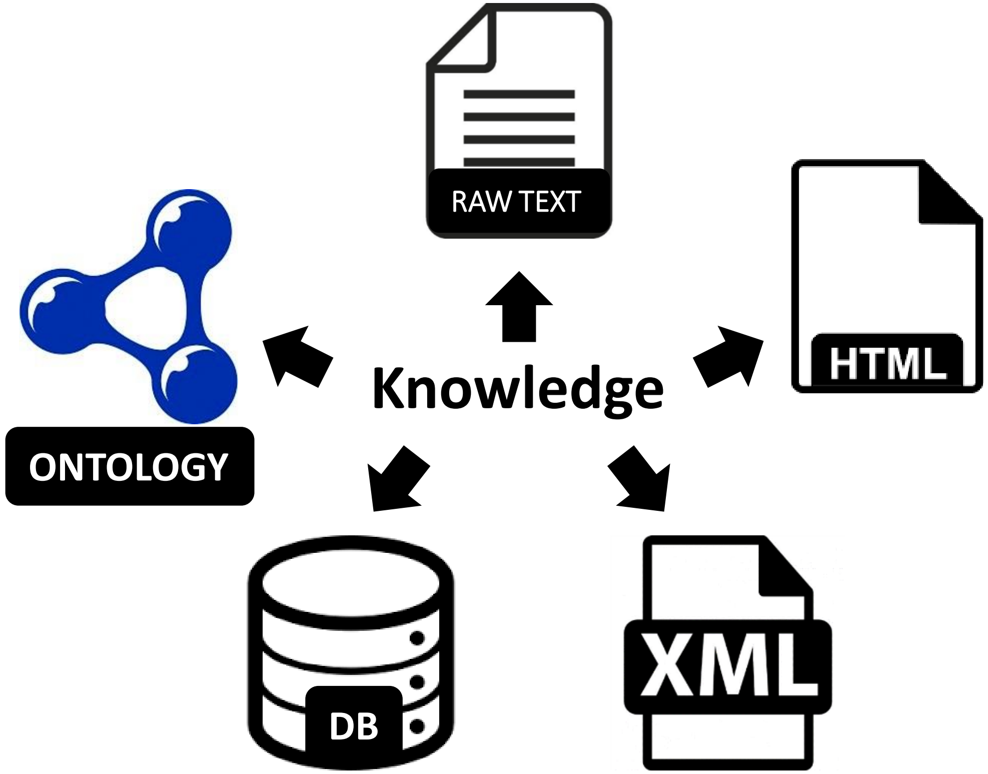
Ontology-based Question Generation Systems

