



16 PhD Dissertation Topics on Information Extraction

Mohit Bansal (UNC & Amazon Scholar)
Dan Bikel (Google)
Luna Dong (Amazon)
Eduard Hovy (CMU & DoD)
Heng Ji (UIUC & Amazon Scholar)
Dan Roth (UPenn & Amazon Scholar)

Eduard Hovy (CMU & DOD)

- Keeping IE results in the (neural) IE model
- Extracting propositional semantics from images+text
- IE with real-world semantics



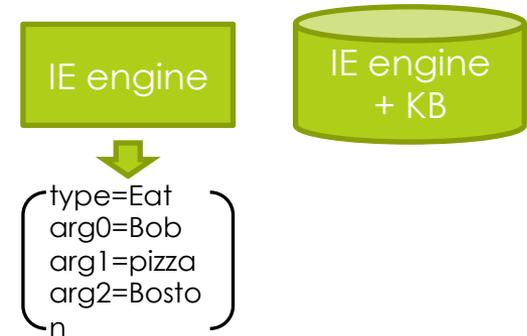
3 thesis topics in IE

Eduard Hovy

CMU and DARPA

1. Keeping IE results in the (neural) IE model

- The output of an IE engine is an instantiated relation or frame. This goes to a user or a KB
- Problems:
 - The IE engine doesn't change after what it has just read
 - The KB is static (just additive and doesn't generalize)
 - The user cannot change many IE parameters in a feedback/ repeat cycle; has to re-train the engine
- Idea: Store IE output 'inside' the DNN extractor!
 - It never forgets a fact it has read
 - It updates its new reading output based on what it has read
 - It even generalizes past knowledge based on new reading
- **Thesis topic: Self-extending reading+learning.** Create a joint neural model for IE+KB
 - Architecture: one deep NN or two?
 - How to formally define the structure/frame to be learned and extracted?
 - How to define the formal filler types? Is it necessary?
 - How can a user access/retrieve the results?
 - How can a user engine change IE parameters for re-extraction?
 - How can extracted frames be assembled, composed into bigger ones, and re-stored?



2. Extracting propositional semantics from images+text

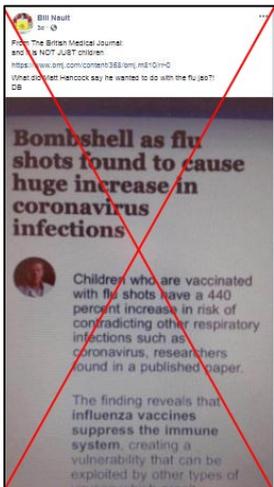
Images add a lot of info to text

Thesis topic: joint vision+text semantics extraction

- How to identify the author? Read and interpret logo (icons, names)
- How to identify epistemic values? Recognize and interpret visual negators (icons, text) plus the background semantics of what is being negated. False claims and their refutations may be overlaid on Informational Graphics
- How to extract timeline info? (title, axis numbering, axis headings, values)
- How to integrate results extracted from images and text? What result?

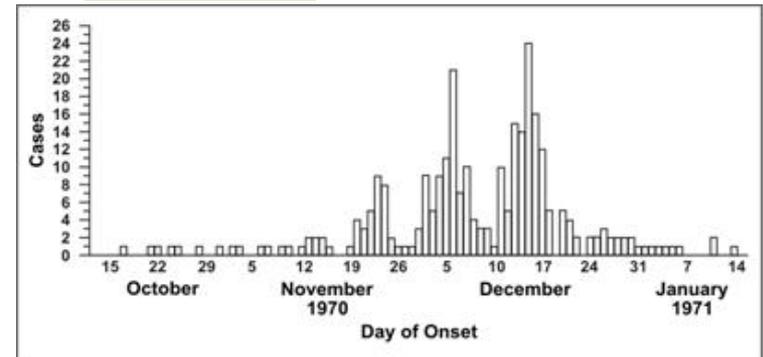


<https://twitter.com/fema/status/1253801441598353412>



<https://leadstories.com/hoax-alert/2020/06/fact-check-wearing-a-mask-does-not-remove-your-ability-to-conceal-weapon.html>

<https://factcheck.afp.com/no-evidence-flu-vaccines-make-children-more-vulnerable-coronavirus-infections>



