

# Intelligent Search

Svitlana Vakulenko

Postdoctoral researcher  
Informatics Institute  
University of Amsterdam

October 1, 2020



# Me

- Postdoctoral researcher in ILPS, Ivi, UvA
  - Conversational Search
- PhD at the Vienna University of Technology, Austria
  - [Knowledge-based Conversational Search](#)
- Internship at Apple Siri, Seattle WA USA
  - [Question Rewriting for Conversational QA](#)

# You?

- MSc? courses?
- IRO? ML/DL? KR/KG?
- year? thesis?

# Search



# Search results

[Content pages](#) [Multimedia](#) [Help and Project pages](#) [Everything](#) [Advanced](#) Results **1 - 20** of **31** for **Seneca Rocks**

► [Categories](#) [Other tools](#)

Create the page "**Seneca Rocks**" on this wiki!

Category: **Seneca Rocks**

United States Category: Spruce Knob-**Seneca Rocks** National Recreation Area

Category: River Knobs (West Virginia) Category: Climbing in West Virginia ...

16 members (0 subcategories, 16 files) - 02:28, April 6, 2011



File: **Seneca Rocks** - Homestead.jpg

Description en: Sites Homestead | Sites Homestead at en: **Seneca Rocks** | ...

**Seneca Rocks** in Monongahela National Forest, West Virginia, USA | ...

(3,790×1,950 (7,262,163 bytes)) - 16:14, February 25, 2010



File: **Seneca Rocks** - View from platform.jpg

0 files selected.

Select [all](#) / [none](#)

Cat-a-lot

# Search

- Information need



# Search

- Information need
- Query formulation



**hockey pocky**

# Search

- Information need
- Query formulation
- Index



**hockey pocky**



# Search

- Information need
- Query formulation
- Index
- Document collection

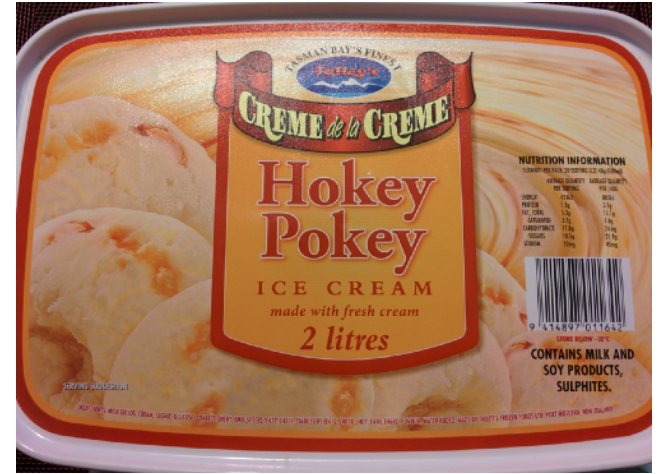


**hockey pocky**



# Search

- Information need
- Query formulation
- Index
- Document collection
- Ranking by relevance

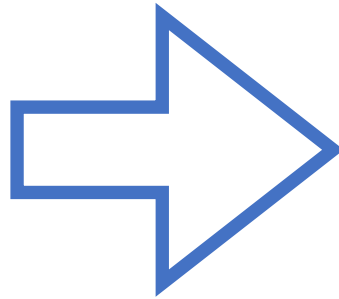
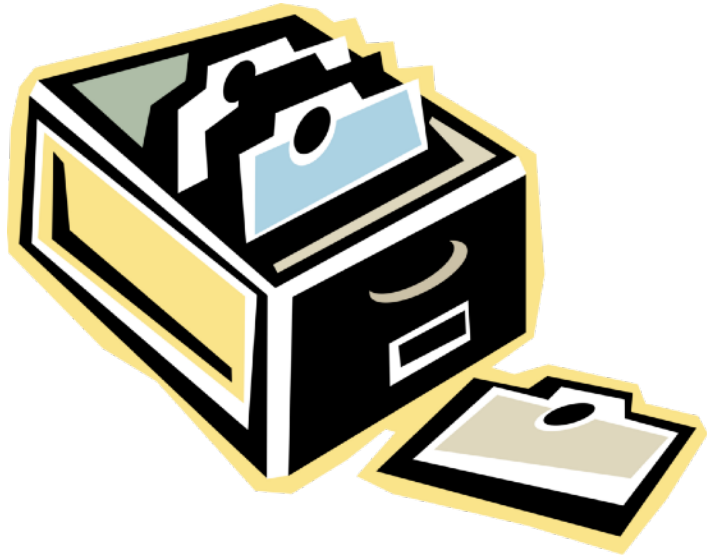


hochey pocky



# Conversational Search

# Conversational Search





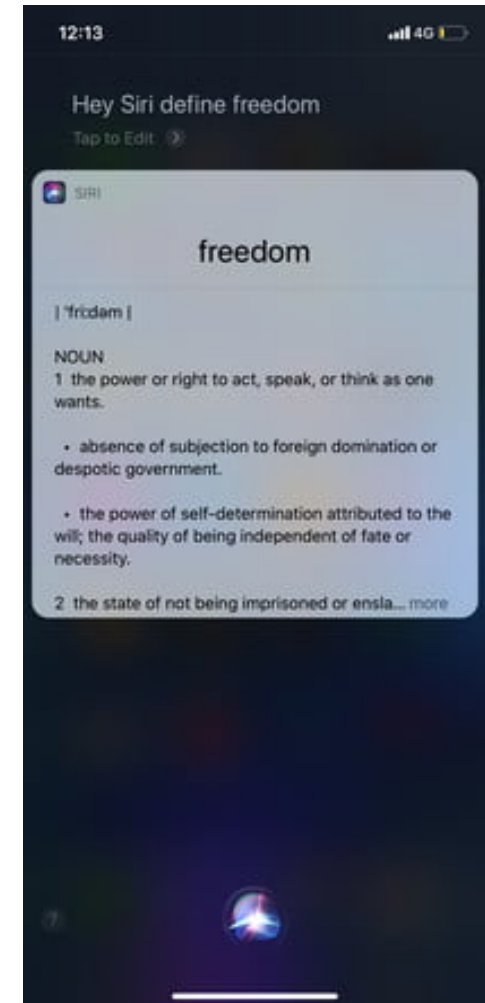
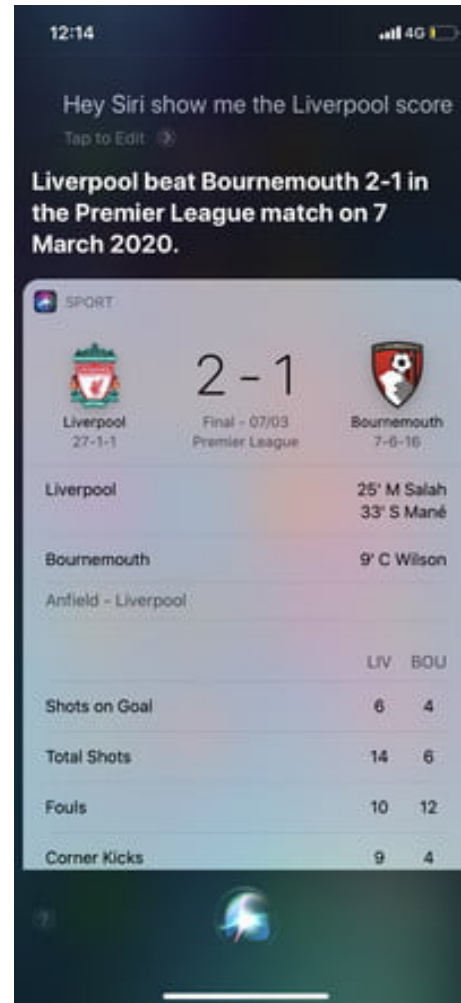
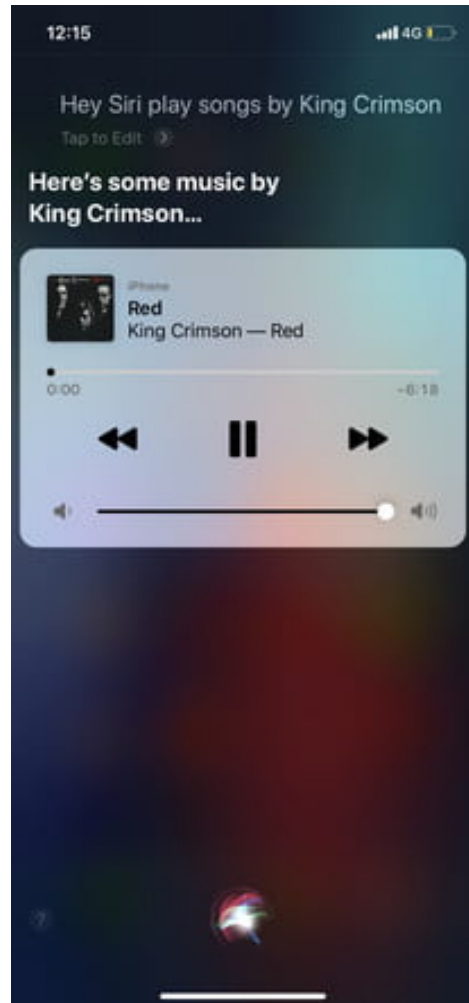
# Conversational Search

