TechFest
the & in R&D

Microsoft Research

Conversational Knowledge Graphs

Larry Heck
Microsoft Research

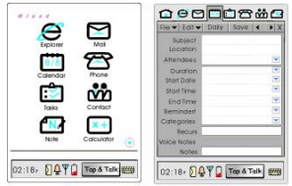
Microsoft

Conversational Systems

Brief History

TV Voice Search
e.g., Bing on Xbox

Multi-modal systems
e.g., Microsoft MiPad, Pocket PC



2014

Task-specific argument extraction
(e.g., Nuance, SpeechWorks)
User: "I want to fly from Boston to New York next week."

Early 2000s

IBM WATSON

Intent Determination
(e.g. Nuance's Emily™, AT&T HMIHY)
User: "Uh…we want to move…we want to change our phone line from this house to another house"

Virtual Personal Assistants:
e.g., Siri, Google now

DARPA
CALO Project

Early 1990s

Keyword Spotting
(e.g., AT&T)
System: "*Please say collect, calling card, person, third number, or operator*"

The Need: *Where are we now?*

Conversational systems crafted for each domain

- Select domain
- Manually construct schema/ontology
- Manually collect and annotate data
- Train models/build grammars

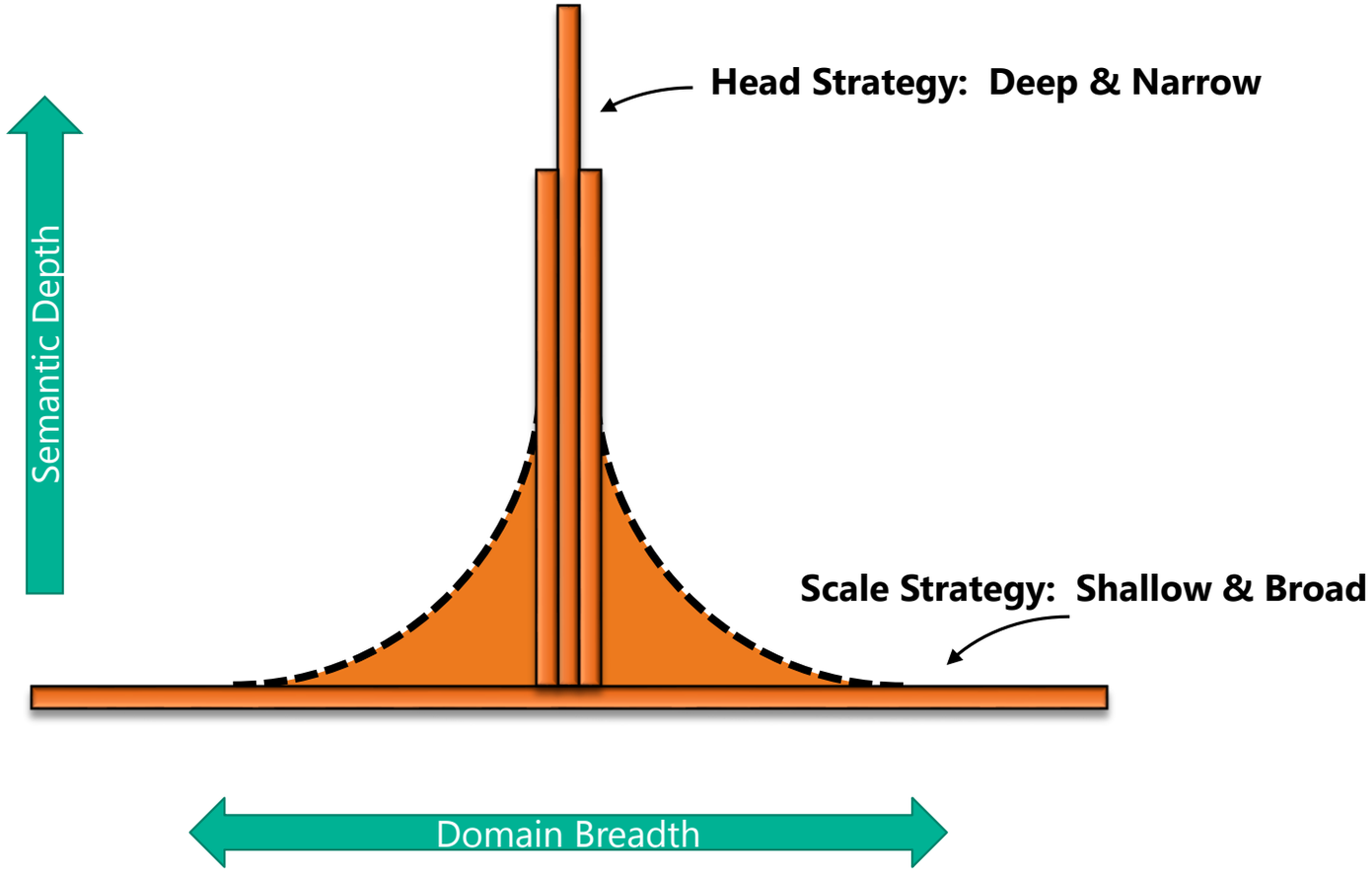
Result

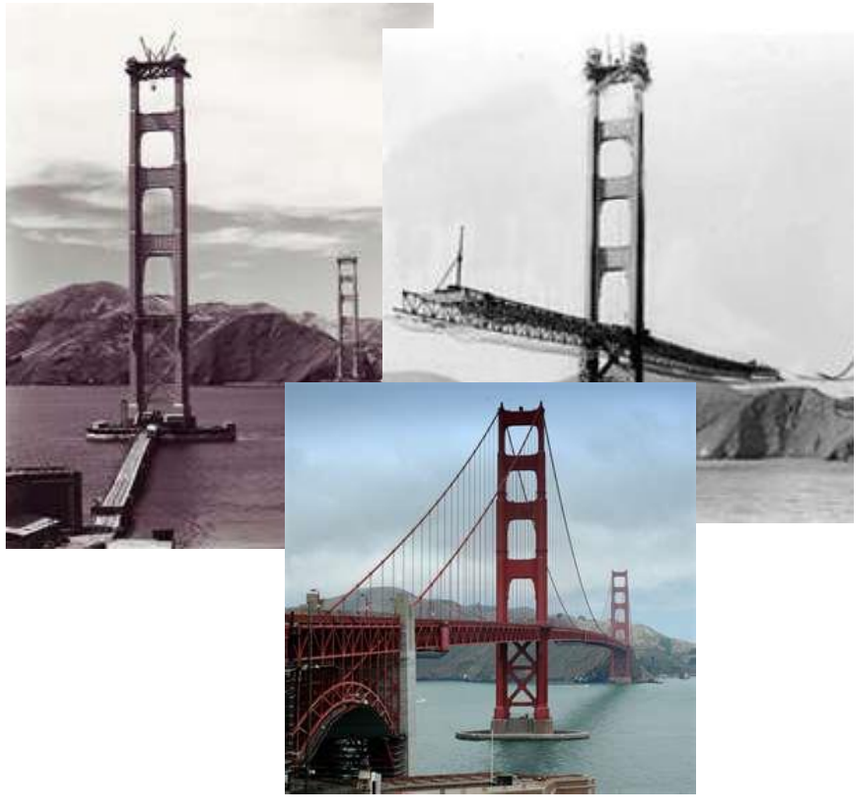
- Narrow breadth of domains
- Limited sharing of data/schemas between domains
- Limited ability to incorporate disparate knowledge sources
- Inflexible to changes in task definition

