An Ontology-Based Conversation System for Knowledge Bases

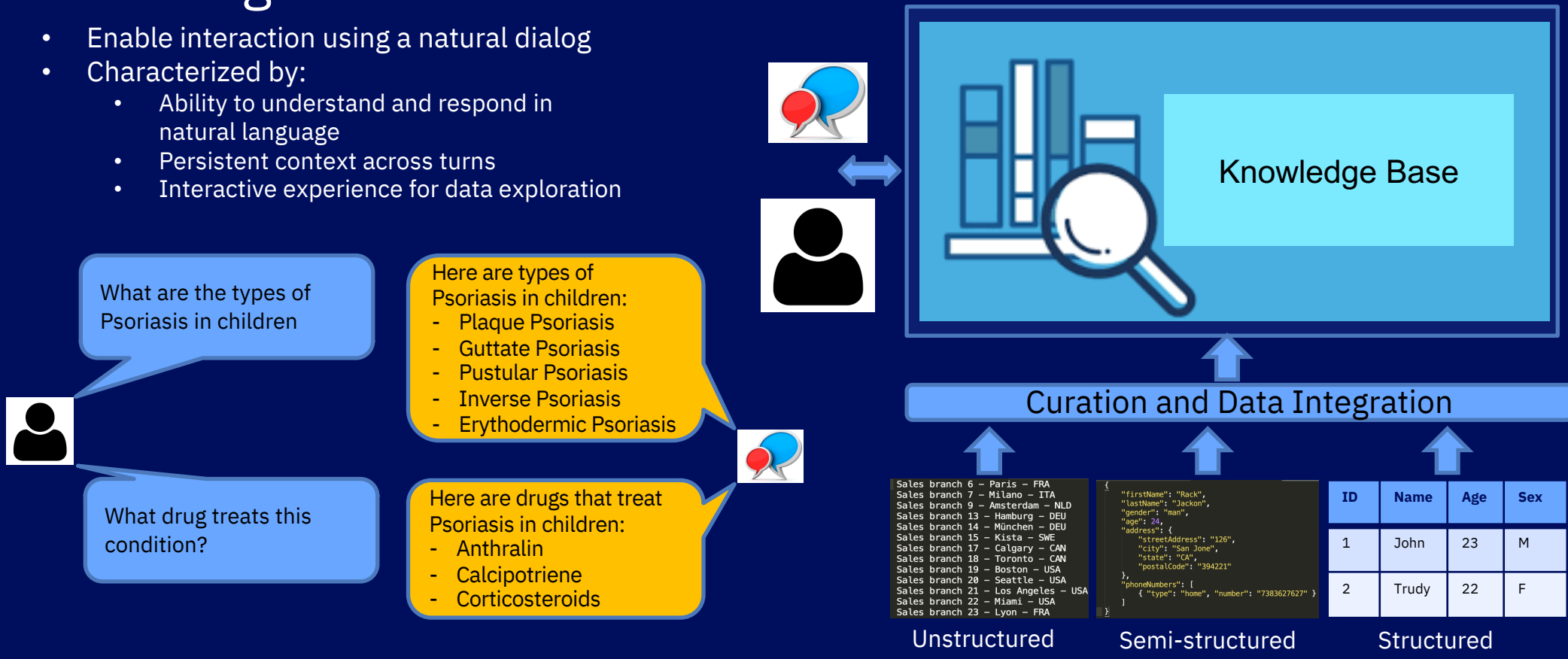
Abdul Quamar[1], Chuan Lei[1], Dorian Miller[2], Fatma Özcan[1], Jeffrey Kreulen[2], Robert J Moore[1], Vasilis Efthymiou[1]

[1]IBM Research-AI, [2]IBM WatsonHealth

IBM

Conversational system for exploring domain specific knowledge bases

- Enable interaction using a natural dialog
- Characterized by:
  - Ability to understand and respond in natural language
  - Persistent context across turns
  - Interactive experience for data exploration