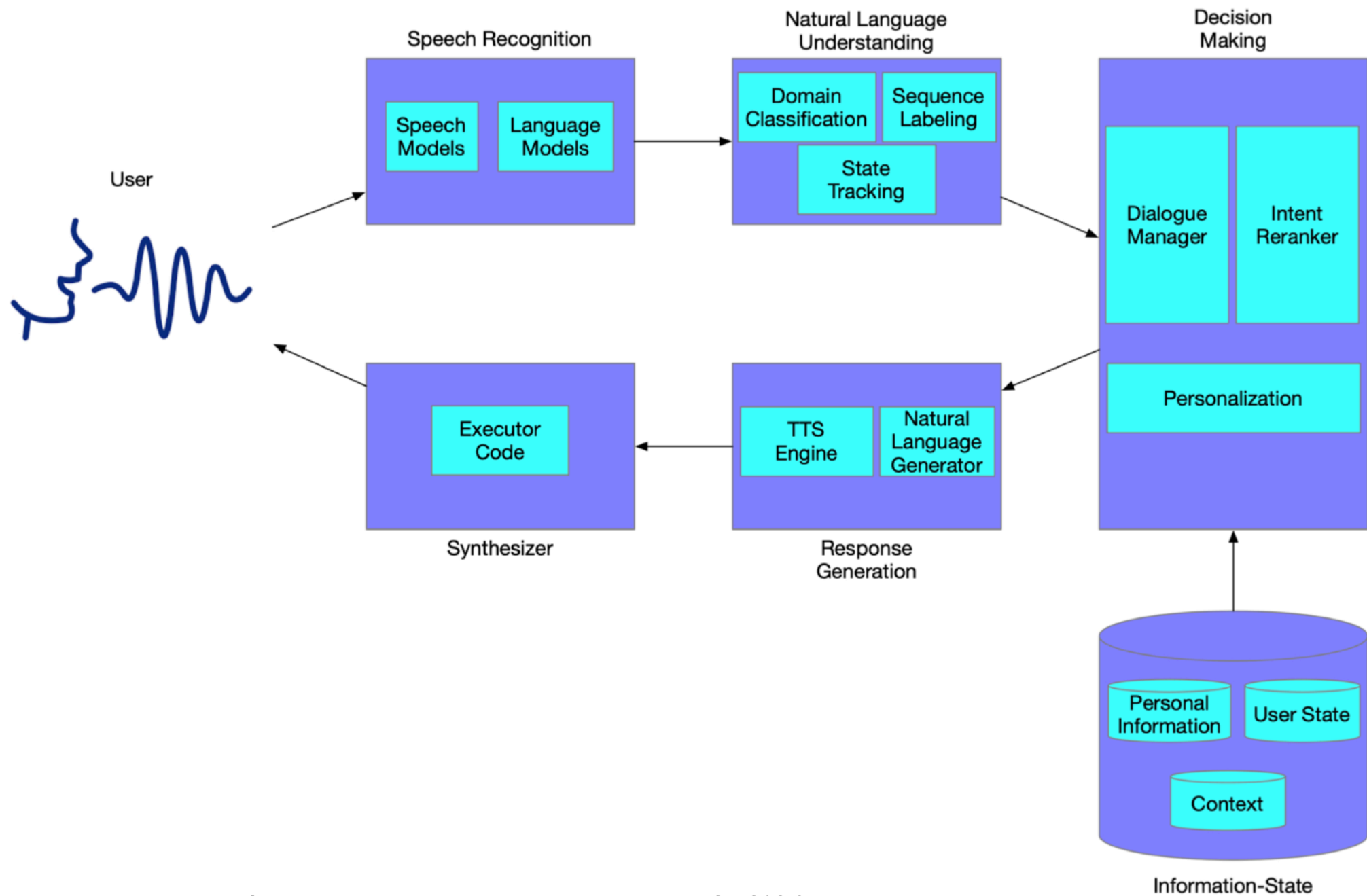
Information access using dialogue interface

# Conversational Search

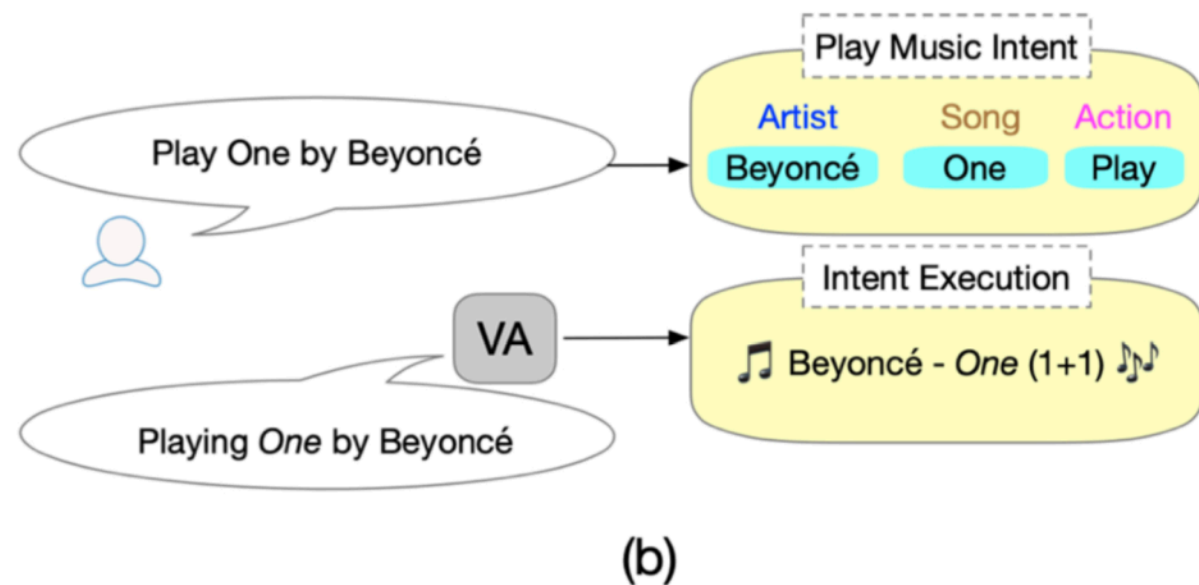
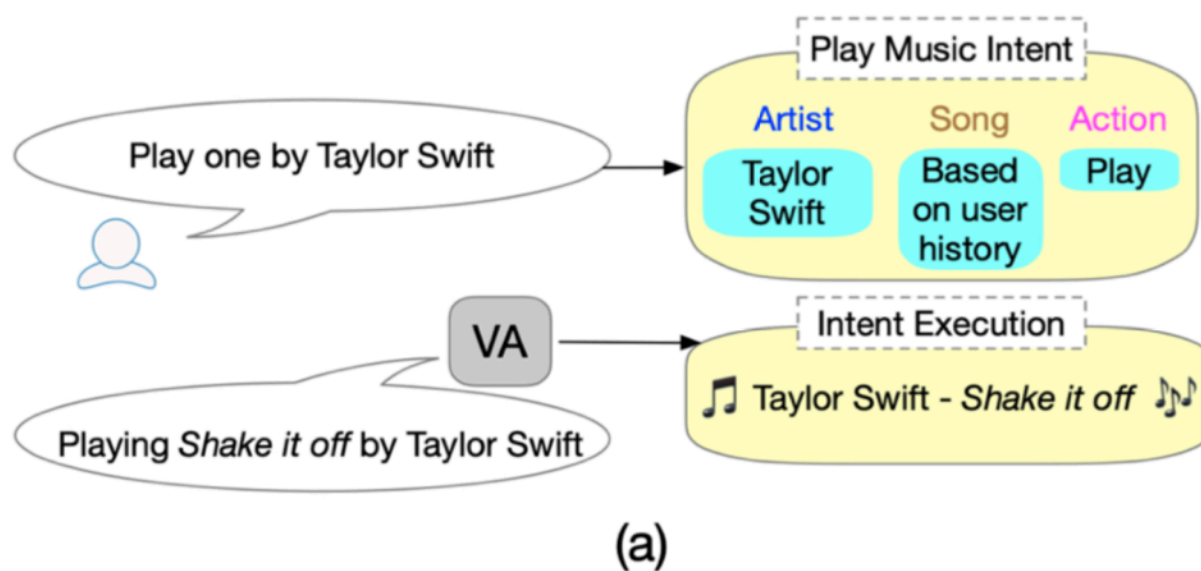