Measles Cases by Date of Onset, October 15, 1970—January 16, 1971

3. IE with real-world semantics

- The IE we do today typically has 10–200 simple types/classes, nicely defined in a simple ontology
- But the semantics of the real world is messy and continuous. No realistic large ontologies are used in IE

□ Thesis topic: Doing IE with real-world semantics

- Use [an extract of] Wikidata as the ontology

- DWD is messy, changing, and not a clean taxonomy

- Use a continuous semantic distance function. Candidates:

- Embedding-based
- Hierarchy-based
- Comparison-based (tf.idf of associated words)
- Combination
- Machine learned, using WSD353 and others

- Use semantic composition of the IE output

- Extract and 'collapse' different representations with 'identical' semantics:
- Accept various frames as output
- How to score composed parts? e.g., *prevent = not(cause)*

Masks prevent Covid ≡ Wearing a mask causes the wearer to not be infected with Covid

Masks prevent Covid
Wearing a mask causes the wearer to not become infected with Covid

Goal: score them as identical

ID=E1
type=event.prevent_{Q1717246}
arg0=DWD.mask_{Q161542}
arg1=DWD.covid_{Q18975243}

ID=E2
type=event.wear_{Q?}
arg0=?var.X1
arg1=DWD.mask_{Q161542}

ID=E3
type=event.statechange.become-ill_{Q?}
arg0=?var.X1
arg1=DWD.covid_{Q18975243}
polarity=NEG_{Q190558}

ID=R1
type=relation.cause_{Q2574811}
arg0=E2
arg1=E3

DWD scores
mask—mask => 1.0
covid—covid => 1.0
prevention—cause => 0.63
prevention—negation => 0.51

Assessor
Matches all essential aspects => 1.0
Many parts match but not all => 0.8
A small amount of match only => 0.3
No match or I don't understand it => 0

N = ½ number of Qnodes counting both sides

Final score
= 1/N x assessor x sum(DWD scores for the sets)
= 1/4 x 1.0 x 3.14
= 0.78

Dan Roth (UPenn & Amazon Scholar)

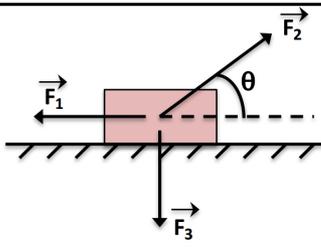
- Fact or Fiction: understanding and targeted summarization of long texts
- A broader notion of Grounding-Quantitative Reasoning from Natural Language Text
- Quantitative Reasoning from Natural Language Text



Reasoning about Quantities

- Extraction, contextualizing, scoping
 - Still open
- Reasoning about quantities
 - Mapping to Equations [Roy & Roth, 2015-2018]
- How to supervise to support this level of understanding?
 - Too many (ill-defined) latent decisions
 - End-to-End is unrealistic
- How to explain the reasoning?

The figure shows three forces applied to a trunk that moves leftward by 3.00 m over a frictionless floor. The force magnitudes are $F_1 = 5.00\text{N}$, $F_2 = 9.00\text{N}$, and $F_3 = 3.00\text{N}$, and the indicated angle is $\theta = 60.0^\circ$. During the displacement, what is the net work done on the trunk by the three forces?



The diagram shows a rectangular trunk on a horizontal surface. Three force vectors are applied to it: \vec{F}_1 points horizontally to the left, \vec{F}_2 points up and to the right at an angle θ from a horizontal dashed line, and \vec{F}_3 points vertically downwards.

What is her seizure frequency?

She reports working with a seizure frequency of 10 seizures now occurring up to 10/week, in clusters about 2-3 day/week. Previously reported seizures occurring about 2-3 times per month, often around the time of menses,...

How much did his challengers raise?