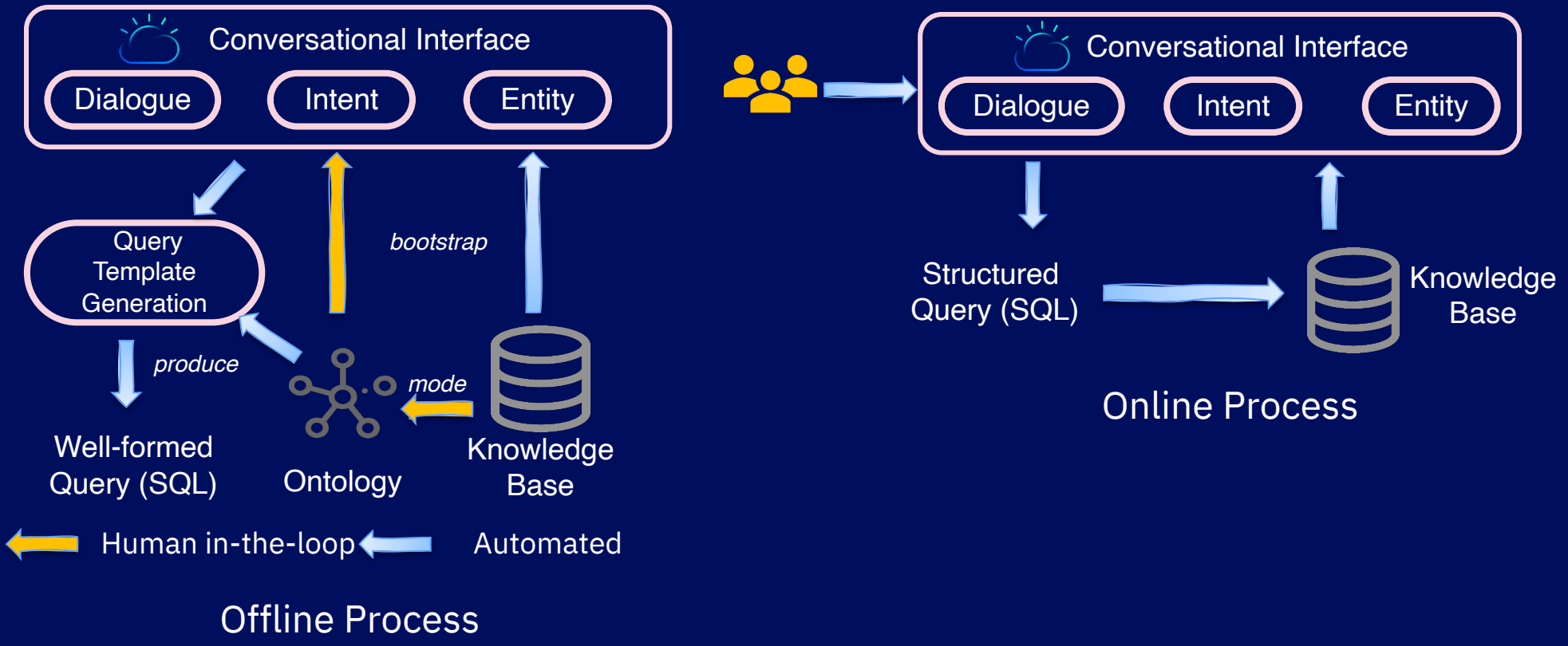


Building conversational systems for domain specific knowledge bases

- Semantic understanding of the knowledge base data
  - Entities, relationships, hierarchies
- Design and automatic population of conversational artifacts
  - User intents: queries that can be answered by the knowledge base
  - Entities: domain vocabulary
  - Dialog: patterns of interaction with the user

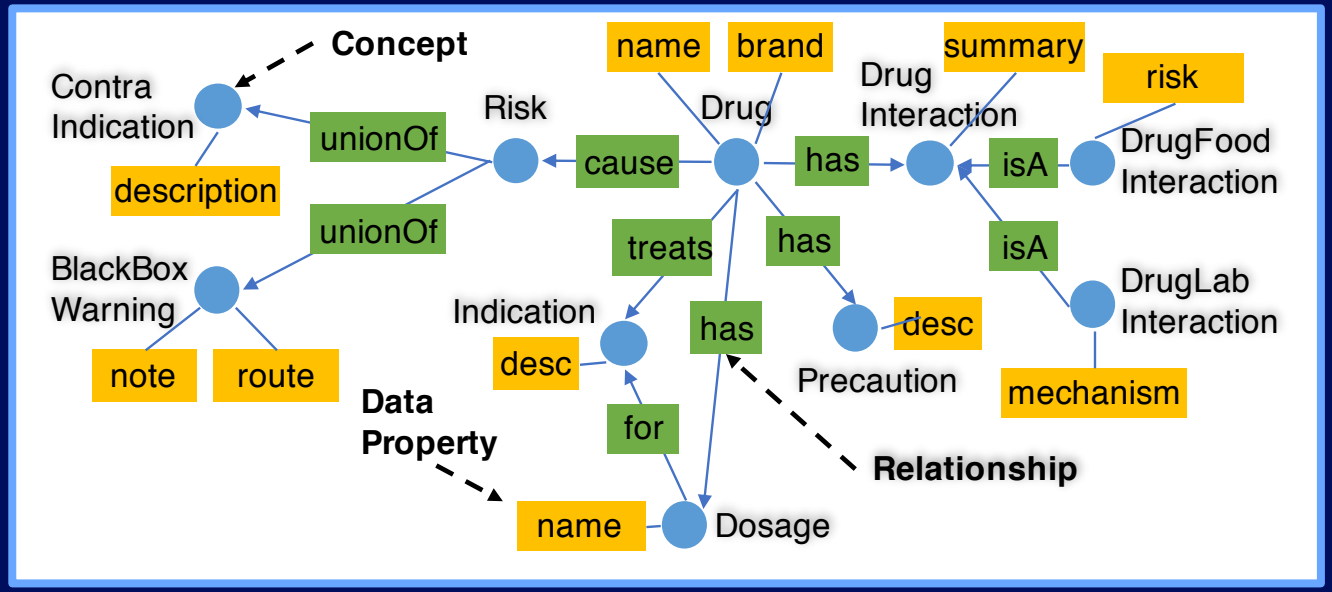
Goal: Develop a principled methodology to build conversation system for querying domain-specific KBs, in a domain-agnostic way

