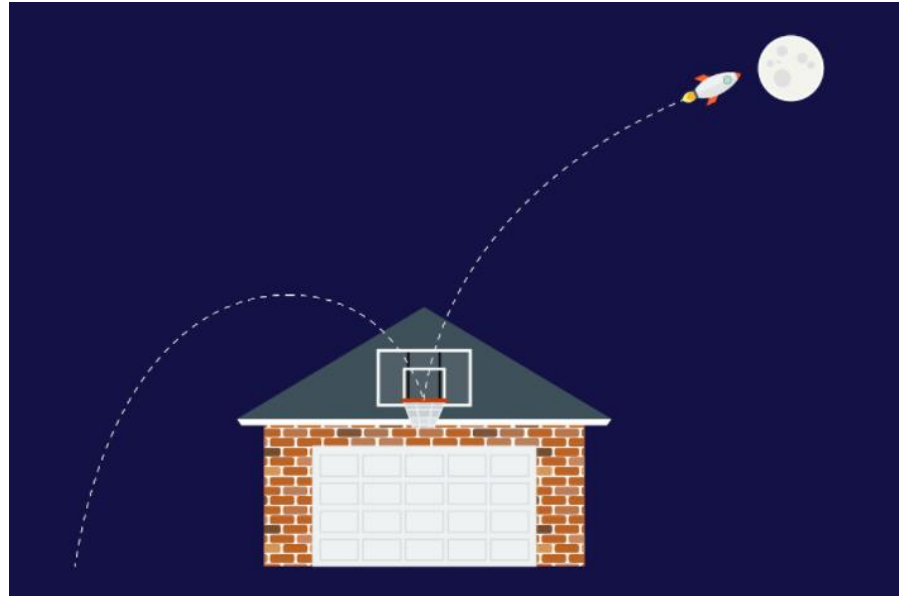
Exploring some other nodes

| Rank | Score | Name |
|------|-------|-----------------|
| 228 | 726 | hans zimmer |
| 267 | 668 | eminem |
| 282 | 650 | michael jackson |
| 369 | 559 | beatles |
| 632 | 412 | cher |
| 678 | 398 | taylor swift |
| 2184 | | |

Semi-supervised
learning with GCN

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
- ❑ The next-generation of KG could be a combination of **rich** graph and **broad** graph
- ❑ **Learning with limited labels** is the rescue to get DI & DC to the next level of success



Thank You!

QUESTIONS?