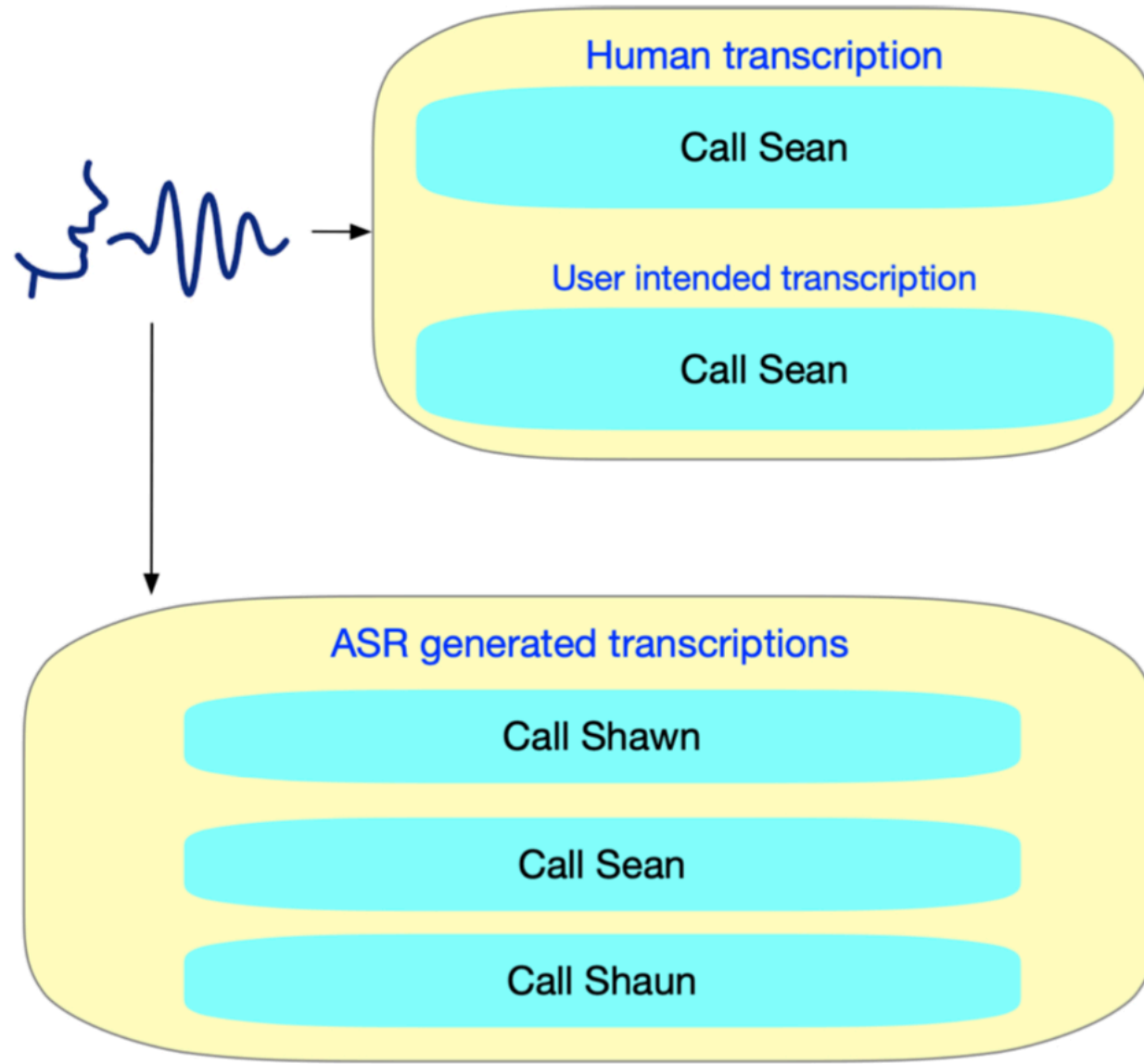
# Siri



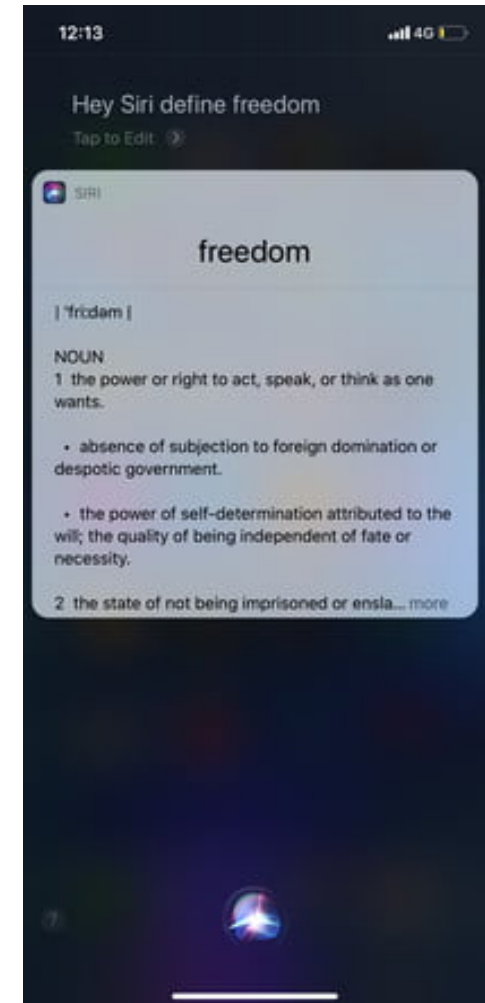
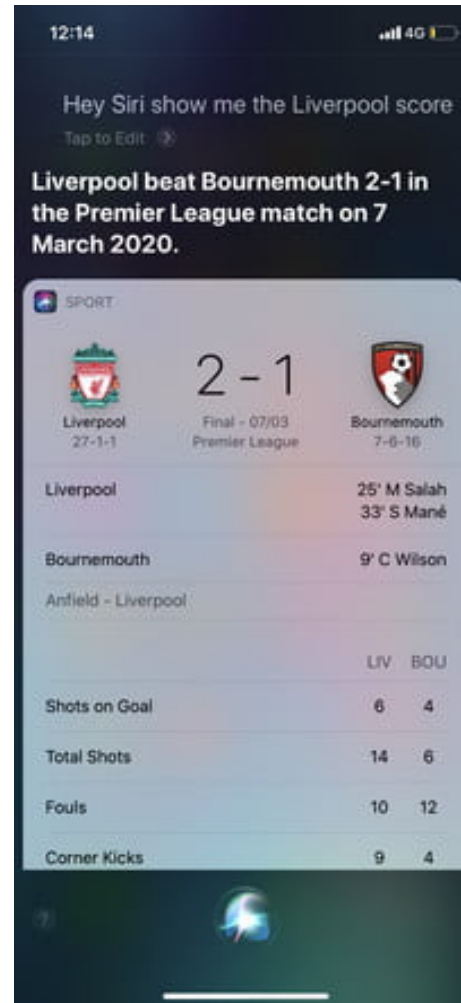
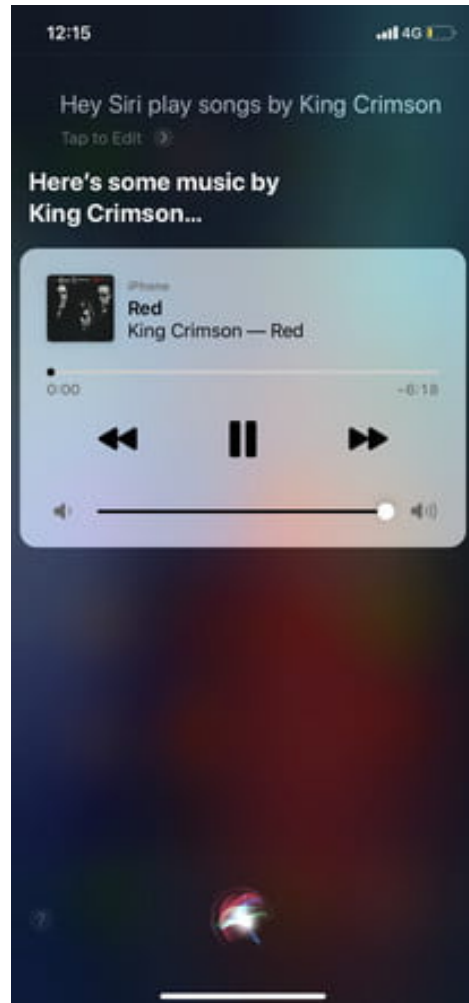


# Alternative Interpretations

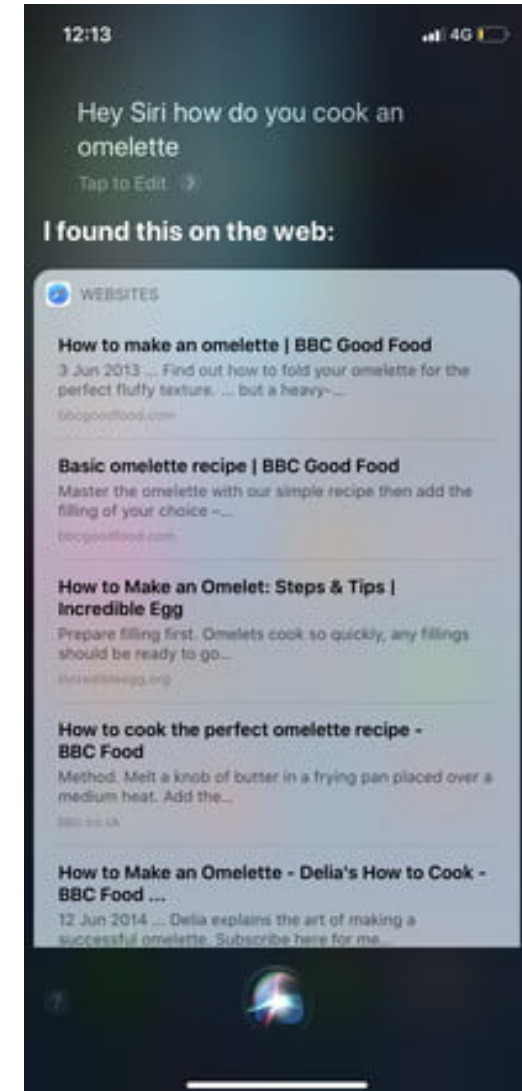
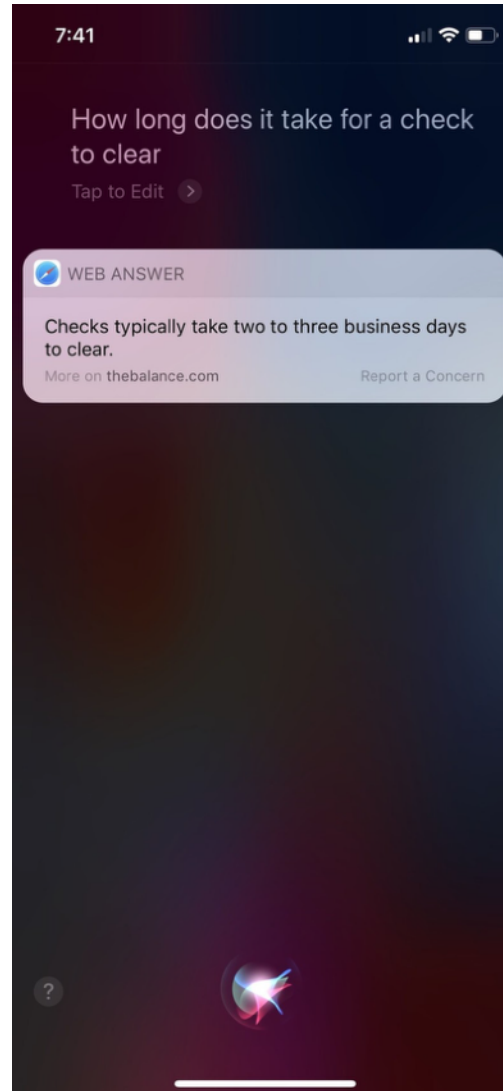
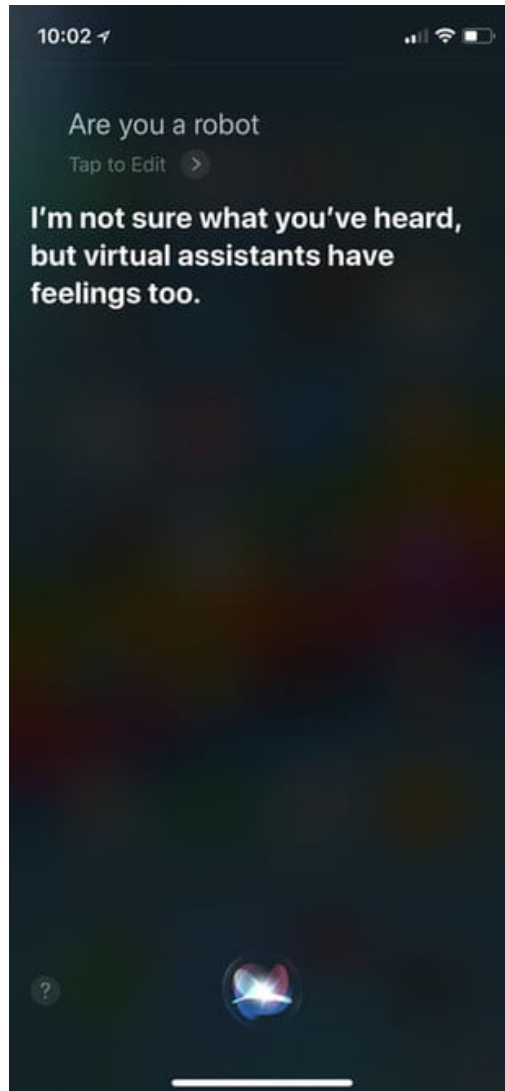




# Siri



# Question Answering





# Question Answering

- QA from text collections
  - Machine Reading

# Question Answering

- QA from text collections
  - Machine Reading
- QA from knowledge graphs
  - KGQA