Ontology-based system architecture



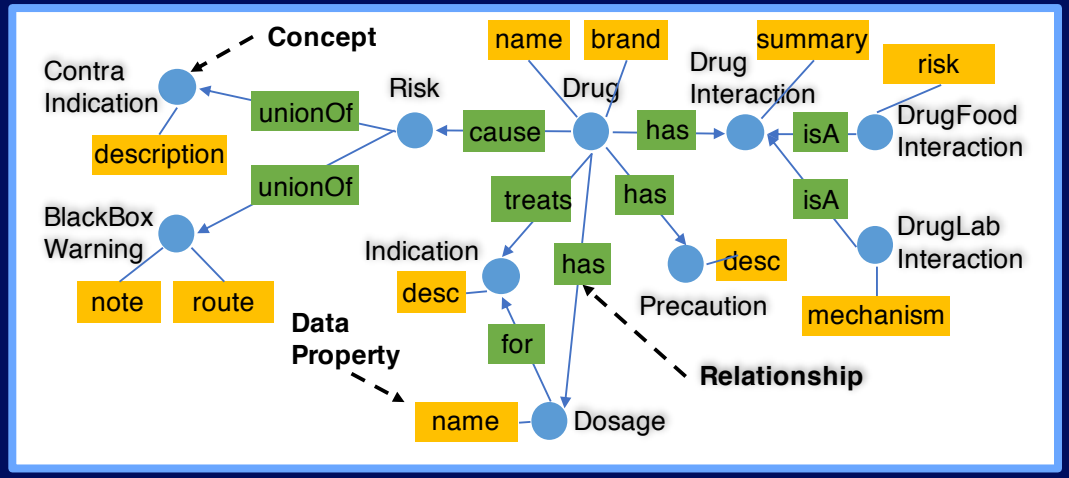
Ontologies to describe the domain schema

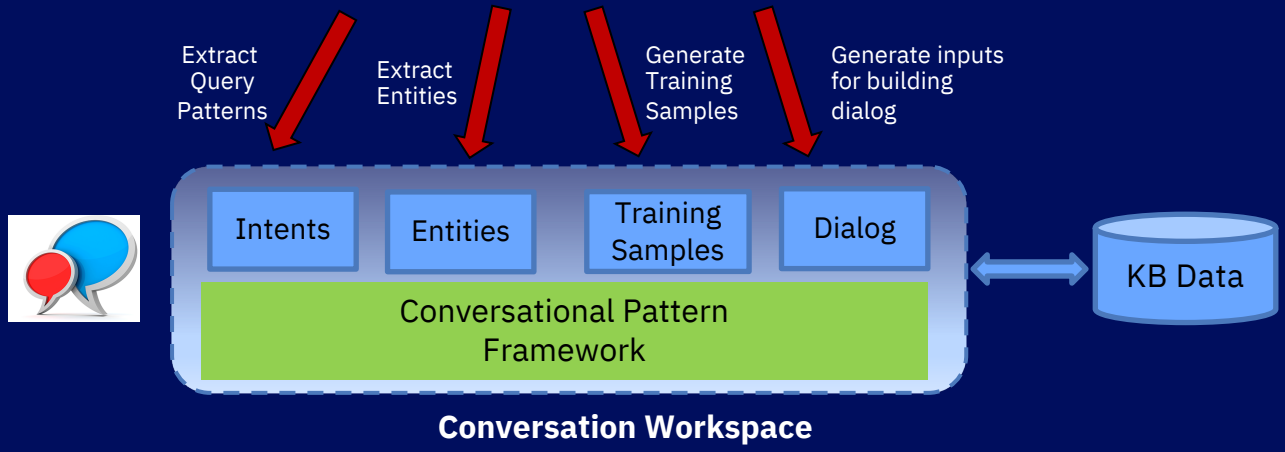
- Ontologies capture the semantics of the domain schema of the knowledge base
  - Concepts
  - Data Properties
  - Relationships
- Provide a rich and expressive data model
- Powerful-object-oriented paradigm
  - Capture real world relationships: Inheritance, Union, Functional



Domain knowledge captured in the ontology + SME feedback -> Enable bootstrapping the conversation space

Bootstrapping the conversation space





Intents:

- Intents express the purpose or goal expressed in the user query/input
- System uses ML Classifiers/Deep NNs to identify intents

Entities:

- Represent real world objects relevant in the context of a user query

Dialog:

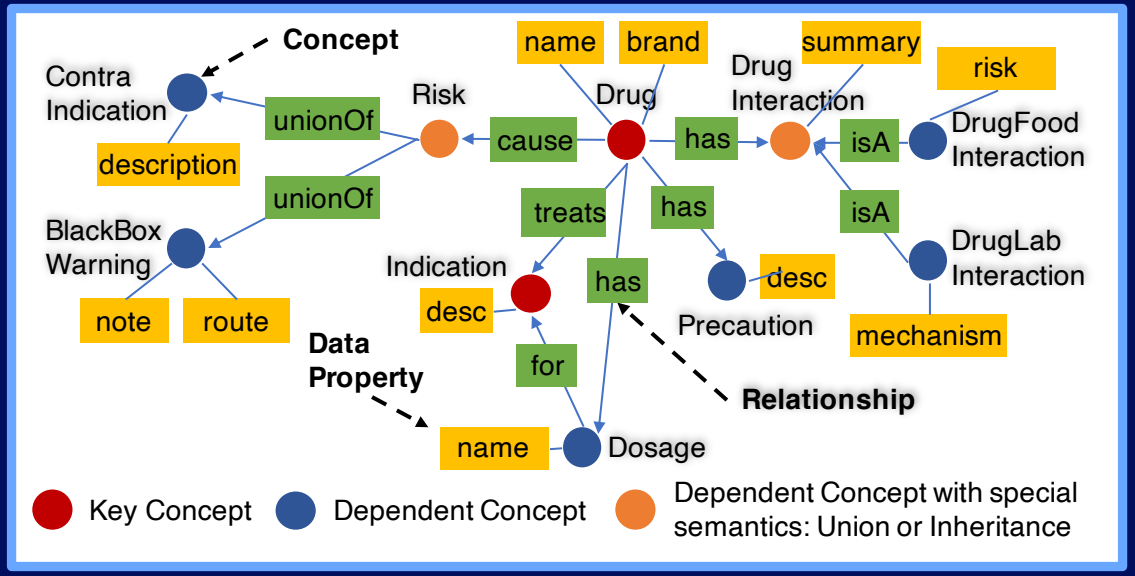
- Uses discovered intents, entities and context from the application to respond to the user

Knowledge Base data:

- Interaction with knowledge base data through structured queries
- Stored under different data models: Json, Relational

Bootstrapping the conversation space: Intent generation

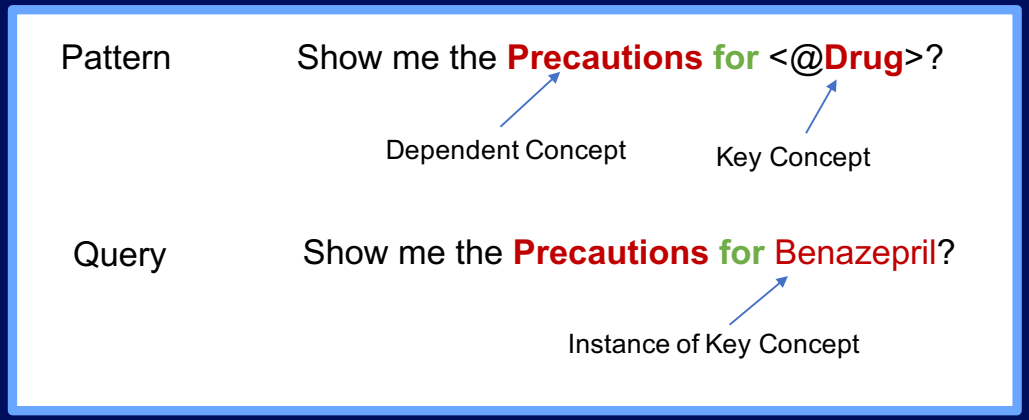
- Key concepts
  - Stand on their own
  - Represent domain entities that a user is interested in
  - Identified using centrality analysis and statistical segregation
- Dependent concepts
  - Concepts in the neighbor-hood of key concepts
  - Use data statistics to identify categorical attributes
  - Union and Inheritance have special semantics
- Intents: query patterns
  - Described as subgraphs around key and dependent concepts



Leveraging the ontology structure

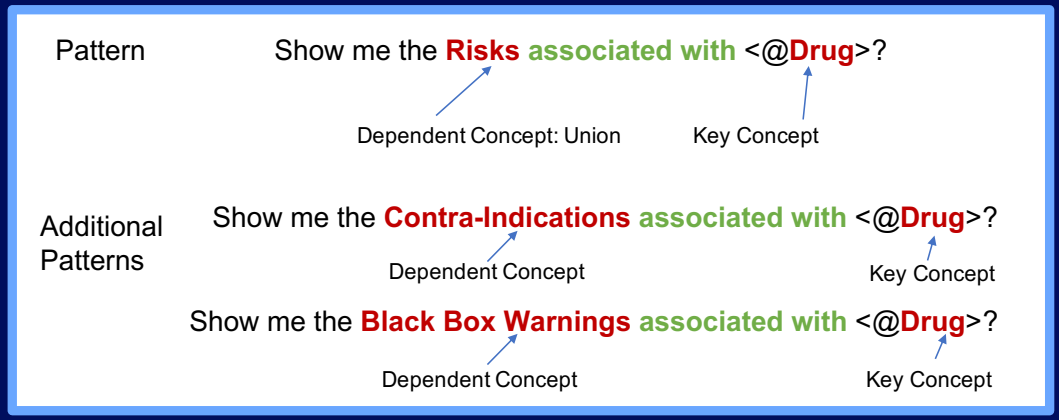
Bootstrapping the conversation space: Intent generation

Query Patterns



Look-up Pattern

Finds information about the key concept with reference to a dependent concept.