How will we ever create a "NUI to the world's knowledge"? …

Conversational Systems Challenge

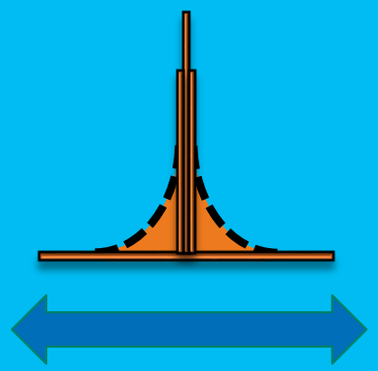
Scaling *Depth and Breadth*





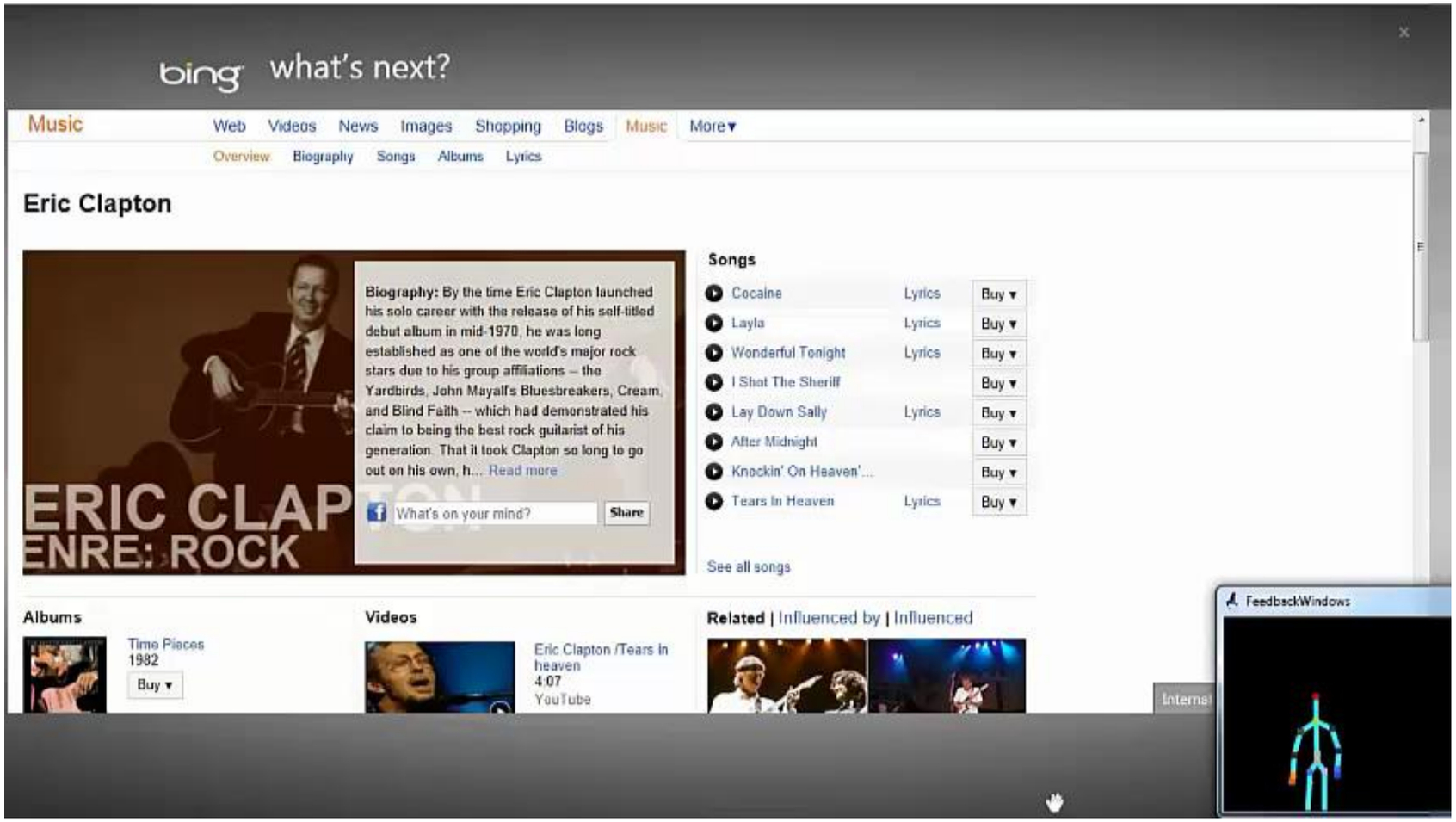
Video

Domain Breadth

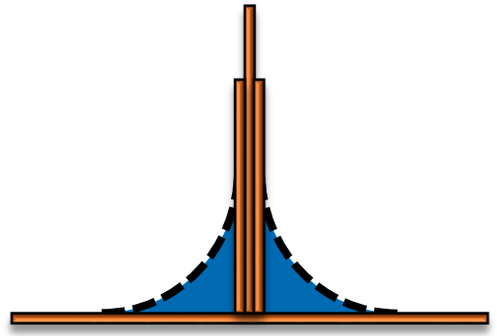


TechFest 2012
Conversational Search and Browse

TechFest 2012
Conversational Search and Browse



The Opportunity



Knowledge is the Foundation of Conversations

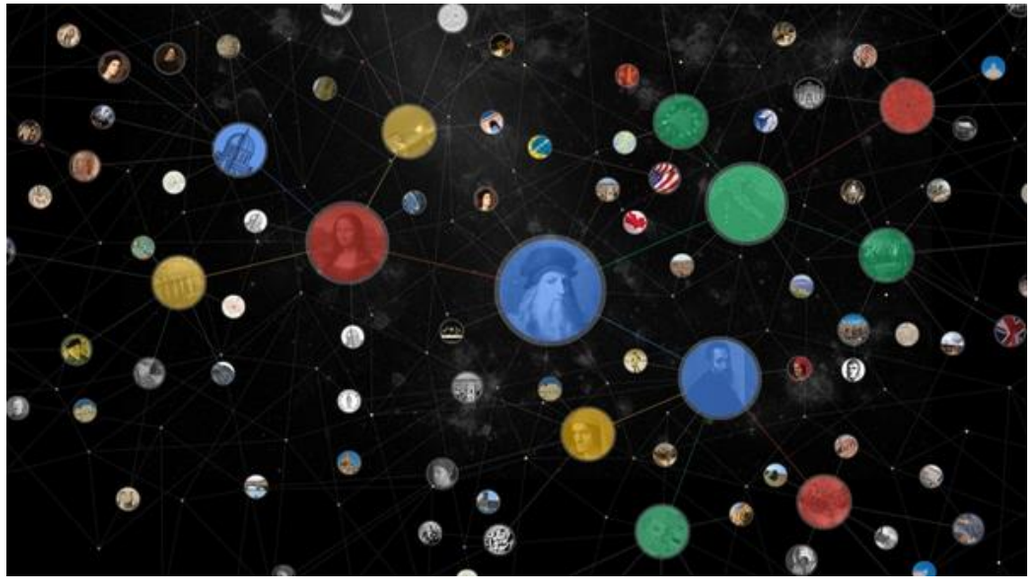
A vast majority of user interactions are with people, locations, things (**entities**).

Knowledge refers to these **entities**/concepts and to how they are interrelated.

The dual-role of knowledge

People seek to **browse** and **find information** about **entities** and to **transact** on them.

Knowledge serves as a **grounding for conversations**.



Semantic Knowledge Graphs

What are knowledge graphs?

Graphs of strongly typed and uniquely identified **entities** (nodes) and facts/literals connected by **relations** (edges)

Examples

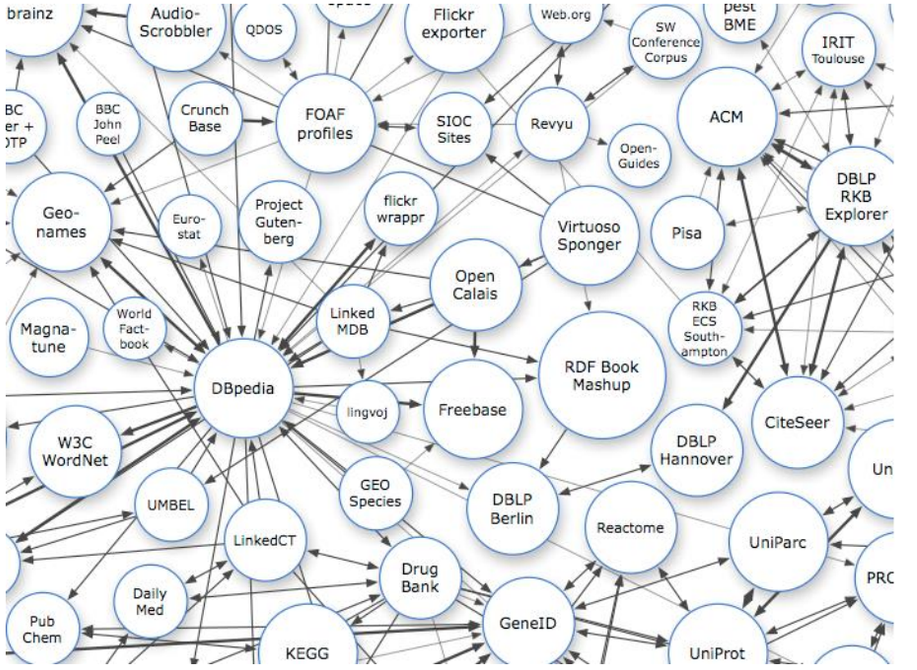
Satori, Google KG, Facebook Open Graph, Freebase

How large?