# Machine Reading

1. Find relevant web page

# Machine Reading

1. Find relevant web page
2. Extract answer span from this page

# Machine Reading

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

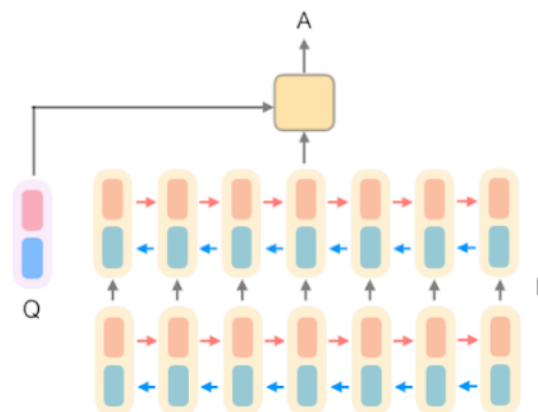


**Document  
Retriever**



**Document  
Reader**

833,500



# Demos

<https://huggingface.co/rag/>

<https://yjernite.github.io/lfqa.html>

# KGQA

1. Find relevant subgraph

# KGQA

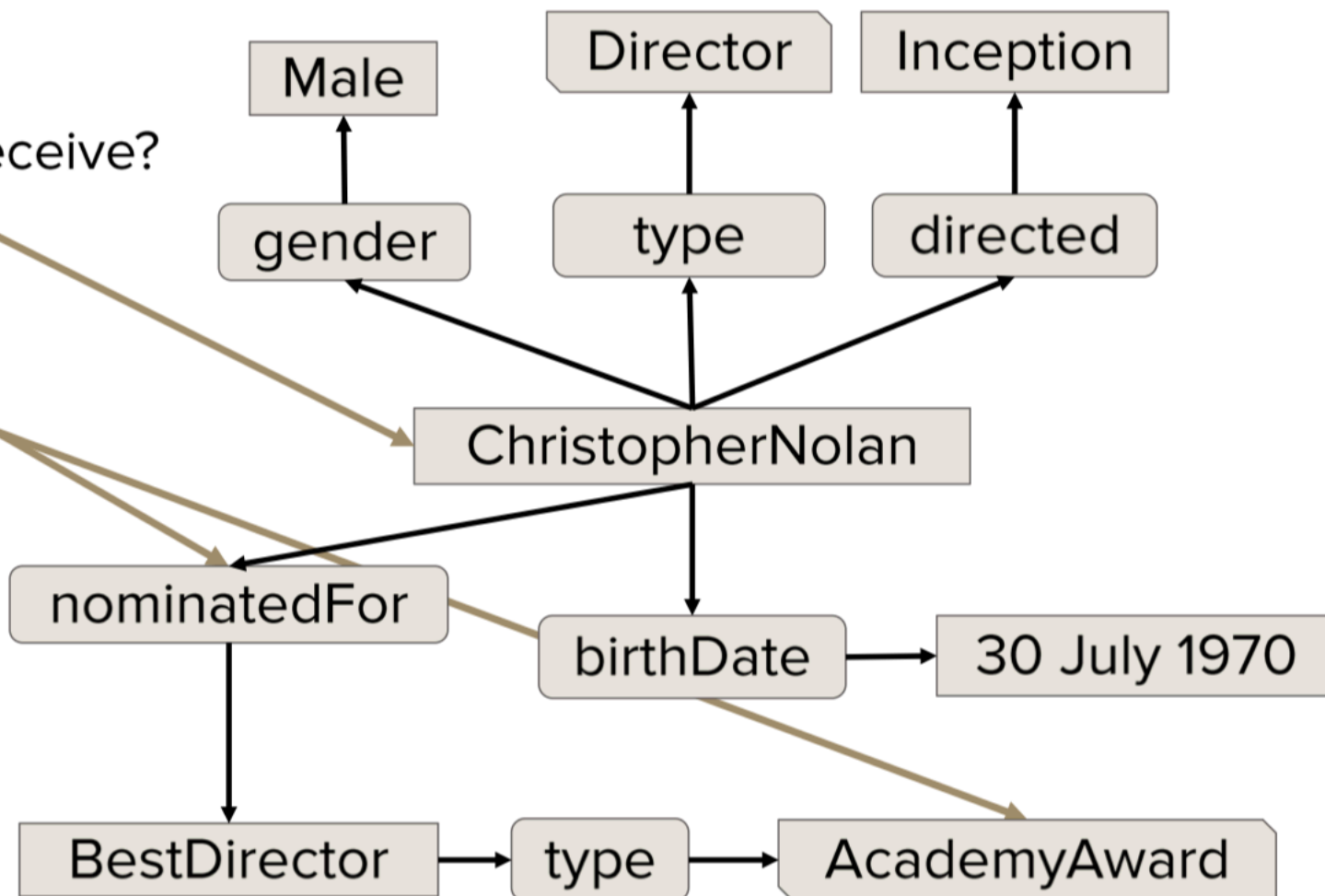
1. Find relevant subgraph
2. Extract answer from this subgraph



# KGQA

Which Oscar nominations did Nolan receive?


<ChristopherNolan, gender, Male>  
<ChristopherNolan, type, Director>  
<ChristopherNolan, directed, Inception>  
<ChristopherNolan, nominatedFor, BestDirector>  
<BestDirector, type, AcademyAward>  
<ChristopherNolan, birthDate, 30 July 1970>




# Conversational QA

**Section:**  **Daffy Duck, Origin & History**



**STUDENT: What is the origin of Daffy Duck?**

**TEACHER:**  first appeared in Porky's Duck Hunt


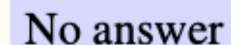
**STUDENT: What was he like in that episode?**

**TEACHER:**  assertive, unrestrained, combative



**STUDENT: Was he the star?**

**TEACHER:**   No, barely more than an unnamed  
bit player in this short

**STUDENT: Who was the star?**

**TEACHER:**   No answer

**STUDENT: Did he change a lot from that first episode in future episodes?**

**TEACHER:**   Yes, the only aspects of the character that have remained consistent (...) are his voice characterization by Mel Blanc

# Question Rewriting

# Question Rewriting

- Where is Xi'an?

- Shaanxi, China

- What is its GDP?

- 95 Billion USD

- What is the share in the province GDP?

- 41.8%

# Question Rewriting

- Where is Xi'an?

- Shaanxi, China

- What is its GDP? - What is Xi'an's GDP?

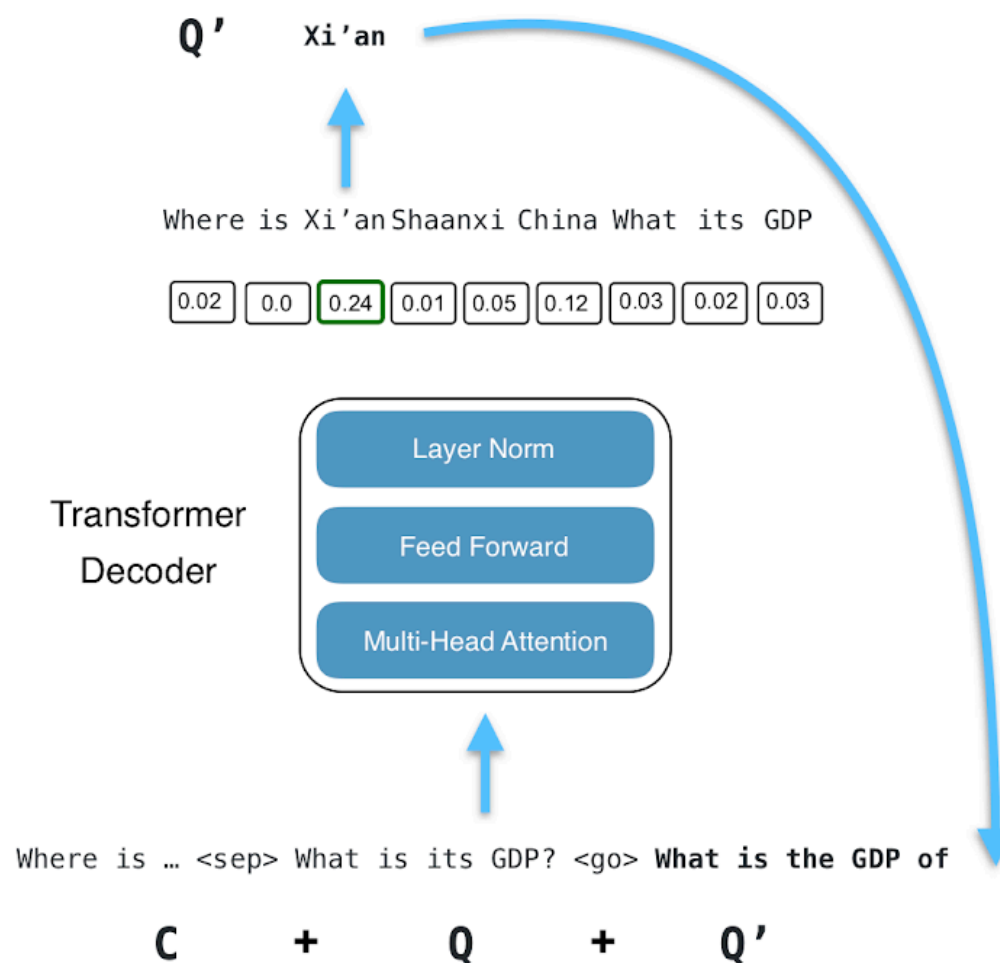
- 95 Billion USD

- What is the share in the province GDP?

- What is the share of Xi'an in the Shaanxi province GDP?

- 41.8%

# Question Rewriting



# Questions?

# Recommender Systems

NETFLIX Watch Instantly ▾ Just for Kids ▾ Instant Queue Personalize DVDs

Movies, TV shows, actors, directors, genres 🔍

lan ▾

### Recently Watched

ORANGE IS THE NEW BLACK  
A NETFLIX ORIGINAL  
FROM THE CREATOR OF WEEDS

### Top 10 for Ian

Disney HERCULES  
Super WHY!  
With the timeless characters of Richard Scarry  
Busytown Mysteries  
THE BEST OUTSIDE FUN EVER!  
Curious George 2  
Leap Frog  
Scout & Friends  
Adventures in Shapeville Park  
Leap Frog  
Scout & Friends  
Number Land  
CONTINUUM  
little tikes. LAND  
Leap Frog  
Scout & Friends  
PHONIC FARM

### Popular on Netflix

MARVEL'S AVENGERS  
THE VERY HUNGRY CATERPILLAR  
New Girl  
the KILLING  
WHO KILLED ROSIE LARSEN?  
THE WEST WING  
123 SESAME STREET  
LIFE HACK TED TALKS  
PROHIBITION  
Disney Pixar TOON MATR'S TALL TALES  
Dr. Seuss The LORAX


### Romantic Independent Comedies

Your taste preferences created this row.

friends  
STRAIGHT A's  
THE GIANT  
THE PILL  
New York, I Love You  
Ira & Abby Cadabby  
When will you find the one?