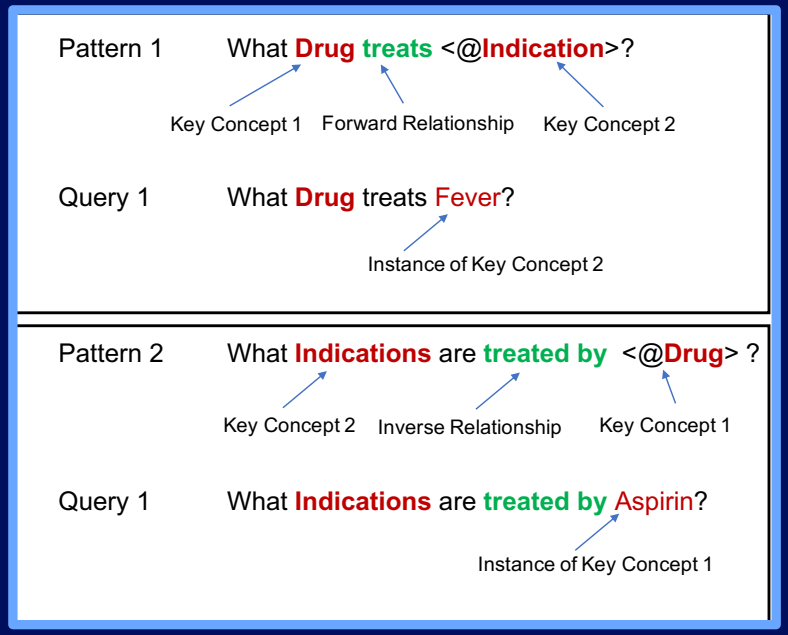


Look-up Pattern Special Semantics

Handles dependent concepts with special semantics:
- Union
- Inheritance

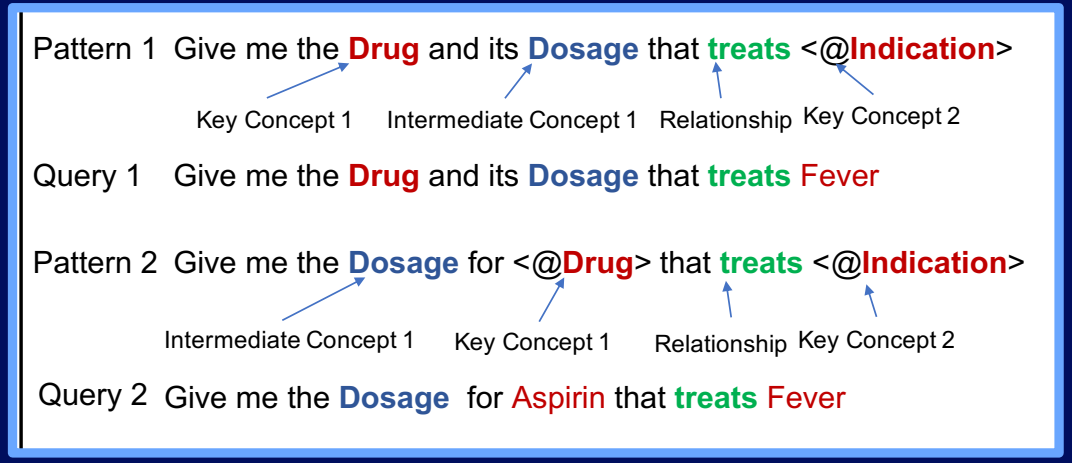
Bootstrapping the conversation space: Intent generation

Query Patterns



Direct Relationship Pattern

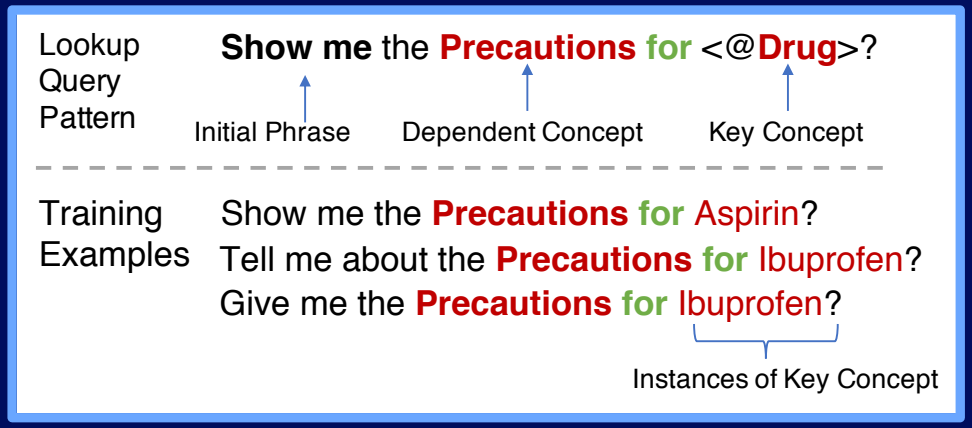
The pair of identified key concepts are connected via at least one direct (one-hop) relationship between them.