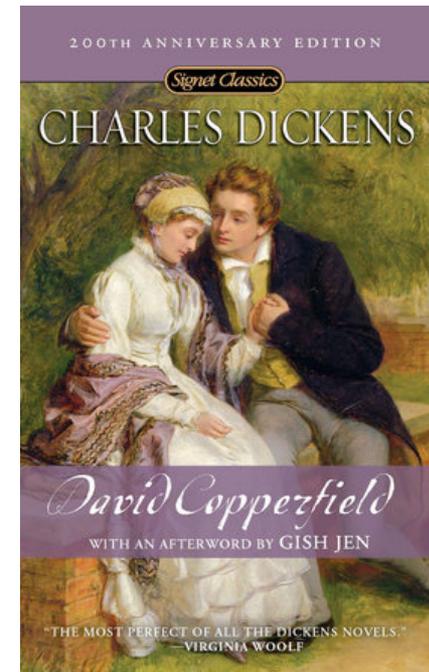
Mayor Rahm Emanuel has raised more than \$10 million toward his bid for a third term – more than five times the total raised by his 10 challengers combined, campaign finance records show.

What have you learned from this?

The COVID-19 pandemic in the United States is part of the As of October 2020, there were more than 9,000,000 cases and 230,000 COVID-19-related deaths in the U.S., representing 20% of the world's known deaths, the most deaths of any country.

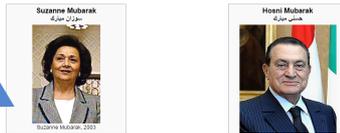
Facts and Fiction

- Reading a book
 - Written from a single or multiple perspectives
 - **Follow** events/situations/people **over time**
- A range of challenges to NLU and IE
 - The novel features the character [David Copperfield](#), his journey of change and growth from infancy to maturity, as many people enter and leave his life and he passes through the stages of his development. (Fiction, and you know it)
 - London and England in the 19-th century; socio-economic state, child exploitation; schools, prisons, emigration to Australia (true historical facts)
- Questions:
 - Identify Facts and Fiction
 - The Audible experience: What happened in the last 5 minutes?
 - Remind me:
 - Who is....(but don't spoil it for me): e.g, Steerforth
 - What does X think/feel about Y?



A broader notion of Grounding (Linking)

- A long tail of reasoning about relational information impacts accurate EDL
- Mubarak, the wife of deposed Egyptian President Hosni Mubarak,...



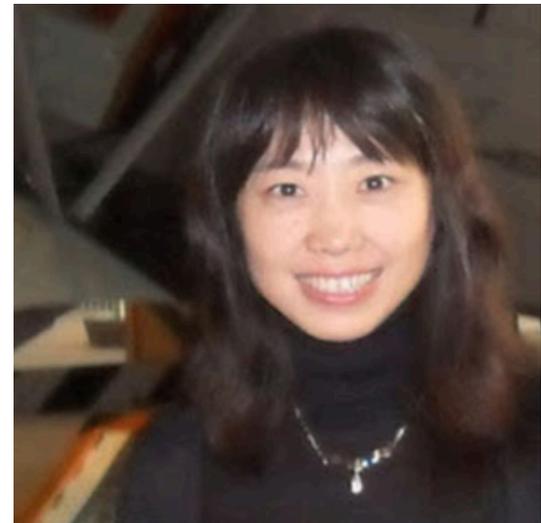
- The US has now managed to upset two of its **closest allies** by allowing the disclosure of sensitive information
 - Who are the allies? (time dependent)
- But understanding text requires grounding many other things

- Who is this?
- Where is it?
- What is happening?
- When did it happen?



Luna Dong (Amazon)

- Few-shot learning knowledge extraction from semi-structured websites in thousands of domains
- Multi-modal knowledge extraction across texts, images, tables, and semi-structured data



SEMI-STRUCTURED DATA

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 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

Reviews
401 user | 173 critic

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

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

Aamir Khan is receiving rave reviews for Dangal.

f Dangal
t Cast: Aamir Khan, Sakshi Tanwar, Fa Khurana, Sanya Malhotra
in Director: Nitesh Tiwari
G+ Rating: 4/5

Xin Luna Dong, Amazon→Facebook, lunadong@gmail.com

MINE THE GOLD

FULL CA

+

Yolie Cin

Watch | From \$2

As studi
class, or
in the cl

Directo
Writers
Stars:

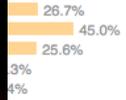
50 Me
Fro

Aamir Kh

卧虎藏龙 臥虎藏龍 (2000)

导演: 李安
编剧: 王惠玲 / 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...