# Recommender Systems

[Join Prime](#)


[Ashley's Amazon.com](#) | [Today's Deals](#) | [Gift Cards](#) | [Sell](#) | [Help](#)

Shop by Department ▾

Search

Hello, **Ashley**  
Your Account ▾

Join Prime ▾

 Cart ▾

Wish List ▾

[Your Amazon.com](#) | [Your Browsing History](#) | [Recommended For You](#) | [Amazon Betterizer](#) | [Improve Your Recommendations](#) | [Your Profile](#) | [Learn More](#)

## Your Amazon.com

- Featured Recommendations
- Books
- Health & Personal Care
- Beauty
- Industrial & Scientific
- Electronics
- Home & Kitchen
- See All Recommendations

### Books

Page 1 of 12





**New Release**  
**The Silver Star**  
Jeannette Walls  
Hardcover  
~~\$26.00~~ **\$15.76**  
[Why recommended?](#)



**New Release**  
**Inferno**  
Dan Brown  
Hardcover  
~~\$29.95~~ **\$17.15**  
[Why recommended?](#)



**The Woman Upstairs**  
Claire Messud  
★★★★☆ (113)  
Hardcover  
~~\$25.95~~ **\$14.99**  
[Why recommended?](#)



**New Release**  
**The Burgess Boys**  
Elizabeth Strout  
Hardcover  
~~\$26.00~~ **\$15.51**  
[Why recommended?](#)



**A Delicate Truth**  
John le Carré  
★★★★☆ (87)  
Hardcover  
~~\$28.95~~ **\$16.25**  
[Why recommended?](#)



**New Release**  
**Paris**  
Edward Rutherfurd  
Hardcover  
~~\$32.50~~ **\$19.04**  
[Why recommended?](#)



**New Release**  
**Maya's Notebook**  
Isabel Allende  
Hardcover  
~~\$27.99~~ **\$19.47**  
[Why recommended?](#)



► See all recommendations in Books

### Health & Personal Care

Page 1 of 12

# Recommender Systems

Google news

Search News

Search the Web

[Advanced news search](#)

## Top Stories

Kobe Bryant  
Sathya Sai Baba  
New York Knicks  
Philadelphia  
76ers  
Tyler Perry  
Muammar  
al-Gaddafi  
Zaza Pachulia  
Prince William of  
Wales  
Taliban  
Tsunami

Starred ★

England, UK

World

U.S.

Business

Politics


## Top Stories

[Sathya Sai Baba »](#)

### [Devotees pay homage to Sai](#)

Times of India - 1 hour ago

Devotees scatter rose petals during a memorial for Hindu guru Sathya Sai Baba at the Sri Satya Sai Baba temple in Amritsar on April 24, 2011.

[+ Video: On birthday, Sachin grieves for Sai Baba](#)  NDTV.com

Thousands of devotees rush to pay homage to Sathya Sai [Hindustan Times](#)

[Indian Express](#) - [The Hindu](#) - [Xinhua](#) - [Daily News & Analysis](#) - [Wikipedia: Sathya Sai Baba](#)


[all 1865 news articles »](#)

[Taliban »](#)

### [Hundreds escape from Afghan prison, NATO says](#)

Los Angeles Times - [Molly Hennessy-Fiske](#) - 19 minutes ago

The prisoners in the mass jailbreak in the southern city of Kandahar fled through a tunnel they had dug, a Taliban spokesman says.

[+ Video: Hundreds of Taliban in tunnel jail escape](#)  [Al Jazeera](#)

[Hundreds of Taliban in tunnel jail escape](#) [Aljazeera.net](#)

[The Hindu](#) - [CNN International](#) - [New York Times](#) - [9&10 News](#)

[all 354 news articles »](#)

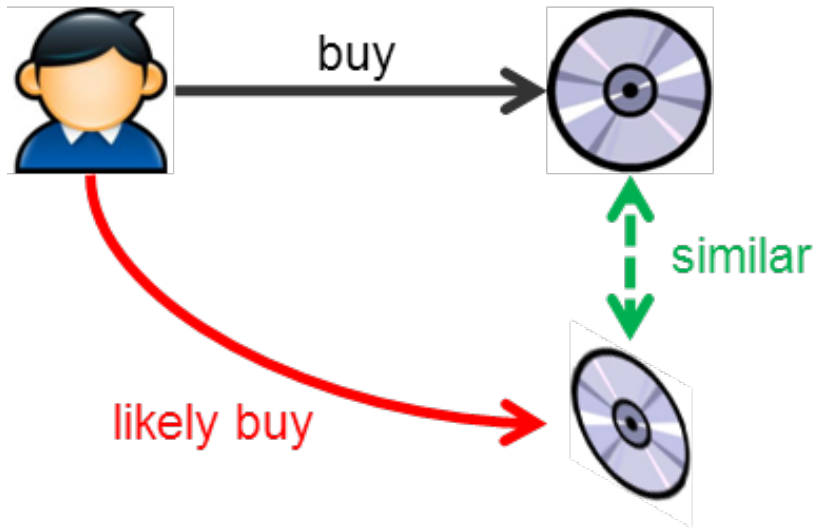
[Arizona Congresswoman Gabrielle Giffords »](#)

### [Giffords cleared to attend shuttle launch: husband](#)

Reuters - [Joe Skipper](#), [Steve Gorman](#) - 1 hour ago

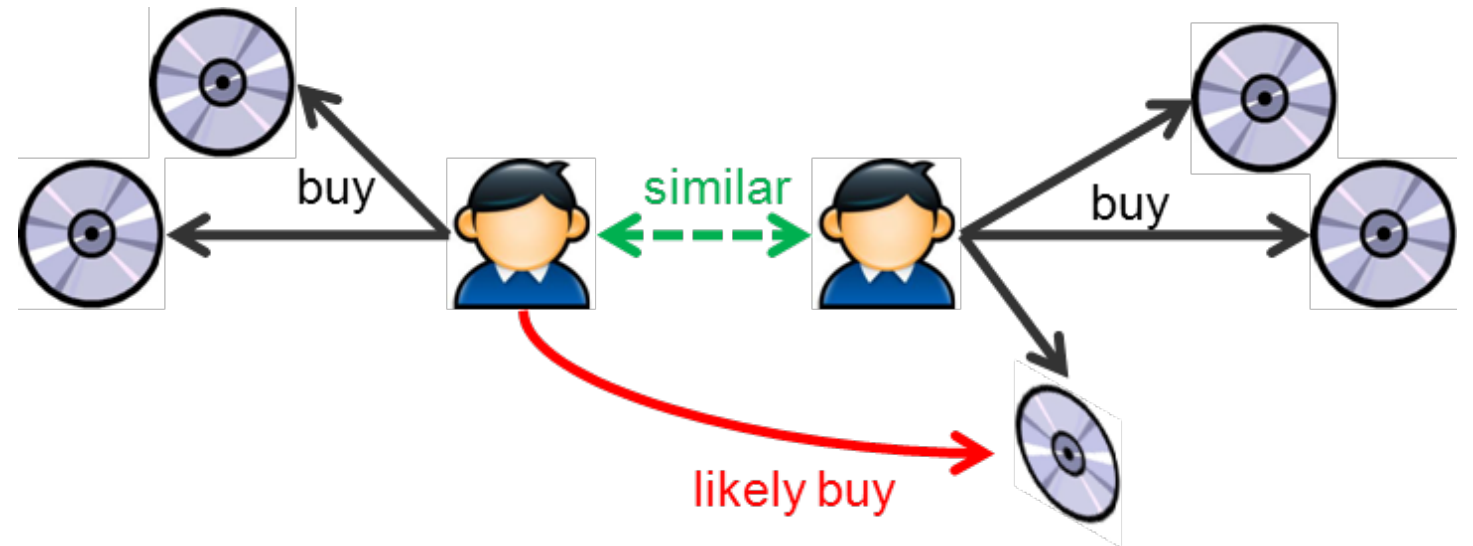
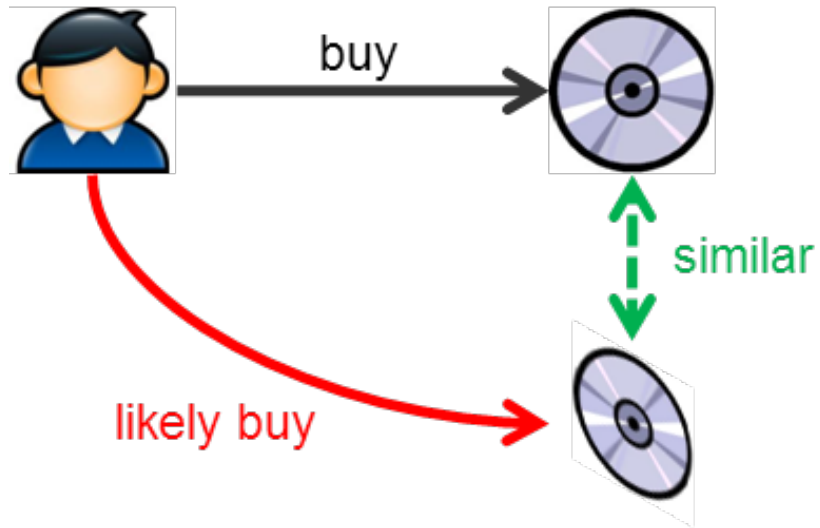
Space shuttle Endeavour sits atop launch pad 39A after arriving from the Vehicle Assembly Building to prepare for Mission STS-134 at the Kennedy Space Center.