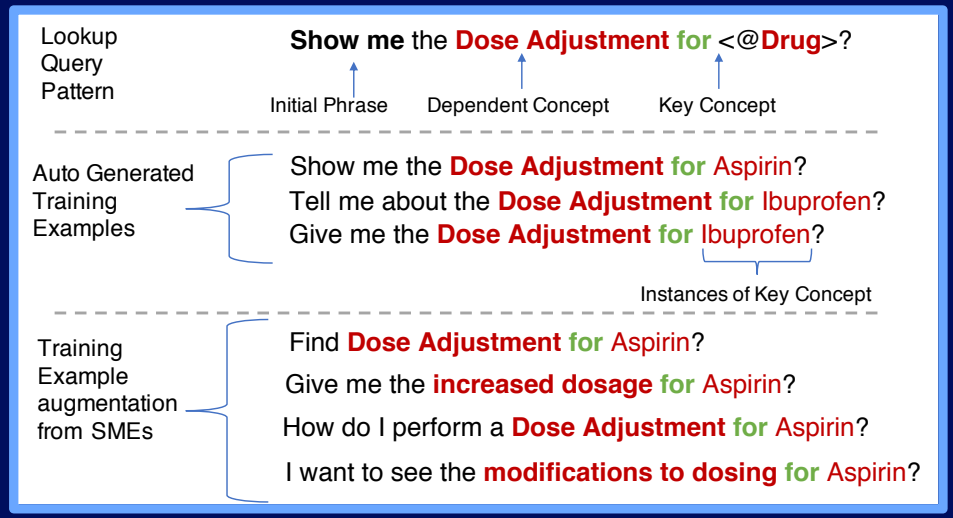


In-Direct Relationship Pattern

The pair of identified key concepts are connected via multiple hops of relationships

Bootstrapping the conversation space: Generating training samples

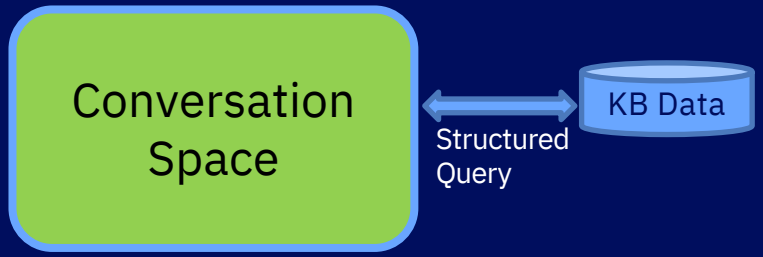




Query patterns viewed as subgraphs over the domain ontology
Entities and corresponding data instances used to automate generation of training examples

Augmentation of training examples using SMEs and prior user experience

Bootstrapping the conversation space: Structured query template generation



- Each intent mapped to a structured query template
- Template populated with identified entities to generate structured query
- Structured query used to retrieve results from the KB

Lookup Query Pattern

Show me the **Precautions** for <@**Drug**>?

Initial Phrase | Dependent Concept | Key Concept

Training Example

Give me the **Precautions** for Ibuprofen?

Structured Query(SQL)

```
SELECT oPrecautions.description
 FROM Precautions oPrecautions  INNER JOIN Drug oDrug
 WHERE oPrecautions.for=oDrug.DrugID
        AND oDrug.name = 'Ibuprofen'
```

Structured Query Template

```
SELECT oPrecautions.description
 FROM Precautions oPrecautions  INNER JOIN Drug oDrug
 WHERE oPrecautions.for=oDrug.DrugID
        AND oDrug.name = '<@Drug>'
```

Bootstrapping the conversation space: Entity extraction

Entities represent the domain vocabulary of the conversational system

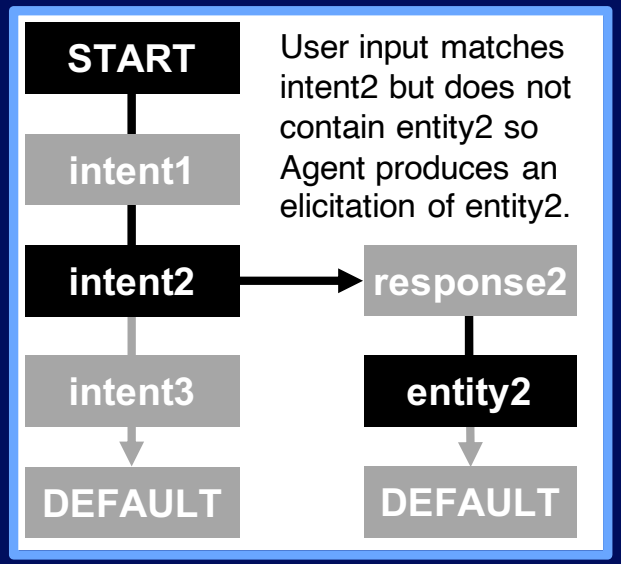
- All concepts from the ontology
- Categorical attributes
  - Data instances added from KB
- Domain specific synonyms for
  - Ontology concepts
  - Data instances