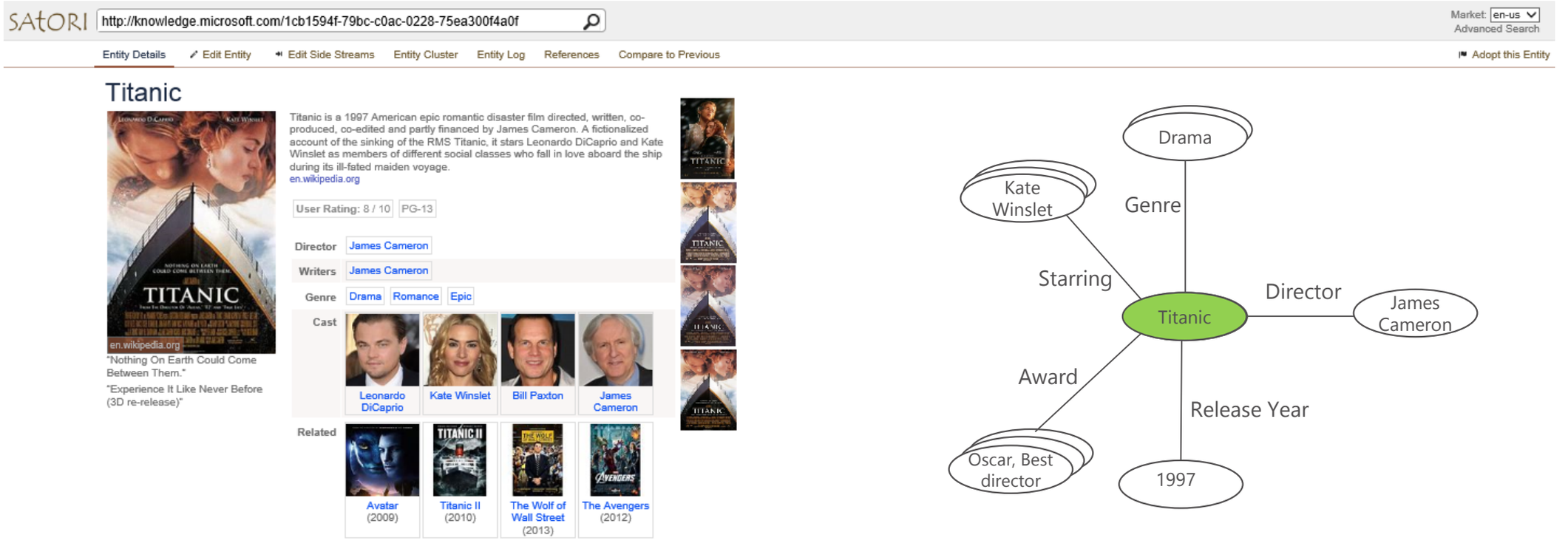
> 500M entities, >1.5B relations, >5B facts

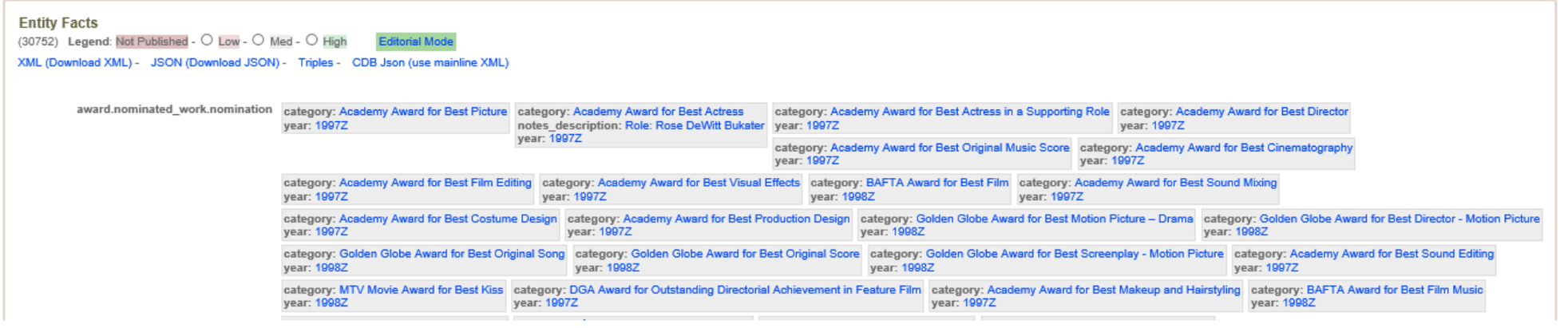
How broad is the knowledge?

Wikipedia-breadth: "American Football" ← → "Zoos"