# Recommender Systems



Images: <https://dzone.com/articles/recommendation-engine-models>

# Recommender Systems



Images: <https://dzone.com/articles/recommendation-engine-models>

# Collaborative Filtering



Harry Potter



The Triplets of  
Belleville



Shrek



The Dark  
Knight Rises



Memento

✓		✓	✓	
	✓			✓
✓	✓	✓		
			✓	✓

# Matrix Factorization

$$A \approx UV^T$$



Harry Potter



The Triplets of  
Belleville



Shrek



The Dark  
Knight Rises



Memento

✓		✓	✓	
	✓			✓
✓	✓	✓		
			✓	✓

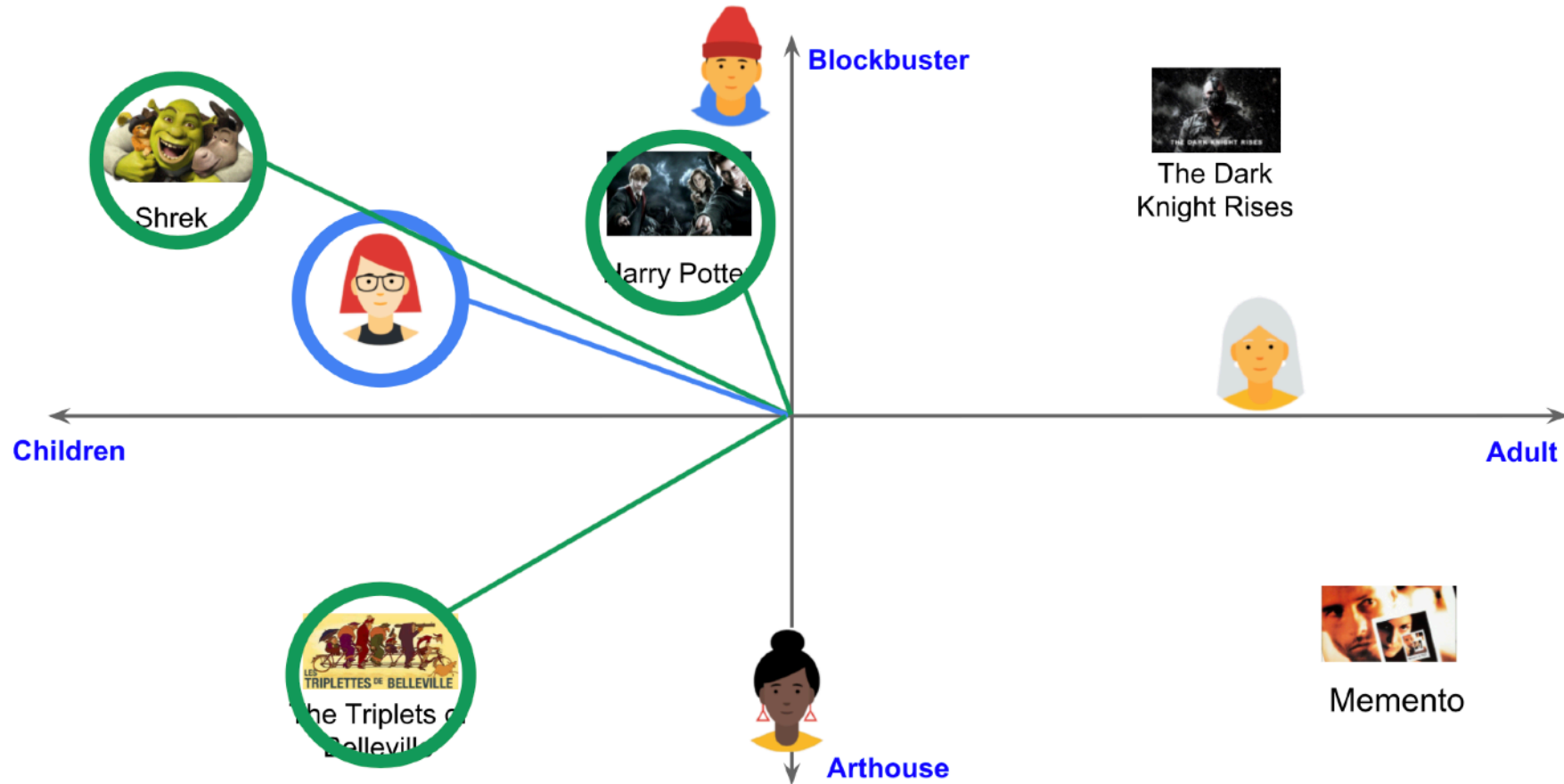
≈

.9	-1	1	1	-.9
-.2	-.8	-1	.9	1

1	.1
-1	0
.2	-1
.1	1


.88	-1.08	0.9	1.09	-0.8
-0.9	1.0	-1.0	-1.0	0.9
0.38	0.6	1.2	-0.7	-1.18
-0.11	-0.9	-0.9	1.0	0.91


# $k$ -NN search








# Conversational Recommendation


 Can you find me a **mobile phone** on Amazon?


 Sure, what **operating system** do you prefer?


 I want an **Android** one.


 OK, and any preference on **screen size**?


 Better larger than **5 inches**.


 Do you have requirements on **storage capacity**?


 I want it to be at least **64 Gigabytes**.


 And any preference on **phone color**?


 **Not particularly**.


 Sure, then what about the following choices?





 I don't like them very much...


 OK, do you have any preference on the **brand**?


 Better be **Samsung or Huawei**.


 Any requirement on **price**?

 Should be **within 700 dollars**.

 OK, then what about these ones?

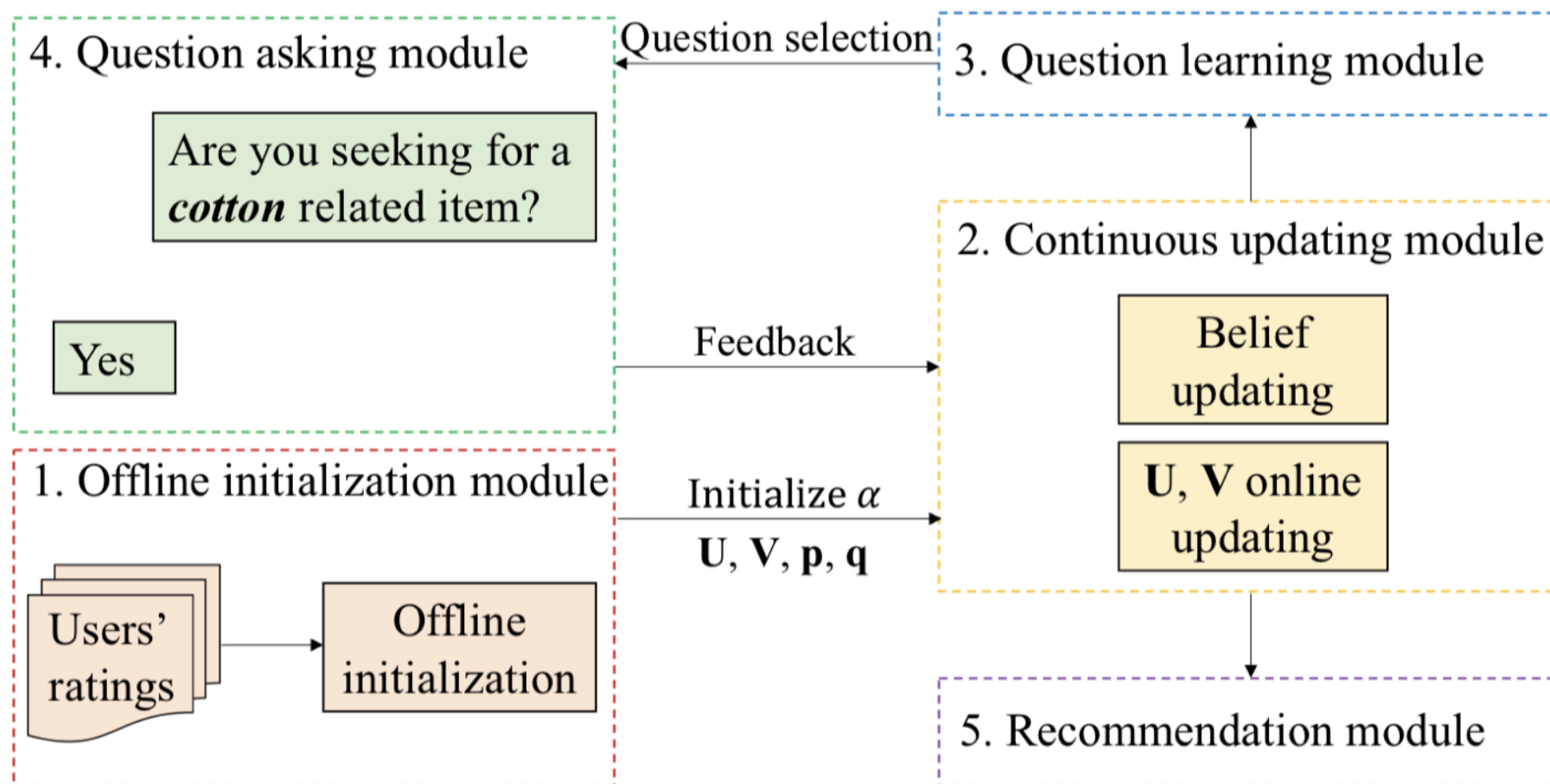


 Great, I want the first one, can you order it for me?

 Sure, I have placed the order for you, enjoy!



# Conversational Recommendation



Conversation?