豆瓣评分
7.9  166740人评价





推荐



卧虎藏龙的剧情介绍

Da 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
Ca: (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
Kh 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
Dir 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
Rat 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

- Industry success:
>90% extraction precision
- Limitations
 - Low recall
 - Restricted scale

I. DIVERSE DOMAINS

Movie

My Fair Lady

1964 · G · 2h 50min

IMDb RATING **7.8** / 10
59K

YOUR RATING ☆ Rate

POPULARITY **1,586** · 210

Cast & crew · User reviews · Trivia · IMDbPro · All topics



Trailer 1:08

8 VIDEOS

99+ PHOTOS

Drama Family Musical

See Showtimes and tickets

More watch options

Add to Watchlist

Director [George Cukor](#)

Writers [Alan Jay Lerner](#) (book of musical play) · [George Bernard Shaw](#) (from a play by)

Stars [Audrey Hepburn](#) · [Rex Harrison](#) · [Stanley Holloway](#)

354 User reviews · 106 Critic reviews · **95** Metascore

Game

The Tenno: monuments of an ancient warrior caste. For generations you've slept, with no call to awake you. Until now.

—In-game description

Awakening is the introduction quest to **WARFRAME**, designed to allow players to familiarize themselves with the basics of Warframe control and combat. It was introduced with [Update 29.0](#) (2020-08-25). Players are given a choice between various starting Warframes: [Excalibur](#), [Mag](#), and [Volt](#), along with a small selection of weapons.

- Contents [\[hide\]](#)
- 1 [Cinematic Intro](#)
 - 2 [Synopsis](#)
 - 3 [Walkthrough](#)
 - 4 [Notes](#)
 - 5 [Trivia](#)
 - 6 [Media](#)
 - 7 [Patch History](#)

Cinematic Intro [edit](#) | [edit source](#)

The intro, narrated by the [Lotus](#), tells the tale of the [Tenno](#). They were once an ancient warrior caste who controlled the [Warframes](#) and brought ruin to the [Orokin](#) Empire, before abruptly leaving. However, with the rise of the [Grineer](#) Empire, seeds from the ruins of [The Old War](#) that now swallow colonies whole, the Tenno begin to awaken.

Awakening



Information	
Introduced	Update 29.0 (2020-08-25)
Type	Main Quest
Requirement	First Login
Rewards	Starter Warframe Starter Weapons
Chronology	
Next Quest	Vor's Prize Replayable

I. DIVERSE DOMAINS

Podcast
How Your Data Powers Artificial Intelligence

SHOW NOTES

Artificial intelligence isn't something we'll see in the future. Thinking machines are already here, and nine powerful companies in the US and China control their development. The spam filter in your email inbox is AI. So are programs like Google Translate. The next level for thinking machines is when they begin learning the way humans learn. As artificial intelligence gets refined, who's keeping track of whether these machines share our motivations, desires, and hopes for the future of humanity? Amy Webb, author of "The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity," speaks with Alexis Madrigal, a staff writer for The Atlantic who covers Silicon Valley. The views and opinions of the podcast guests are their own and do not necessarily reflect those of the Aspen Institute.

SPEAKERS

- Amy Webb**
Founder of the Future Today Institute
- Alexis Madrigal**
Staff writer at The Atlantic

LEARN MORE

Additional Information

Resources

WEBSITE

Thinking Machines and the Future of Humanity - video

Course

Math and Logic > Math and Logic

Offered By **Duke UNIVERSITY**

Data Science Math Skills

★★★★★ 4.5 9,277 ratings

Daniel Egger [+1 more instructor](#)

Enroll for Free
Starts Jun 8

Financial aid available

263,988 already enrolled

Included with **COURSERA PLUS** Unlimited access to 3,000+ courses, Guided Projects, Specializations, and Professional Certificates. [Learn More](#)

[About](#) [Instructors](#) [Syllabus](#) [Reviews](#) [Enrollment Options](#) [FAQ](#)

About this Course

276,618 recent views

Data science courses contain math—no avoiding that! This course is designed to teach learners the basic math you will need in order to be successful in almost any data science math course and was created for learners who have basic math skills but may not have taken algebra or pre-calculus. Data Science Math Skills introduces the core math that data science is built upon, with no extra complexity, introducing unfamiliar ideas and math symbols one-at-a-time.