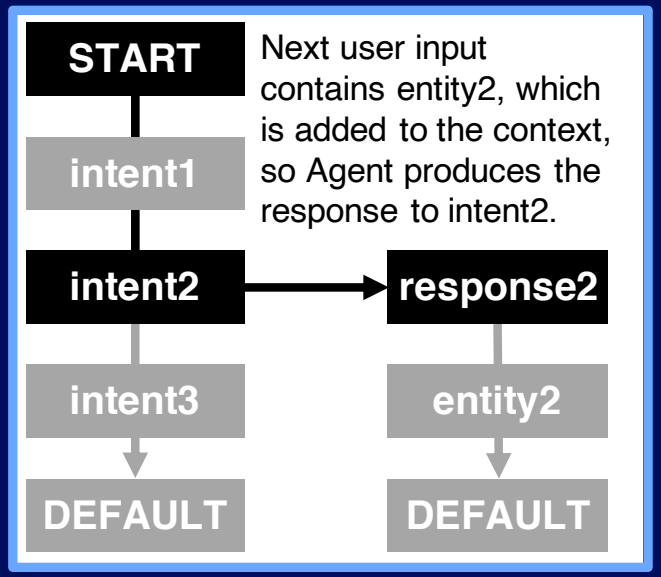
Entities:	Examples	
Concepts:	Drug, Precautions, Dosage, Indication	Ontology concepts
Risk:	Contra-Indication, Black Box Warning	Concepts grouped under Risk
Drug Interaction:	DrugFood Interaction, DrugLab Interaction	Concepts grouped under Drug Interaction
Drug:	Aspirin, Ibuprofen, Citicoline, Pancreatin	Instance values of Drug
Indication:	Fever, Headache, Bronchitis, Diabetes	Instance values of Indication
Contra-Indication:	Cardiovascular disease, Breast carcinoma	Instance values of Contra-Indication

Entity	Synonyms
Adverse Effect:	Side effect, adverse reaction, adverse event, AE
Condition:	disease, finding, disorder
Drug:	medicine, meds, medication, substance
Precaution:	caution, safe to give
Dosage:	dosing
Dose adjustment:	dose modification, dosing modification, dose reduction

Bootstrapping the conversation space: Building dialog

- Dialog tree
  - Defines the space of user utterances the system can recognize and respond to
  - Enables interactive experience for the user
  - Responses conditioned on
    - Combination of intents and entities identified in the user utterance
    - Context captured from previous utterances





Designed to handle both domain specific requests and general conversation management

IBM Micromedex (MDX)

An evidence-based clinical decision support application

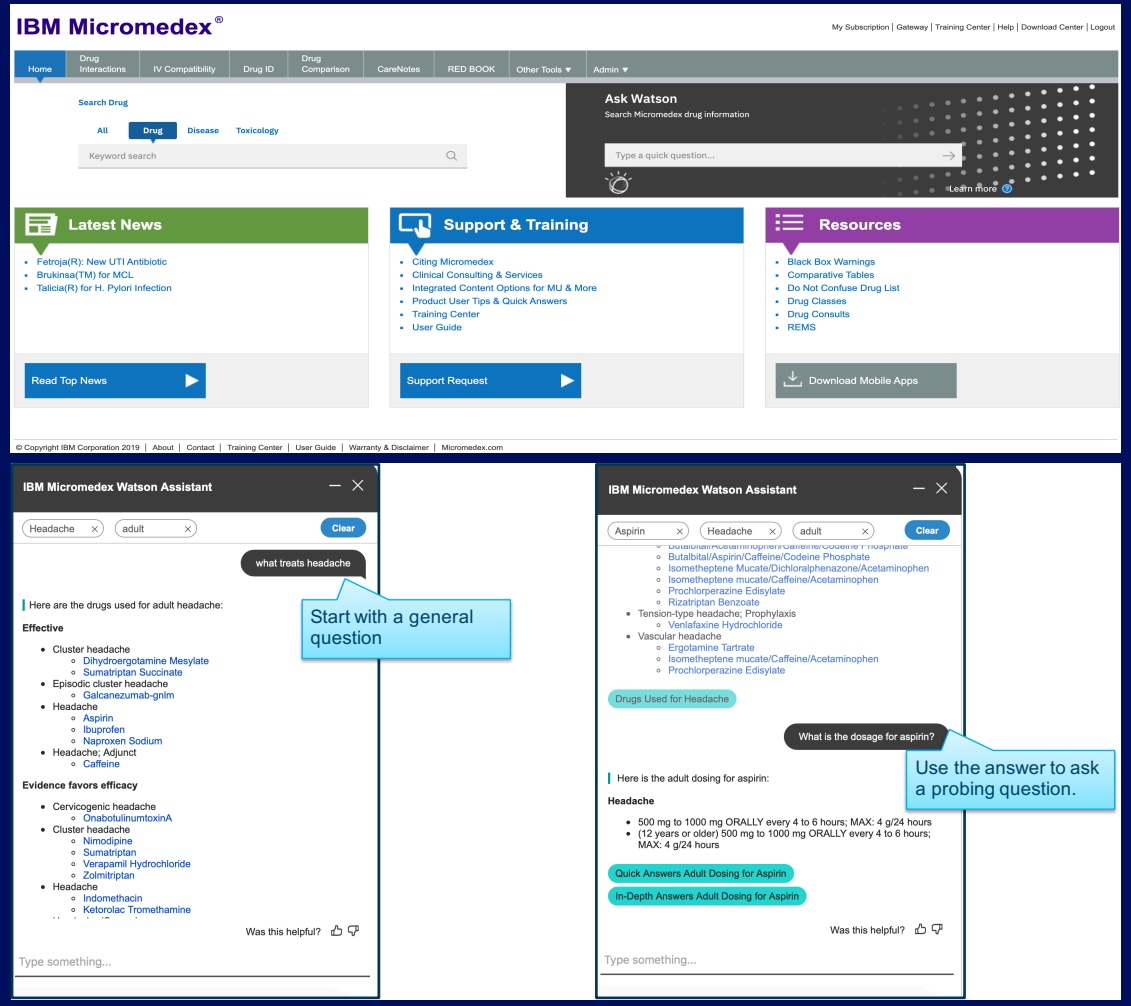
- Summary and expanded content from the world's biomedical literature
- Includes drug information, toxicology, diseases and conditions, and alternative medicine