# Knowledge Grounded Conversations

# Knowledge Grounded Conversations

- External knowledge == Information source

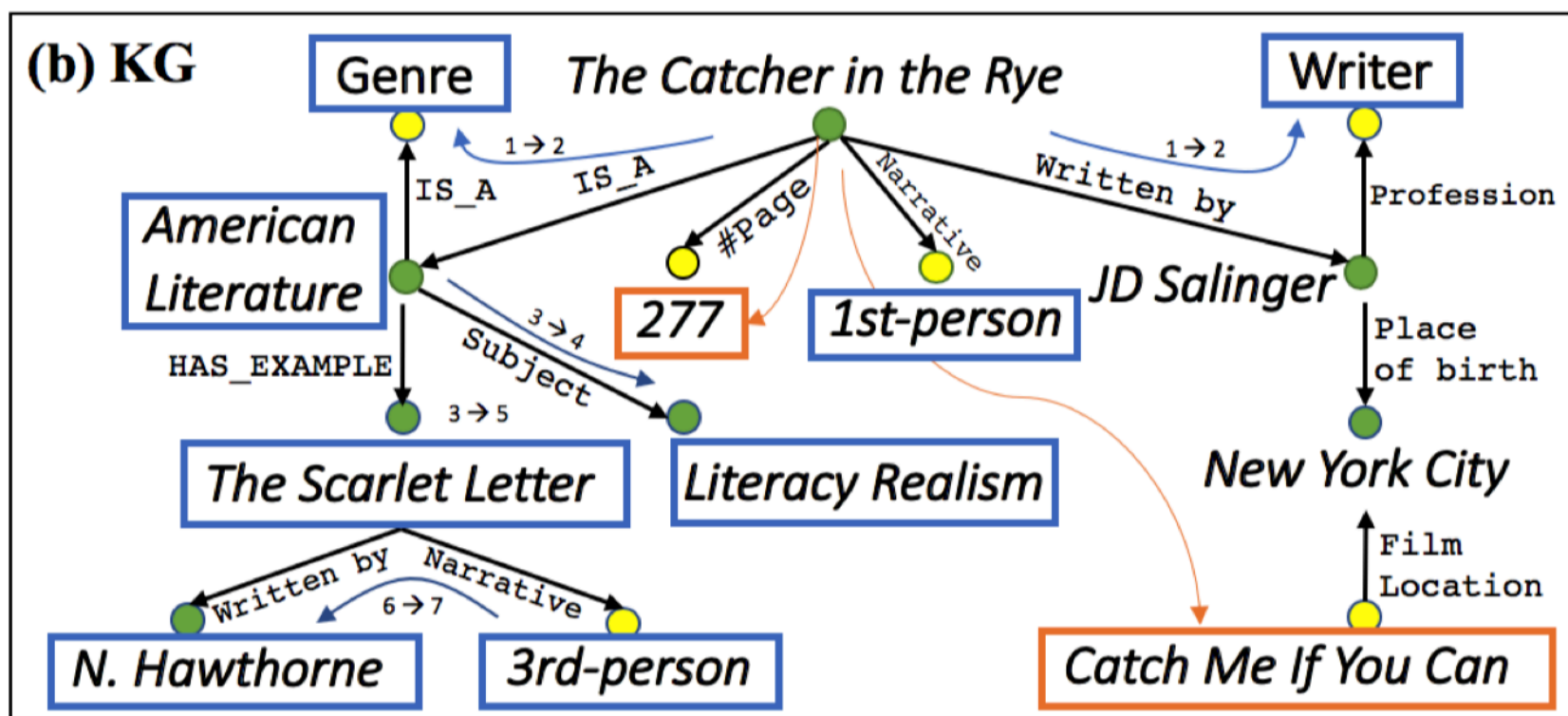
# Knowledge Grounded Conversations

- External knowledge == Information source
  - Text collection, e.g., Wikipedia

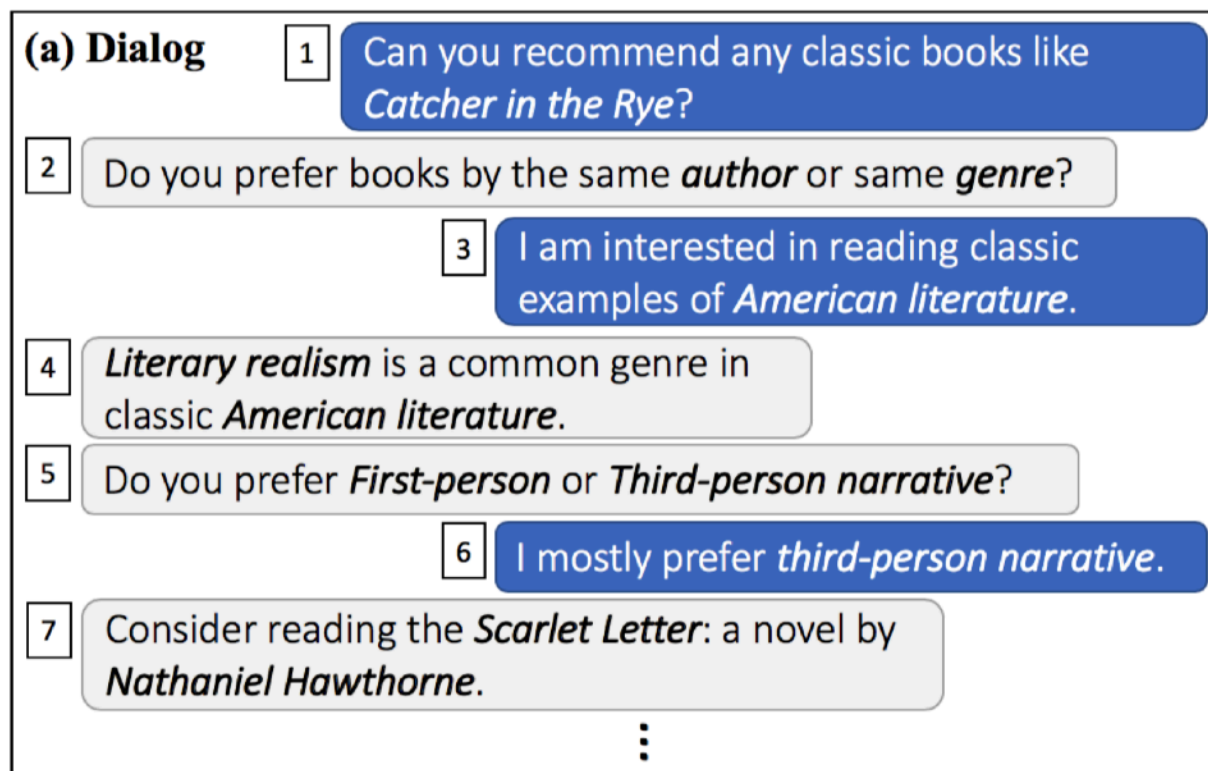
# Knowledge Grounded Conversations

- External knowledge == Information source
  - Text collection, e.g., Wikipedia
  - Knowledge Graphs, e.g., DBpedia or WikiData

# Knowledge Grounded Conversations



# Knowledge Grounded Conversations





# Knowledge Grounded Conversations

```
{
  "message": "Do you like Iron Man",
},
{
  "path": [
    "Iron Man",
    "starred_actors",
    "Robert Downey Jr."
  ]
},
{
  "message": "Sure do! Robert Downey Jr. is a favorite."
},
{
  "path": [
    "Robert Downey Jr.",
    "~starred_actors",
    "Zodiac (Crime Fiction Film)"
  ]
}
```

# Knowledge Grounded Conversations

```
{
  "message": "Do you like Iron Man",
},
{
  "path": [
    "Iron Man",
    "starred_actors",
    "Robert Downey Jr."
  ]
},
{
  "message": "Sure do! Robert Downey Jr. is a favorite."
},
{
  "path": [
    "Robert Downey Jr.",
    "~starred_actors",
    "Zodiac (Crime Fiction Film)"
  ]
}
```

## 1. Entity Linking



# Knowledge Grounded Conversations

```
{
  "message": "Do you like Iron Man",
},
{
  "path": [
    "Iron Man",
    "starred_actors",
    "Robert Downey Jr."
  ]
},
{
  "message": "Sure do! Robert Downey Jr. is a favorite."
},
{
  "path": [
    "Robert Downey Jr.",
    "~starred_actors",
    "Zodiac (Crime Fiction Film)"
  ]
}
```

1. Entity Linking  
2. Entity Selection

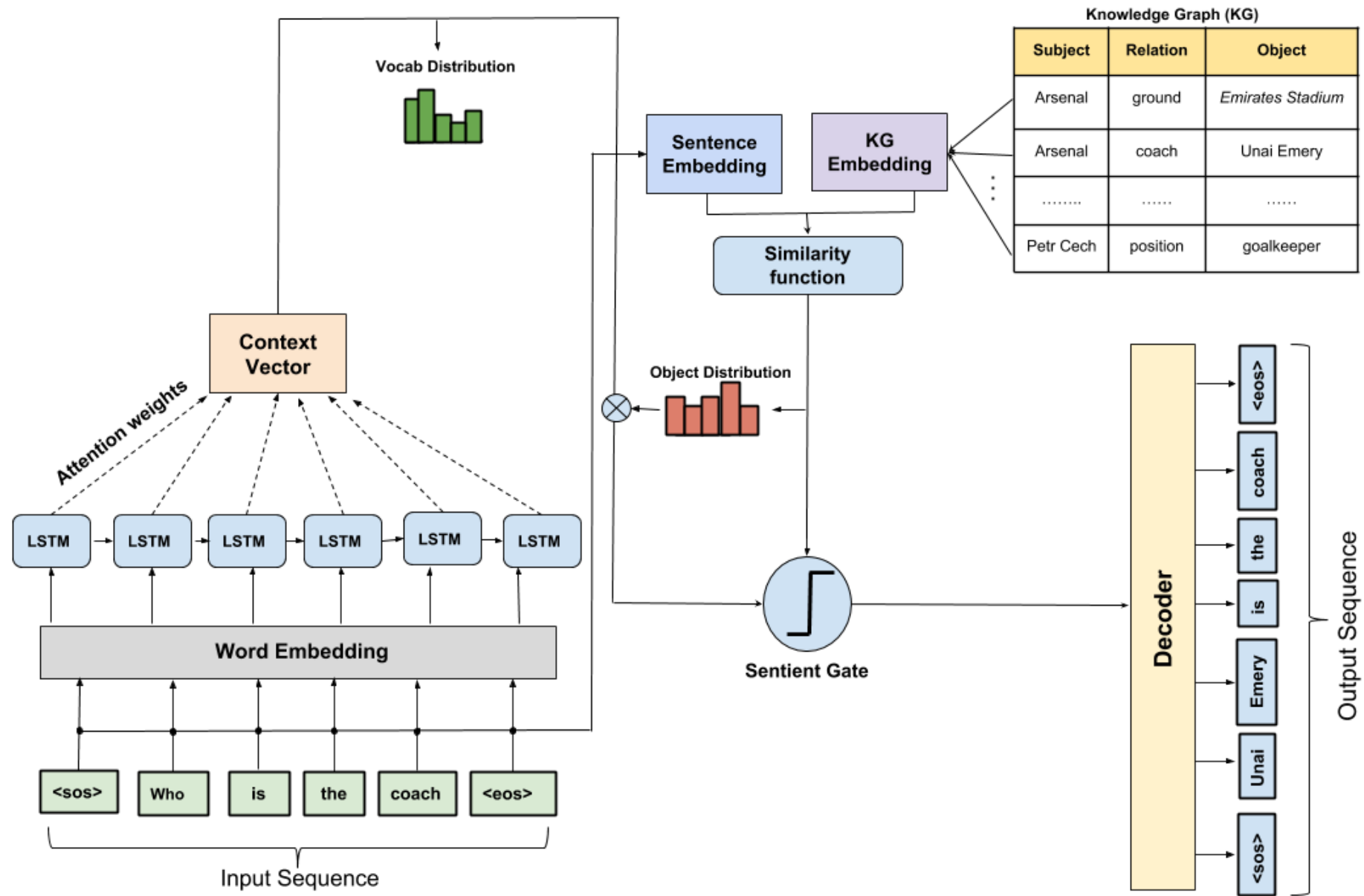


# Knowledge Grounded Conversations

```
{
  "message": "Do you like Iron Man",
},
{
  "path": [
    "Iron Man",
    "starred_actors",
    "Robert Downey Jr."
  ]
},
{
  "message": "Sure do! Robert Downey Jr. is a favorite."
},
{
  "path": [
    "Robert Downey Jr.",
    "~starred_actors",
    "Zodiac (Crime Fiction Film)"
  ]
}
```

1. Entity Linking
2. Entity Selection
3. Response Generation





# Questions?

# Questions

- How do we know that our search system is intelligent?
  - i.e. How should we evaluate it?
- What should we measure and how?
  - i.e. What is the **evaluation metric** we should use?

# Evaluation

- Response relevance
- Correctness
- Completeness
- Mixed Initiative



# Evaluation

- Response relevance

# Evaluation

- Response relevance
  - BLEU
  - METEOR
  - ROUGE
  - cosine

# Evaluation

	Twitter				Ubuntu			
Metric	<b>Spearman</b>	p-value	<b>Pearson</b>	p-value	<b>Spearman</b>	p-value	<b>Pearson</b>	p-value
Greedy	0.2119	0.034	0.1994	0.047	0.05276	0.6	0.02049	0.84
Average	0.2259	0.024	0.1971	0.049	-0.1387	0.17	-0.1631	0.10
Extrema	0.2103	0.036	0.1842	0.067	0.09243	0.36	-0.002903	0.98
METEOR	0.1887	0.06	0.1927	0.055	0.06314	0.53	0.1419	0.16
BLEU-1	0.1665	0.098	0.1288	0.2	-0.02552	0.8	0.01929	0.85
BLEU-2	0.3576	< 0.01	0.3874	< 0.01	0.03819	0.71	0.0586	0.56
BLEU-3	0.3423	< 0.01	0.1443	0.15	0.0878	0.38	0.1116	0.27
BLEU-4	0.3417	< 0.01	0.1392	0.17	0.1218	0.23	0.1132	0.26
ROUGE	0.1235	0.22	0.09714	0.34	0.05405	0.5933	0.06401	0.53
Human	0.9476	< 0.01	1.0	0.0	0.9550	< 0.01	1.0	0.0

Liu, et al. How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation. EMNLP 2016:

# Evaluation

- Response relevance
  - Automated
  - Human

# Evaluation

- Correctness

# Evaluation

- Correctness

Q: Who was president of the United States in 1700?