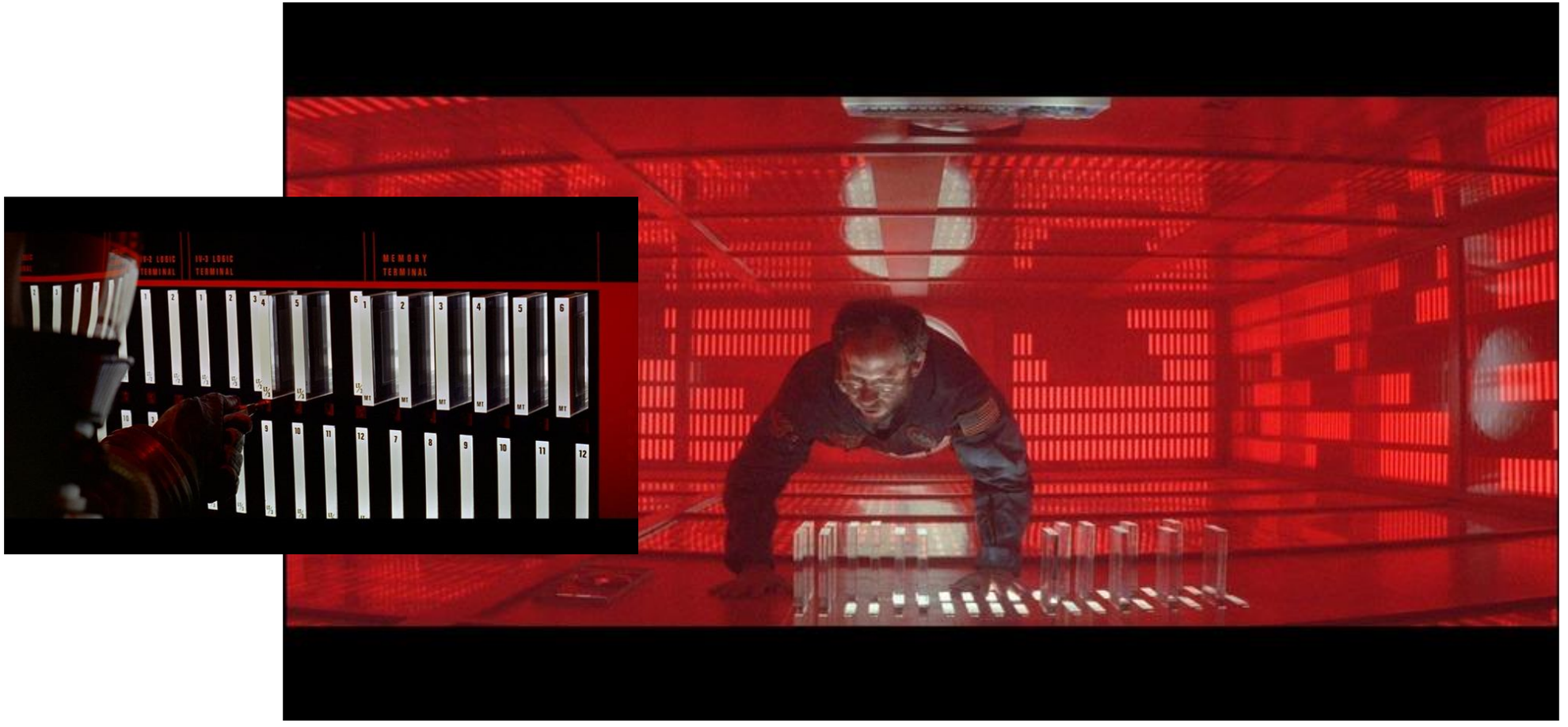
Semantic Knowledge Graphs





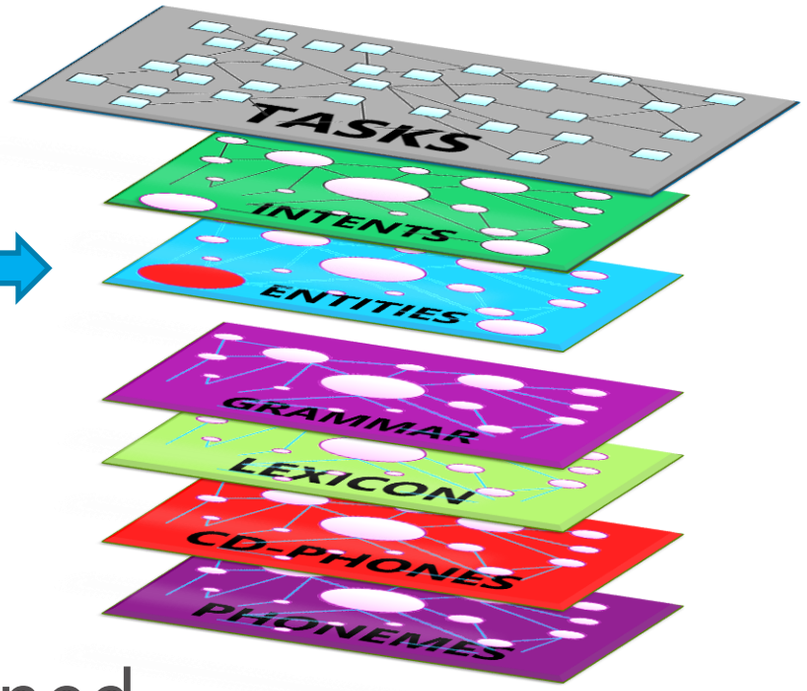
Knowledge “Crystals”

Vision: *Push-button NUI from Knowledge Graph*

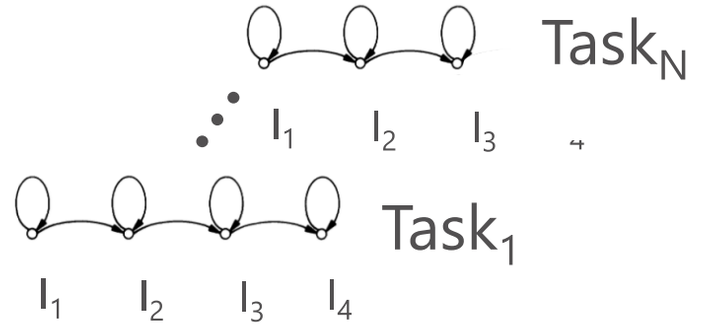


Conversational Knowledge Graphs

Compositionality: Waveforms ➔ Tasks



Pivot on Knowledge

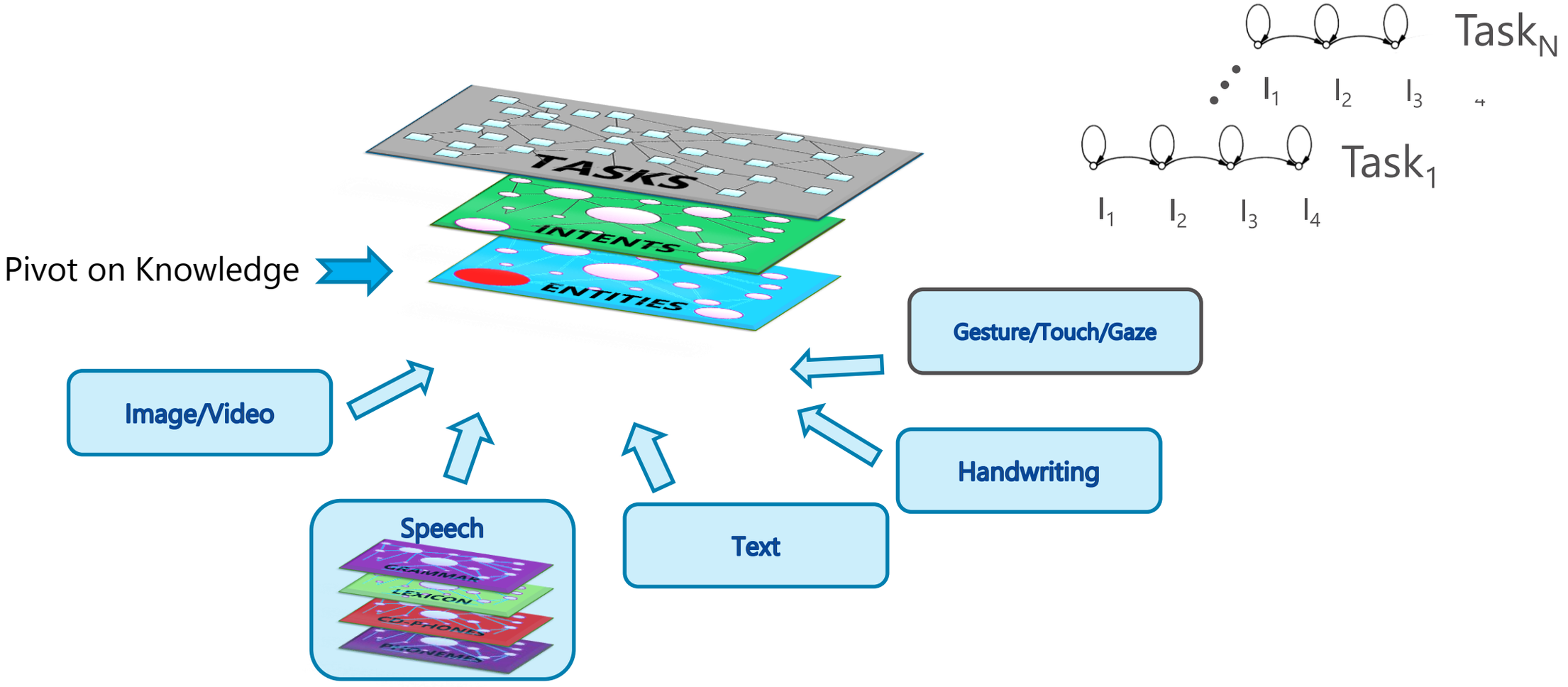


Machine Learned
waveforms → tasks

Waveform

Conversational Knowledge Graphs

Multi-Modal



Semantic Parsing: Unsupervised Learning

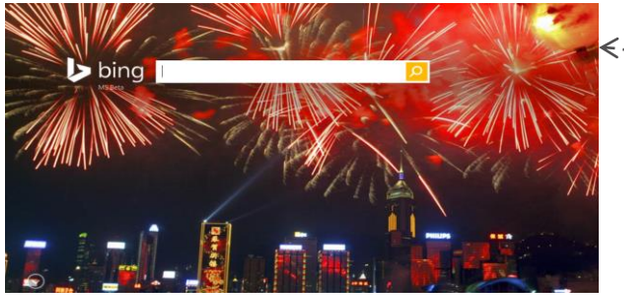
Entity Spotting and Linking

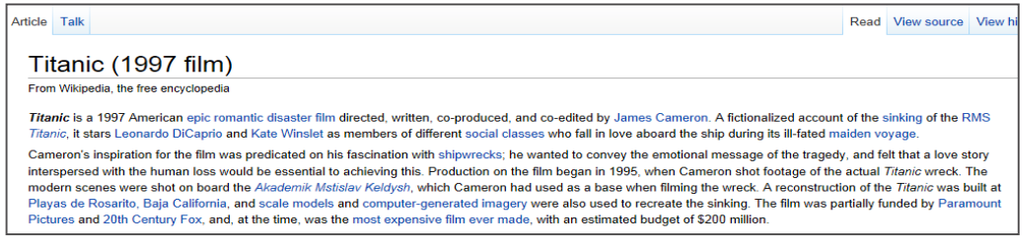
Goal Precise & robust (high-recall) entity linking over a broad knowledge base