MDX ontology

- 59 Concepts, 178 Properties, and 58 Relationships

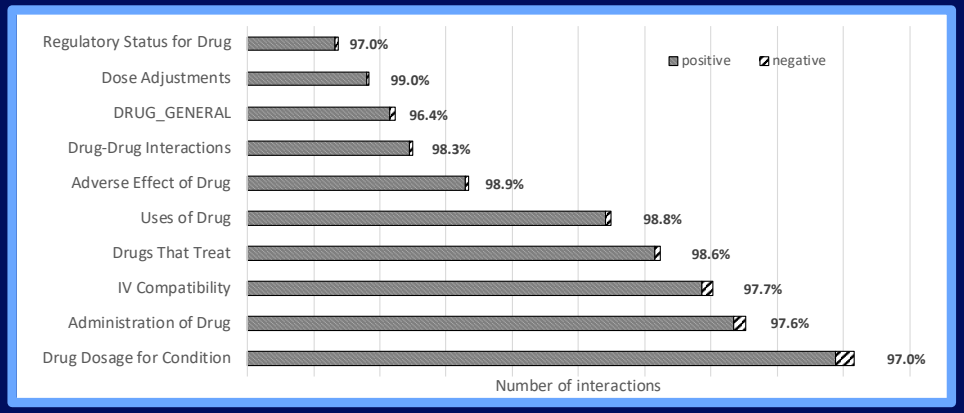
Conversation space

- 36 intents
  - 14 lookup and 8 relationship patterns
  - 14 intents for conversation management
- 52 Entities and corresponding data instances

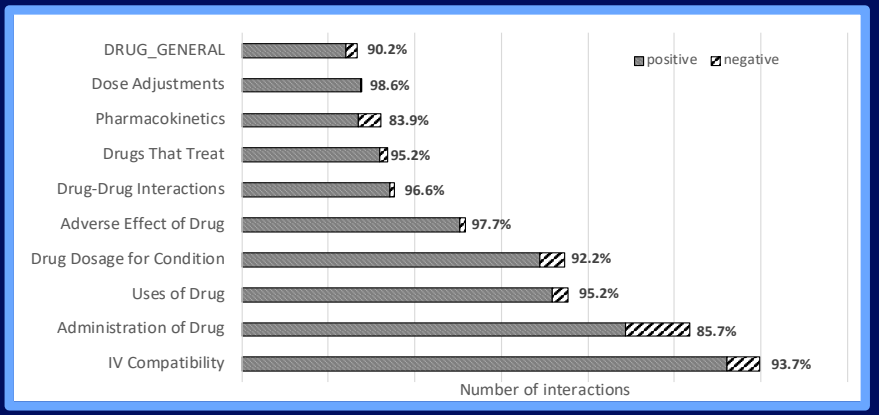


Conversational Micromedex Works Great!!

Success rate per intent provided by users for top 10 intents



Success rate per intent provided by SMEs for top 10 intents



Average success rate across all intents 96.3%

Top 10 intents account for 75% or workload

Overall F1-Score of Intent identification 85%

Conclusions

- Our ontology-based framework provides sufficient semantic information for building an effective conversation system
- Our bootstrapping mechanism creates a rich conversation space
- Access to prior user queries and SME knowledge further improves the precision
- Our results show that the overall conversational system is very effective in exploring knowledge bases
  - Average success rate across all intents is 96.3%


thank you
danke
рахмат
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
bedankt
dziękuję
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
terima kasih
감사합니다
ευχαριστώ
ありがとう
спасибо
Баярлалаа
spas
mersi
kia ora
barka
welalin
tack
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
хвала
asante
manana
obrigada
tenki
murakoze
vinaka
спасибі
blagodaram
faafetai lava
kiitos
dankie
nanni
nandri
dhanyavad
mauruuru
hvala
köszönöm
gràçie
bayarlalaa
enkosi
dankon
akun
aćiū
djiere dieuf
tau
дякую
mamnun
chokrane
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
sobodi
děkuji
sagolun
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmat
tanemirt
rahmet
grazie
arigatô
takk
dakujem
trugarez
diolch
dhanyavadagalu
shukriya
mercê
мерси
xièxie
তোমাকে ধন্যবাদ

IBM