A: William Penn was president of the United States in 1700.

Q: Who was president of the United States in 1650?

A: Richard Bennett was president of the United States in 1650.

Q: Who was president of the United States in 1620?

A: James I was president of the United States in 1620.

Q: Who was president of the United States in 1600?

A: Queen Elizabeth I was president of the United States in 1600.

# Evaluation

- Correctness
  - Machine Reading
    - answer text span overlap: F-measure

# Evaluation

- Correctness
  - Machine Reading
    - answer text span overlap: F-measure
  - Retrieval
    - page/passage ranking: Precision, nDCG



# Evaluation

- Correctness
  - Machine Reading
    - answer text span overlap: F-measure
  - Retrieval
    - page/passage ranking: Precision, nDCG
- KGQA
  - answer set overlap: Precision, nDCG, Recall, F-measure

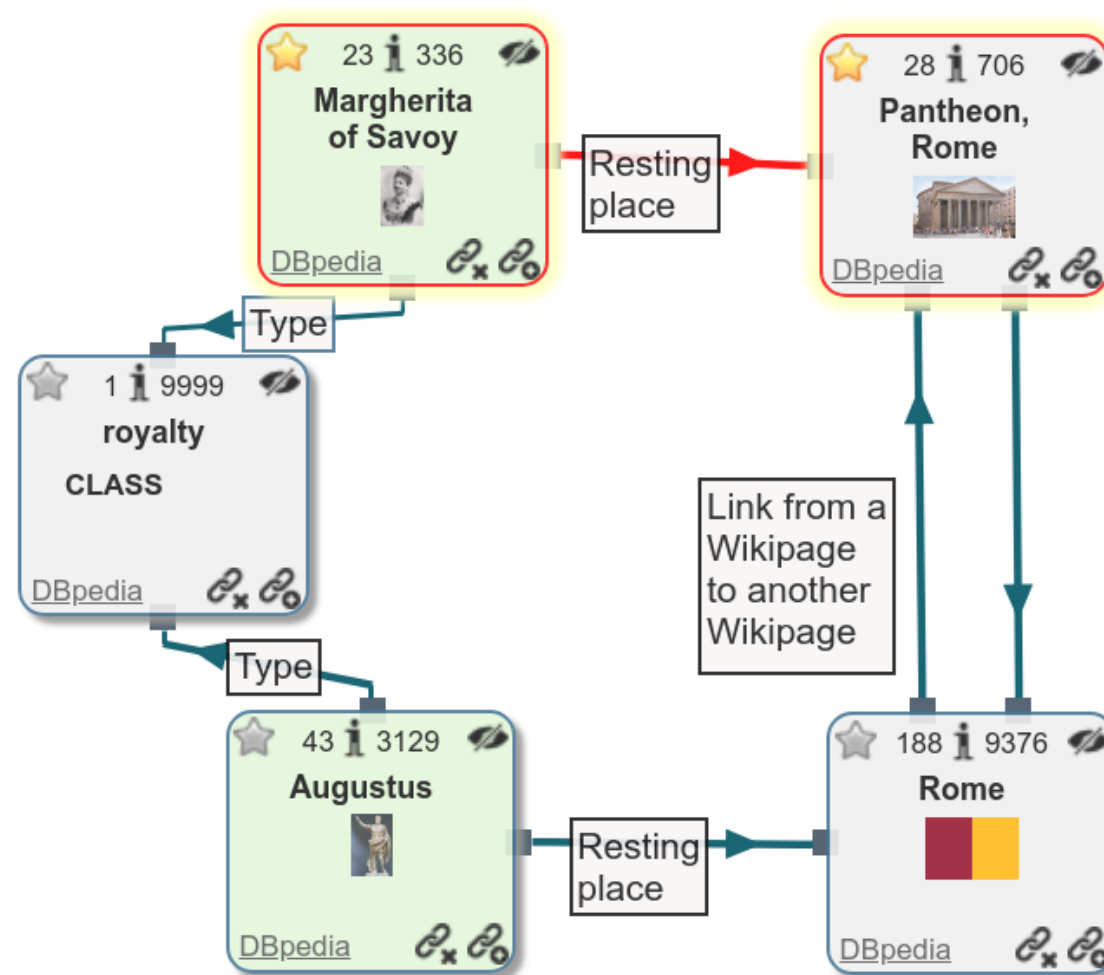
# Evaluation

- Completeness

# Evaluation

- Completeness

Q: Give me a count of royalties buried in Rome?



# Evaluation

- Completeness
  - pooling
  - relevance assessment
  - TREC NIST <https://trec.nist.gov>

# Evaluation

- Response relevance
- Correctness
- Completeness
- Mixed Initiative

# Evaluation

- Mixed Initiative

# Evaluation

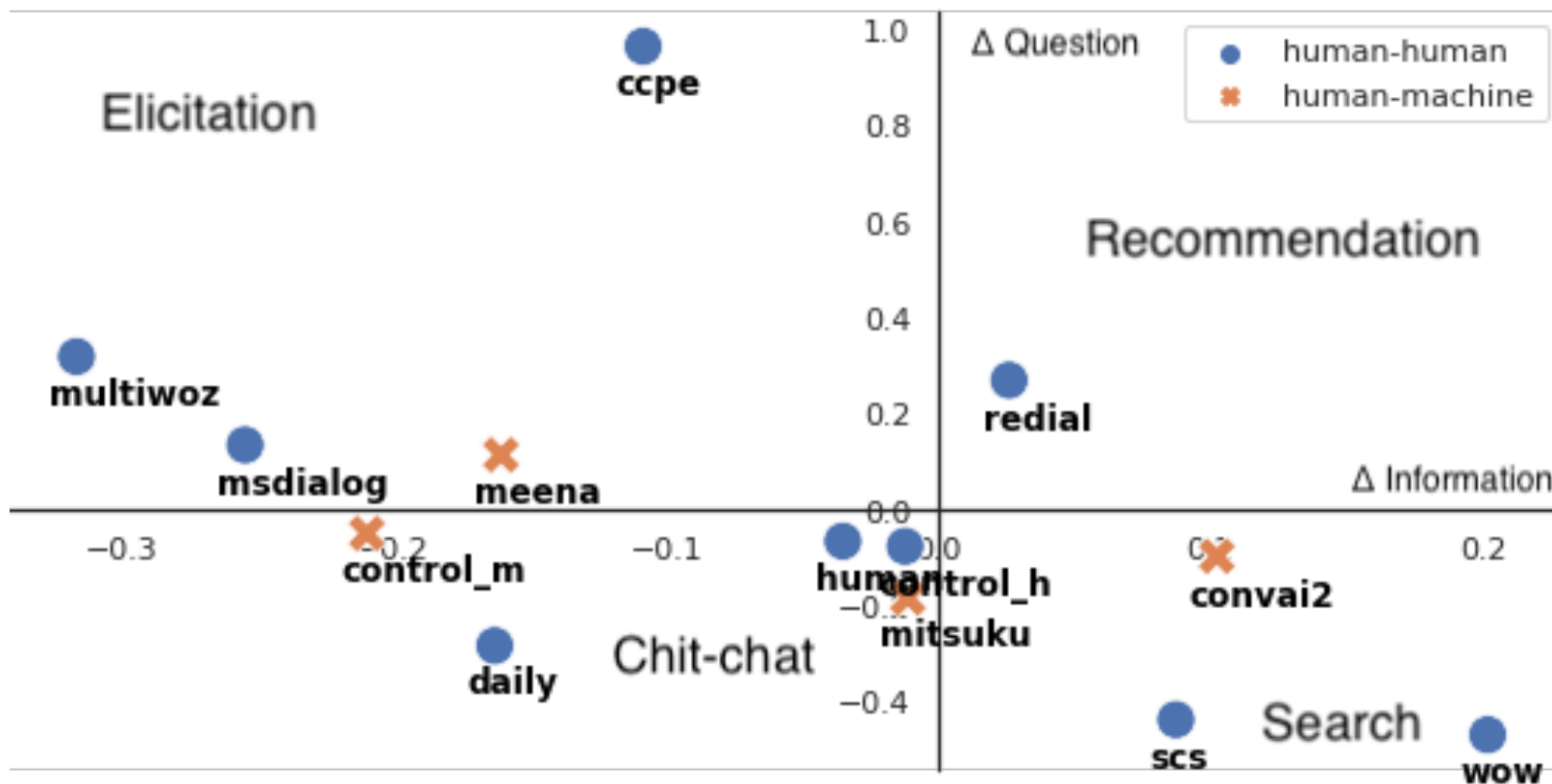
- Mixed Initiative
  - Question
    - Asking questions
    - Answering questions

# Evaluation

- Mixed Initiative
  - Question
    - Asking questions
    - Answering questions
  - Topic (Information)
    - Introduction
    - Follow-up



# Mixed Initiative



# Questions?

# Information Retrieval

- SIGIR
- ECIR
- ICTIR
- WSDM
- CIKM
- CHIIR
- TOIS

# ILPS

information and language processing systems

[Home](#) [People](#) [Publications](#) [Research](#) [Teaching](#) [Events](#) [About](#)

Research at ILPS is focused on technology that connects people to information: search engines, recommender systems, and conversational agents. We are organized in four labs.



## Information Retrieval Lab



## Language Technology Lab



## Human Machine Intelligence Lab



### Recent Posts

[A New Year, A New Team](#)

[Yang Fang Joins ILPS](#)

[ILPS @ SIGIR 2020](#)

Tweets by [@ILPS\\_Amsterdam](#)



**ILPS Amsterdam**  
[@ILPS\\_Amsterdam](#)



Great to see so many first timers signing up for the online version of SEA! 😊 Want to join as

# Courses

- **IR1** February
  - Asst. Prof. Dr. Ilya Markov
  - + Learning to rank, Evaluation, Entity Search
- **IR2** September
  - Assoc. Prof. Dr. Evangelos Kanoulas
  - team-projects in Conversational Search





# SEA: Search Engines Amsterdam

Amsterdam, Netherlands

1,955 members · Public group

Organized by **Maartje ter Hoeve** and 1 other

Share: [f](#) [t](#) [in](#)

[About](#)
[Events](#)
[Members](#)
[Photos](#)
[Discussions](#)
[More](#)

You're a member

## What we're about

Information drives the planet. We organize talks around implementations of information retrieval, in search engines, in recommender systems, or in...

[Read more](#)

## Organizers



**Maartje ter Hoeve** and 1 other

[Message](#)

# MSc Theses

- Question Rewriting for Conversational QA
- Knowledge Graph Grounded Conversation Generation
- Integrating Language Style Matching Objective into Generative Dialogue Models
- Conversation Mining (KPN)
- Deep matching of resumes to job descriptions (Randstad)

Svitlana Vakulenko  
[s.vakulenko@uva.nl](mailto:s.vakulenko@uva.nl)