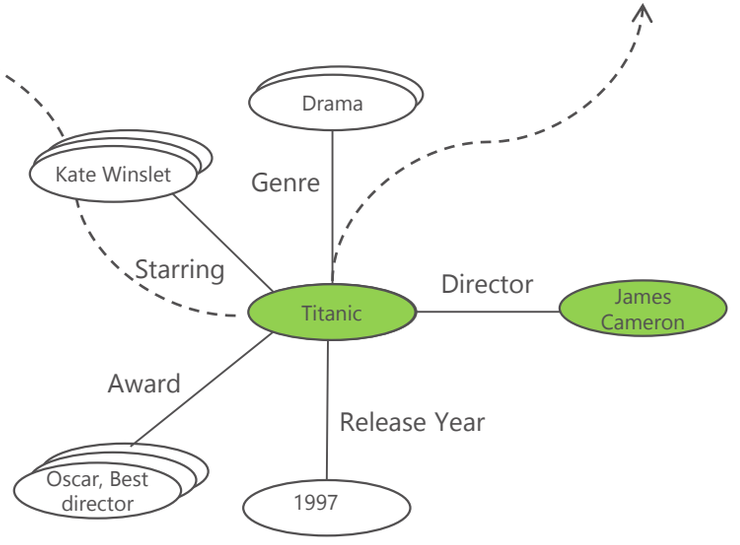
Challenge Requires a lot of *annotated (labeled)* data

Link Satori entities to NL Surface forms:

- Bing queries
- Wikipedia
- Twitter
- MusicBrainz
- IMDB
- etc.







Solution ➔ Start from knowledge graph, mine data, auto-annotate

Semantic Parsing: Unsupervised Learning

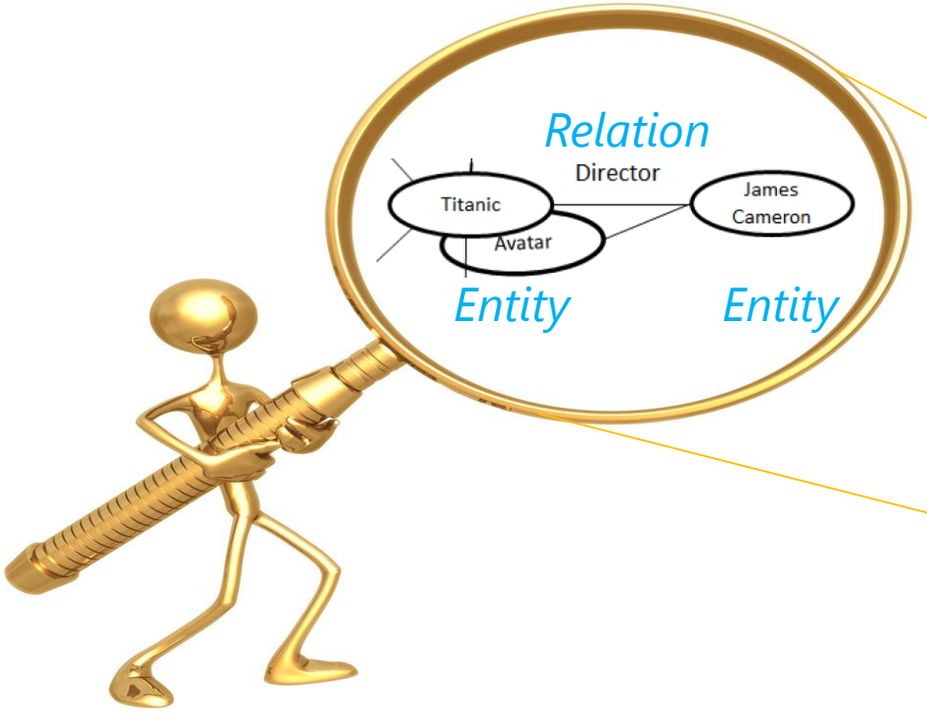
Entity Spotting and Linking

	Manual Transcriptions					ASR Output				
	Movie	**Actor**	**Genre**	**Director**	**All**	**Movie**	**Actor**	**Genre**	**Director**	**All**
Supervised										
CRF Lexical + Gazetteers	51.25%	86.29%	93.26%	64.86%	66.53%	45.15%	82.56%	88.58%	58.59%	60.96%
CRF Lexical only	46.44%	80.22%	92.83%	52.94%	61.72%	39.21%	74.86%	86.21%	45.36%	54.10%
Unsupervised										
Gazetteers only	69.69%	50.70%	15.76%	2.63%	51.14%	59.66%	47.78%	11.80%	2.82%	43.88%
CRF Lexical only	0.19%	9.67%	0.00%	62.83%	5.61%	0.20%	9.67%	0.00%	57.14%	5.27%
+ Gazetteers	1.96%	72.35%	4.73%	79.03%	31.94%	1.74%	69.76%	3.57%	75.00%	30.77%
+ Adaptation	71.72%	58.61%	29.55%	77.42%	60.38%	55.74%	62.70%	30.95%	73.21%	54.69%
+ Relations				**84.62%**	**61.02%**				**80.67%**	**55.40%**

Unsupervised learning ≅ *supervised* (F-measure)

Semantic Parsing: Unsupervised Learning

Induced Relation Grammars



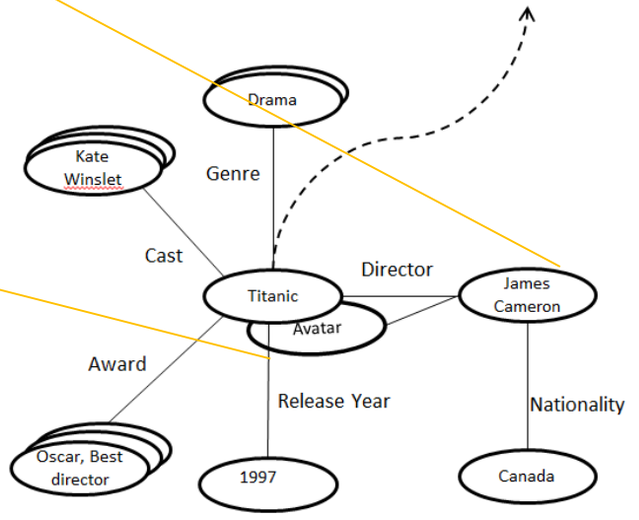
Article Talk Read View source View hi

Titanic (1997 film)

From Wikipedia, the free encyclopedia

Titanic is a 1997 American epic romantic disaster film directed, written, co-produced, and co-edited by James Cameron. A fictionalized account of the sinking of the RMS *Titanic*, it stars Leonardo DiCaprio and Kate Winslet as members of different social classes who fall in love aboard the ship during its ill-fated maiden voyage.

Cameron's inspiration for the film was predicated on his fascination with shipwrecks; he wanted to convey the emotional message of the tragedy, and felt that a love story interspersed with the human loss would be essential to achieving this. Production on the film began in 1995, when Cameron shot footage of the actual *Titanic* wreck. The modern scenes were shot on board the *Akademik Mstislav Keldysh*, which Cameron had used as a base when filming the wreck. A reconstruction of the *Titanic* was built at Playas de Rosarito, Baja California, and scale models and computer-generated imagery were also used to recreate the sinking. The film was partially funded by Paramount Pictures and 20th Century Fox, and, at the time, was the most expensive film ever made, with an estimated budget of $200 million.