[SHOW ALL](#)

SKILLS YOU WILL GAIN

- Bayes' Theorem
- Bayesian Probability
- Probability
- Probability Theory

Xin Luna Dong, Amazon → Facebook, lunadong@gmail.com

I. DIVERSE DOMAINS

Dance

YOUR ULTIMATE GUIDE TO SWAN LAKE

HOME / DISCOVER / BACKSTAGE / YOUR ULTIMATE GUIDE TO SWAN LAKE

High Drama. Gorgeous Music. Ballet's Most Iconic Roles.

Helgi Tomasson's Swan Lake is part of SF Ballet's 2021 Digital Season. It will be performed in Program 07, streaming May 20–June 9.

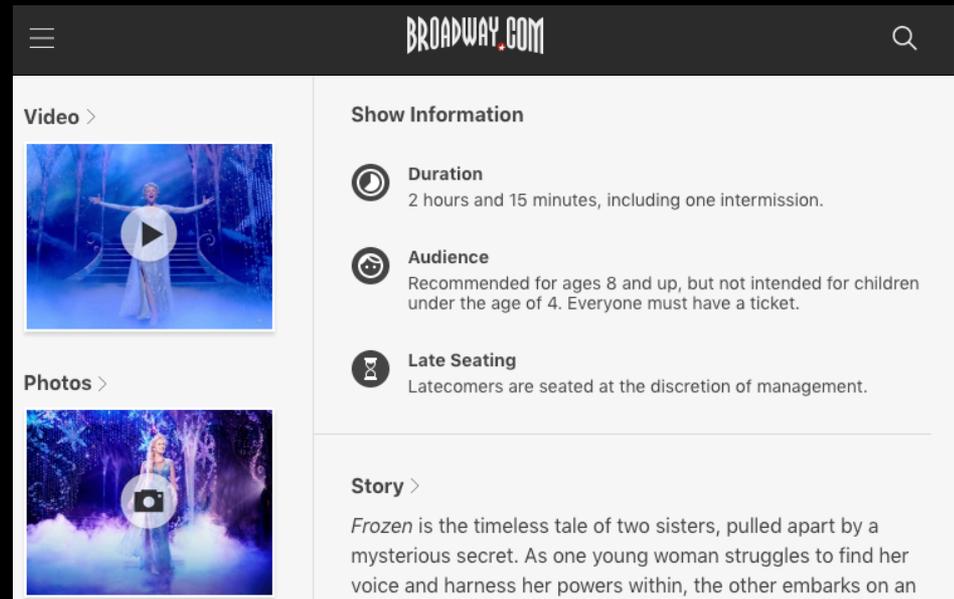
What Is It? Ballet's most famous love story, featuring one of the art form's most iconic roles. Thanks in part to this ballet's prevalence in pop culture (think: Natalie Portman in *Black Swan*, but also, Rudolf Nureyev on *The Muppets* or Taylor Swift's ballerinas in *Shake It Off*), swans and ballet have become nearly synonymous in the cultural imagination. That history—plus the high drama, thrilling dancing, and gorgeous music in Helgi Tomasson's 2009 production of *Swan Lake*—make it one of SF Ballet's most popular productions.

In Short? Tragic love; a Tchaikovsky score; and swans, swans, and more swans.

Who's It For? Fans of movies that make you cry, Olympic-level athletics, and bird-watching.

What Will I See? The story begins with a young woman named Odette . . .

Show



The screenshot shows the Broadway.com mobile interface for the show 'Swan Lake'. At the top, there is a navigation bar with a hamburger menu icon on the left, the 'BROADWAY.COM' logo in the center, and a search icon on the right. Below the navigation bar, the page is divided into two main columns. The left column features a 'Video >' section with a video player showing a ballerina in a white tutu on a stage with a blue background. Below the video is a 'Photos >' section with a photo of the same ballerina. The right column is titled 'Show Information' and contains three sections: 'Duration' (2 hours and 15 minutes, including one intermission), 'Audience' (Recommended for ages 8 and up, but not intended for children under the age of 4. Everyone must have a ticket.), and 'Late Seating' (Latecomers are seated at the discretion of management.). At the bottom of the right column is a 'Story >' section with a short paragraph of text.