Semantic Parsing: Unsupervised Learning

Relation Modeling for Entity Linking

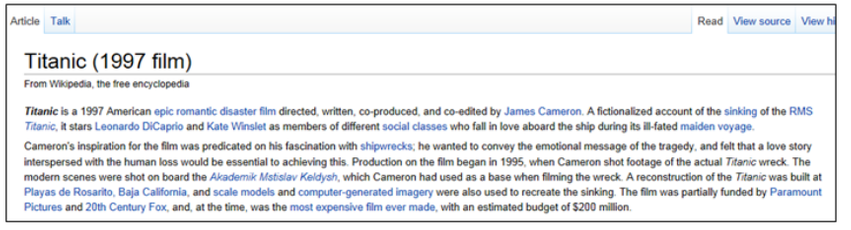
Entities anchor higher-level grammatical structure

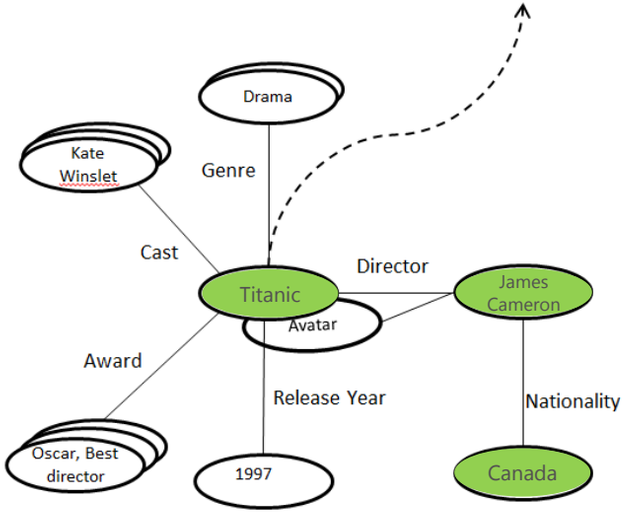
- Induce grammars over high-confidence entities (anchor points)
- "Repair" missing entities

➔ *"Canadian born ___?___ directed titanic"*

Template	Frequency
ent	44.9%
$type \sqcap rel(ent)$	12.8%
$ent_0 \sqcap rel(ent_1)$	7.7%
$ent \sqcap type$	5.8%
$type$	5.8%
$attr(ent)$	3.8%
$ent_1 \sqcap rel(ent_0)$	3.2%
$rel(ent)$	1.9%
$ent_0 \sqcap rel(ent_1, rel(ent_2))$	1.3%
$type_1 \sqcap rel(type_0)$	1.3%

Ten most frequently occurring templates among entity-based queries (Pound et al., CIKM'12)





Semantic Parsing: Unsupervised Learning

Induced Relation Grammars

	Manual Transcriptions					ASR Output				
	Movie	**Actor**	**Genre**	**Director**	**All**	**Movie**	**Actor**	**Genre**	**Director**	**All**
Supervised										
CRF Lexical + Gazetteers	51.25%	86.29%	93.26%	64.86%	66.53%	45.15%	82.56%	88.58%	58.59%	60.96%
CRF Lexical only	46.44%	80.22%	92.83%	52.94%	61.72%	39.21%	74.86%	86.21%	45.36%	54.10%
Unsupervised										
Gazetteers only	69.69%	50.70%	15.76%	2.63%	51.14%	59.66%	47.78%	11.80%	2.82%	43.88%
CRF Lexical only	0.19%	9.67%	0.00%	62.83%	5.61%	0.20%	9.67%	0.00%	57.14%	5.27%
+ Gazetteers	1.96%	72.35%	4.73%	79.03%	31.94%	1.74%	69.76%	3.57%	75.00%	30.77%
+ Adaptation	71.72%	58.61%	29.55%	77.42%	60.38%	55.74%	62.70%	30.95%	73.21%	54.69%
+ Relations				**84.62%**	**61.02%**				**80.67%**	**55.40%**

> *+7% F-measure with induced relation grammars*

Semantic Parsing: Unsupervised Learning

Entity Linking and Relations

To dig deeper…

Larry Heck, Dilek Hakkani-Tur, and Gokhan Tur, Leveraging Knowledge Graphs for Web-Scale Unsupervised Semantic Parsing, in *Proceedings of Interspeech*, International Speech Communication Association, August 2013

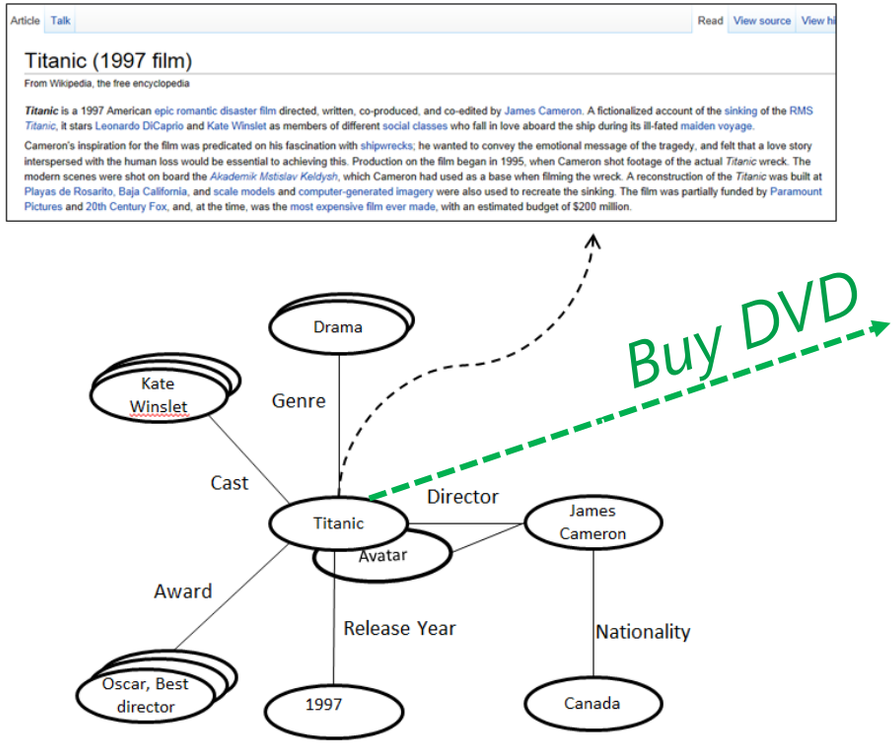
Larry Heck and Dilek Hakkani Tur, Exploiting the Semantic Web for Unsupervised Spoken Language Understanding, IEEE Spoken Language Technology Workshop, December 2012

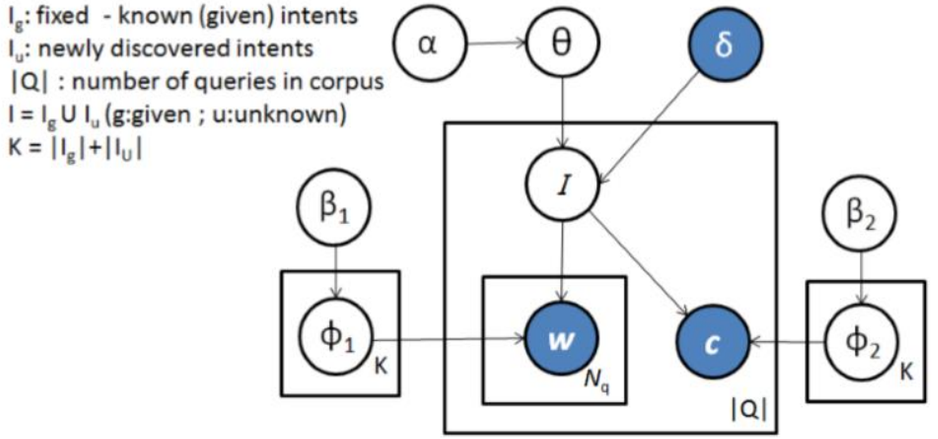
Dilek Hakkani-Tur, Larry Heck, and Gokhan Tur, Using a Knowledge Graph and Query Click Logs for Unsupervised Learning of Relation Detection, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2013

Gokhan Tur, Minwoo Jeong, Ye-Yi Wang, Dilek Hakkani-Tur, and Larry Heck, Exploiting the Semantic Web for Unsupervised Natural Language Semantic Parsing, in *Proceedings of Interspeech*, International Speech Communication Association, 2012