Xin Luna Dong, Amazon→Facebook, lunadong@gmail.com

I. DIVERSE DOMAINS

노다지
A Bonanza (Nodaji)

1961년 · 대한민국 · 127분 · 1961-06-01 (개봉)
제작사 화성영화주식회사
감독 정창화
출연 **김승호** 황해, 엄앵란, 조미령, 허장강 [더보기](#)

Extracted knowledge triples
(**노다지** **출연** **김승호**)

**Topic 1. Few-shot learning
knowledge extraction from semi-
structured websites in thousands of
domains**

II. DIVERSE FORMATS OF INFORMATION

Image

Audrey Hepburn (I) (1929–1993) Top 5000

[Actress](#) | [Soundtrack](#)



1:56 | Trailer 36 VIDEOS | 1120 IMAGES

Audrey Hepburn was born as Audrey Kathleen Ruston on May 4, 1929 in Ixelles, Brussels, Belgium. Her mother, [Baroness Ella Van Heemstra](#), was a Dutch noblewoman, while her father, Joseph Victor Anthony Ruston, was born in Úzice, Bohemia, to English and Austrian parents. After her parents' divorce, Audrey went to London with her mother where she went ... [See full bio](#) »

Born: [May 4, 1929](#) in [Ixelles, Brussels, Belgium](#)

Died: [January 20, 1993](#) (age 63) in [Tolochenaz, Vaud, Switzerland](#)

Text

Semi-structured

Filmography

Jump to: [Actress](#) | [Soundtrack](#) | [Thanks](#) | [Self](#) | [A](#)

Show all | Show by... | Edit

Actress (34 credits)

Hide

Always

Hap

1989

Love Among Thieves (TV Movie)

Baroness Caroline DuLac

1987

They All Laughed

Angela Niotes

1981

Bloodline

Elizabeth Roffe

1979

Personal Details

Edit

Other Works: TV commercial using a combination of archive footage, CGI, and an Audrey Hepburn impersonator for Dove Chocolate Bars (archive footage and archive sound). [See more](#) »

Publicity Listings: [6 Biographical Movies](#) | [27 Print Biographies](#) | [4 Portrayals](#) | [1 Interview](#) | [23 Articles](#) | [12 Pictorials](#) | [77 Magazine Cover Photos](#) | [See more](#) »

Alternate Names: [Audrey](#)

Height: [5' 7"](#) (1.7 m)

Xin Luna Dong, Amazon→Facebook, lunadong@gmail.com

Image

II. DIVERSE FORMATS OF INFORMATION

Purina Pro Plan with Probiotics Shredded Blend High Protein, Adult Dry Dog Food Chicken & Rice (Packaging May Vary)

Visit the [Purina Pro Plan Store](#)
★★★★☆ 10,468 ratings

Price: **\$65.98** (\$1.40 / lb) ✓prime

Coupon Save an extra 20% on your first Subscribe & Save order.
[Details](#)

Style: **Dry Food**
Flavor Name: **Chicken & Rice**
Size: **47 Pound (Pack of 1)**

6 lb. Bag \$14.48 (\$2.41 / lb) ✓prime	18 Pound (Pack of 1) 9 options from \$37.99	35 Pound (Pack of 1) \$49.98 (\$1.43 / lb) ✓prime
47 Pound (Pack of 1) \$65.98 (\$1.40 / lb) ✓prime		

Text

Image text

Ingredients Chicken, Rice Flour, Whole Grain Wheat, Poultry By-Product Meal (Source Of Glucosamine), Soybean Meal, Beef Tallow Preserved With Mixed-Tocopherols, Corn Gluten Meal, Whol...