Growing the Knowledge Graph

Discovering new knowledge



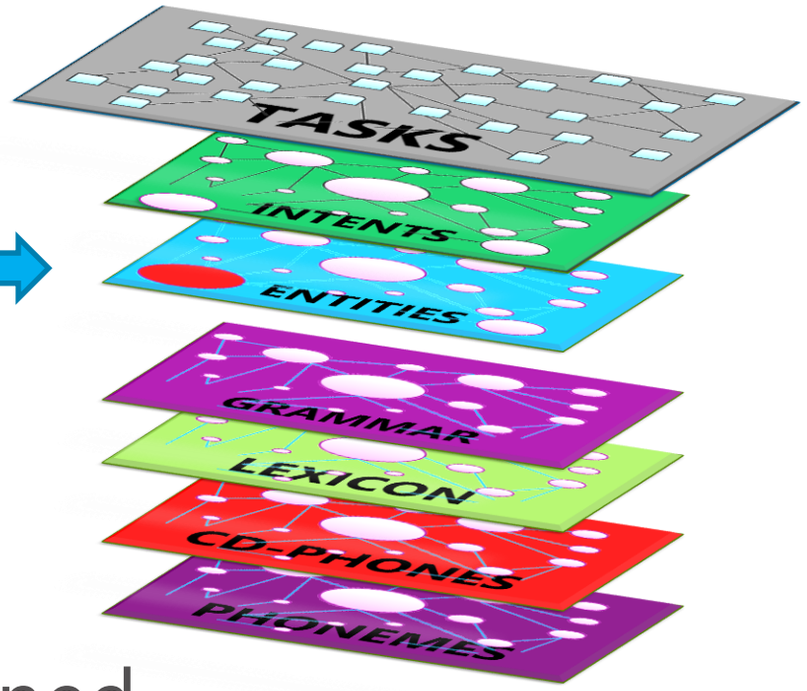


Unsupervised learning ≅ *supervised*

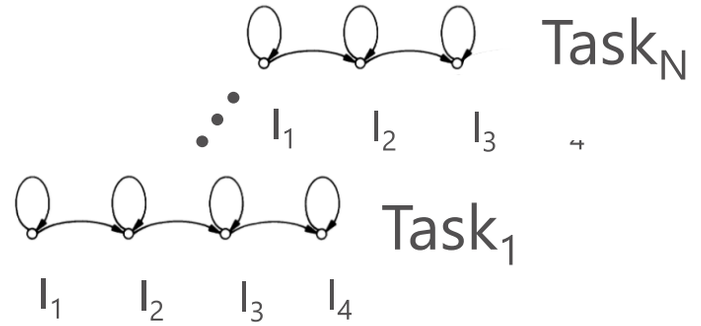
Dilek Hakkani-Tur, Asli Celikyilmaz, Larry Heck, and Gokhan Tur, A Weakly-Supervised Approach for Discovering New User Intents from Search Query Logs, Annual Conference of the International Speech Communication Association (Interspeech), August 2013

Conversational Knowledge Graphs

Compositionality: Waveforms ➔ Tasks



Pivot on Knowledge

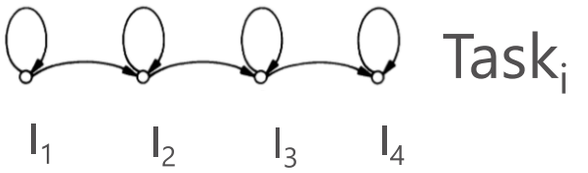


Machine Learned
waveforms → tasks

Waveform

Dialog Modeling with Knowledge Graphs

Dynamic multi-turn conversations



Statistical methods for dialog managers is active research topic (e.g., POMDP)

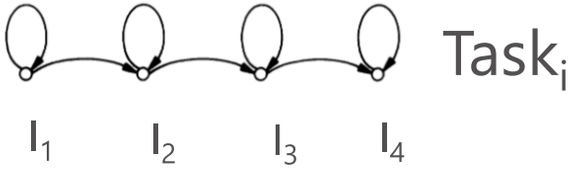
Key Technical Challenge: significant amount of annotated dialogs required for training

Idea: can we leverage Web (IE) session data combined with Knowledge Graphs

Web search & browse ➔ Conversations/dialog
Massive volume of interactions >100M queries/day, Millions of users
Coverage of user interactions is high (broad domains across the web)

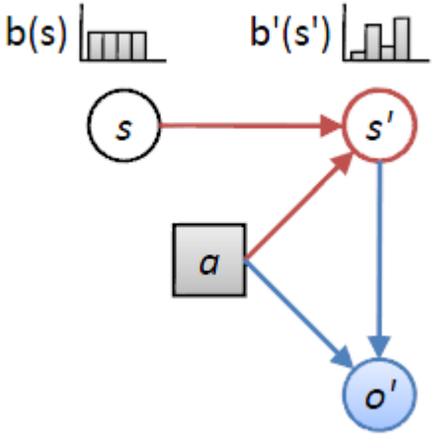
Dialog Modeling with Knowledge Graphs

Dynamic multi-turn conversations



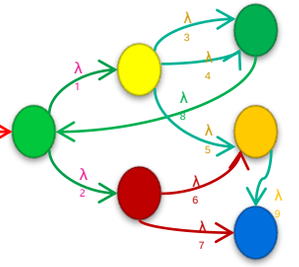
New Approach

Step 1. ***Learn task completion patterns from web*** → IE sessions through Satori Knowledge Graph



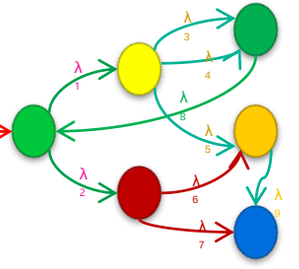
Successful Dialogs

Goal 1: Q 4s RL 1s SR 53s SR 118s END
Goal 2: Q 3s Q 5s SR 10s AD 44s END
Goal 3: Q 4s RL 1s SR 53s SR 118s END
Goal 4: Q 3s Q 5s SR 10s AD 44s END
……
Goal n: Q 4s RL 1s SR 53s SR 118s END
Goal n-1: Q 3s Q 5s SR 10s AD 44s END



Unsuccessful Dialogs

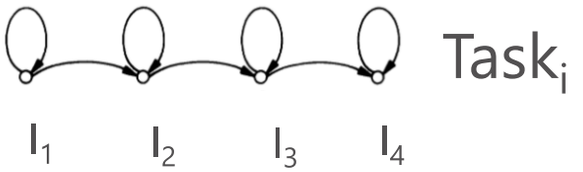
Goal 1: Q 4s RL 1s SR 53s SR 118s END
Goal 2: Q 3s Q 5s SR 10s AD 44s END
Goal 3: Q 4s RL 1s SR 53s SR 118s END
Goal 4: Q 3s Q 5s SR 10s AD 44s END
……
Goal n: Q 4s RL 1s SR 53s SR 118s END
Goal n-1: Q 3s Q 5s SR 10s AD 44s END



Step 2. ***Learn "mapping" of web*** *keyword-click* ***language*** *to natural* ***spoken conversations***

Dialog Modeling with Knowledge Graphs

Dynamic multi-turn conversations



Results

Successfully learned ***conversational search and browse*** models from **IE sessions** + **Satori**

Increased F-measures of semantic parsing by **> 18% (rel.)**

To dig deeper…

Lu Wang, Larry Heck, Dilek Hakkani-Tur, Leveraging Semantic Web Search and Browse Sessions for Multi-Turn Spoken Dialog Systems, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014*

Deep Learning from Structured Knowledge

Current/Future Work

Unsupervised data mining and semantic annotations ➔ *unlimited* training data over knowledge graphs

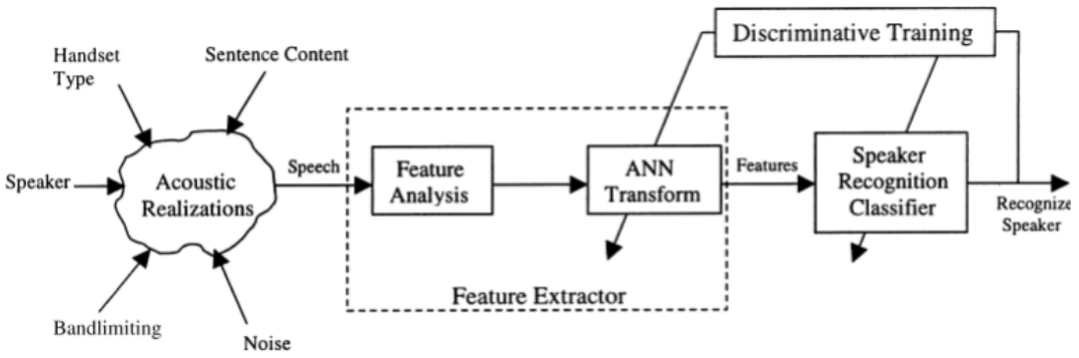
Research Questions:

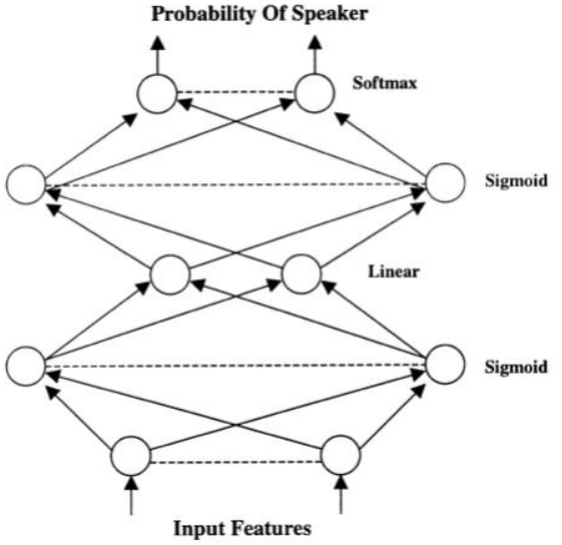
Can deep learning (neural networks) discover the *fundamental features of knowledge?*

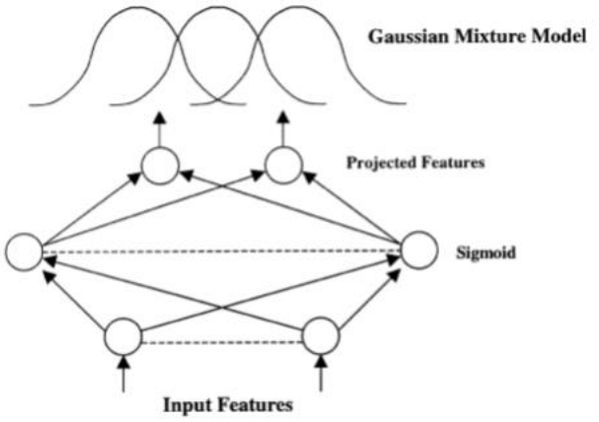
Can we leverage these features to *transfer learning across domains/sub-graphs?*

Back to the Future: Deep Learning for Speaker Recognition

Learning the fundamental *features of speakers*





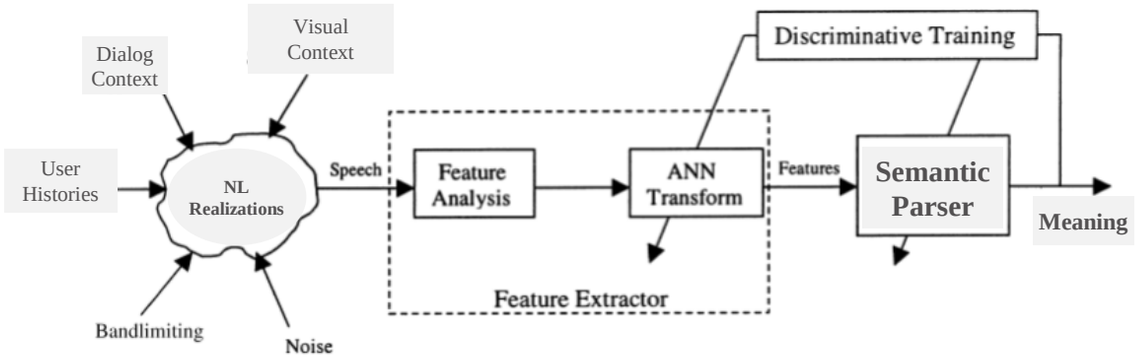


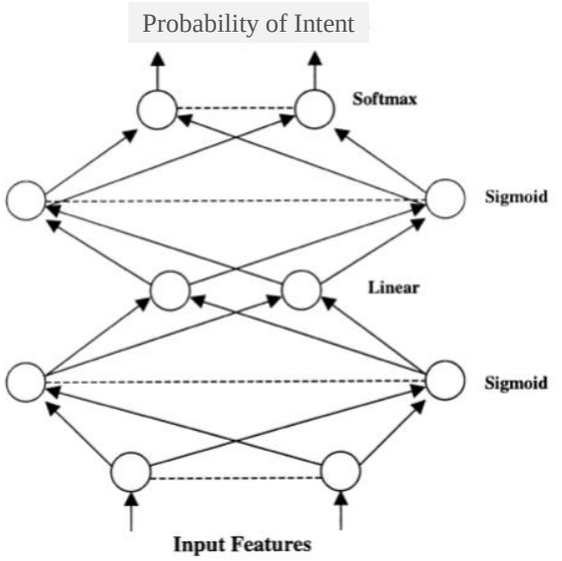
#1 Improvement in NIST 1998 Speaker Recognition Evaluations (+28% ERR)

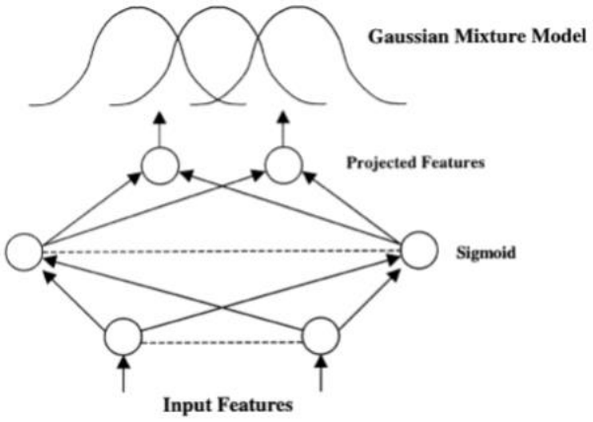
Larry Heck, Yochai Konig, M. Kemal Sonmez, and Mitch Weintraub, Robustness to Telephone Handset Distortion in Speaker Recognition by Discriminative Feature Design, in *Speech Communication*, Elsevier, 2000

Deep Learning for Robust Semantic Parsing

Learn the fundamental *features of natural language*





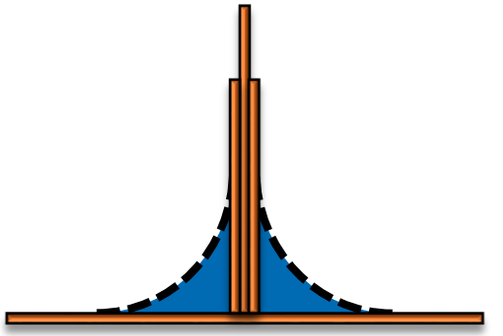


ep Le in ron ledge Graphs

earn al ntal *la* *dge*

Lan chai Konig, M omez, and Mitch Weintraub

Summary
Conversational Systems with Depth & Breadth



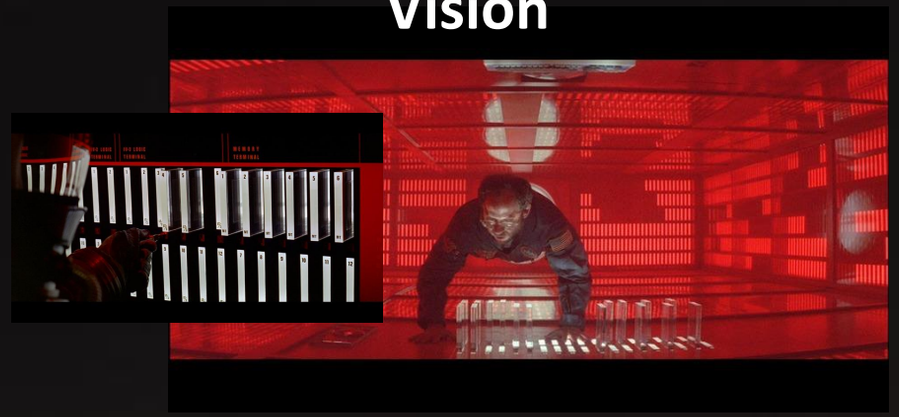
Breadth
Conversational Search and Browse

Depth
Conversational Knowledge Graphs (replacing manually crafted domains)
"Crystals" of knowledge

Breadth *and* Depth
Unsupervised/Weakly-supervised learning methods
Data mining directed by the knowledge graph ➔ *enriched knowledge graphs*
Multi-turn dialog models *compose* entities: learned from Web (IE)
Deep learning from structured knowledge graphs

Microsoft Research

Vision



Strategy

Thank you