[See more](#) ▾

Brand Purina Pro Plan

Flavor Chicken & Rice

Item Weight 47 Pounds

Age Range Adult

Description

About this item

- One (1) 47 lb. Bag - Purina Pro Plan High Protein Dog Food With Probiotics for Dogs, Shredded Blend Chicken & Rice Formula
- Hard kibble combined with tender, shredded pieces for taste and texture dogs love
- High protein formula, with real chicken as the first ingredient
- Fortified with guaranteed live probiotics for digestive and immune health
- Used to be known as SAVOR Shredded Blend Chicken & Rice Formula

Semi-structured

Text

II. DIVERSE FORMATS OF INFORMATION

Image

Purina Pro Plan with Probiotics Shredded Blend High Protein, Adult Dry Dog Food Chicken & Rice (Packaging May Vary)

Visit the Purina Pro Plan Store

★★★★★ 10,468 ratings

Price: **\$65.98** (\$1.40 / lb) ✓prime

Coupon Save an extra 20% on your first Subscribe & Save order. Details

Style: **Dry Food**

Flavor Name: **Chicken & Rice**

Size: **47 Pound (Pack of 1)**

6 lb. Bag \$14.48 (\$2.41 / lb) ✓prime	18 Pound (Pack of 1) 9 options from \$37.99	35 Pound (Pack of 1) \$49.98 (\$1.43 / lb) ✓prime
47 Pound (Pack of 1) \$65.98 (\$1.40 / lb) ✓prime		

Text

Image text

Ingredients Chicken, Rice Flour, Whole Grain Wheat, Poultry By-Product Meal (Source Of Glucosamine), Soybean Meal, Beef Tallow Preserved With Mixed-Tocopherols, Corn Gluten Meal, Whol...

[See more](#) ▾

Brand Purina Pro Plan

Flavor Chicken & Rice

Item Weight 47 Pounds

Age Range Adult

Description

About this item

- One (1) 47 lb. Bag - Purina Pro Plan High Protein Dog Food With Probiotics for Dogs, Shredded Blend Chicken & Rice Formula
- Hard kibble combined with tender, shredded pieces for taste and texture dogs love
- High protein formula, with real chicken as the first ingredient
- Fortified with guaranteed live probiotics for digestive and immune health
- Used to be known as SAVOR Shredded Blend Chicken & Rice Formula

Semi-structured

Text

Topic 2. Multi-modal knowledge extraction across texts, images, tables, and semi-structured data

Xin Luna Dong, Amazon→Facebook, lunadong@gmail.com

RESEARCH QUESTIONS

- **Research question 1.** Can we pre-train over vast number of sources over diverse domains on different types of information?
- **Research question 2.** Shall we do transfer learning or one-size-fits all and how many models do we need to model all kinds of diversity?
- **Research question 3.** How much can different modals leverage each other in multi-modal modeling and in co-training?

Dan Bikel (Google)



- Thesis topic 1: Incremental approaches to entity understanding and discovery
 - Large language models such as BERT, GPT-[23] and T5 are very expensive to train, and yet the information that we would like our models to take advantage of is constantly changing. This thesis explores ways to deal with incrementally arriving data that pertains to both existing and new entities of interest. We explore methods of fine-tuning large LM's in an efficient way, as well as methods of discovering and finding facts about new entities as they arrive in both natural language (unstructured) text and structured knowledge bases.

- Thesis topic 2: Integration of structured KB's with large language models
 - Recent work has shown the promise of building large language models that have access to external knowledge, which is to say, structured knowledge that is independent of a language model's parameters. This thesis explores novel methods for integrating external knowledge into a large LM/NLU model.

- **Bonus** Thesis topic 3 [related to topic 2]: On the synthesis of retrieval and structured queries
 - Building dense representations of entities has been shown to be an accurate and scalable way of doing not only entity linking via retrieval. Similarly, other work has shown how it is possible to build large-scale neural network approaches to encoding structured queries. This Ph.D. thesis explores novel approaches to integrating methods for representation learning and retrieval with methods that can encode structured KB queries. In particular, we hope to do better at queries involving types or quantification, where only some of the entities can be reliably obtained from a KB using structured queries, and others can be obtained only through better retrieval from unstructured corpora.



Mohit Bansal (UNC & Amazon Scholar)

- Dynamic Spatio-Temporal Knowledge Extraction and Reasoning on Videos
- Multi-Hop Reasoning on Documents and Multimodal Information



Heng Ji (UIUC & Amazon Scholar)

- Combining symbolic semantics and distributional semantics for Information Extraction
- Schema-guided generation for Cross-media cross-lingual Information Extraction and Prediction
- Cross-media cross-lingual joint knowledge and text embedding representation