Vinu E.V (CS12D019), Dept. of CSE, IIT Madras
Guide: Prof. P Sreenivasa Kumar

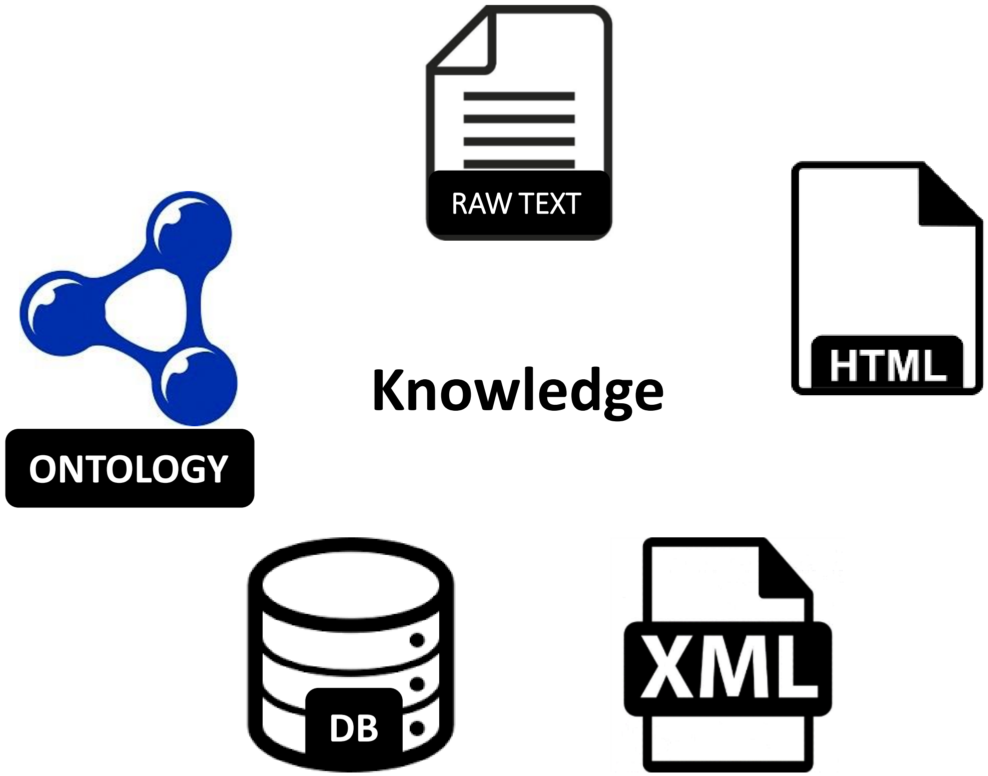
Outline

- ❖ Introduction
- ❖ Fundamentals of Ontology
- ❖ Motivation
- ❖ Existing Approaches
- ❖ Thesis – Statement & Overview
- ❖ Research work and Results
- ❖ Conclusion & Future works
- ❖ Publications & References

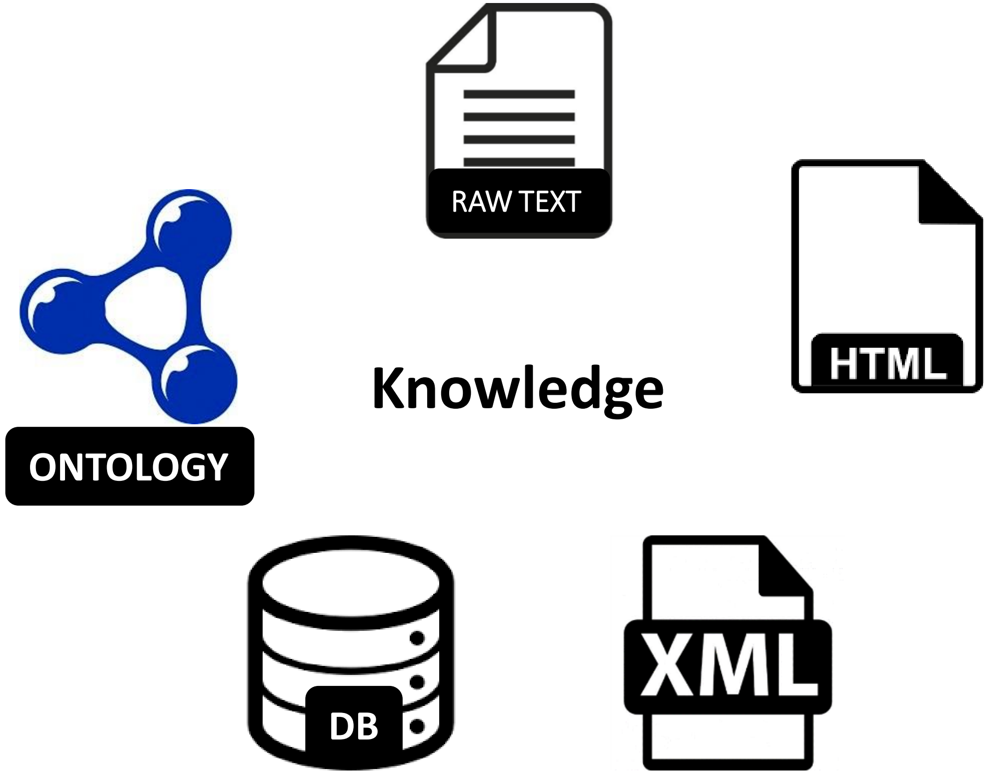
Introduction



Introduction



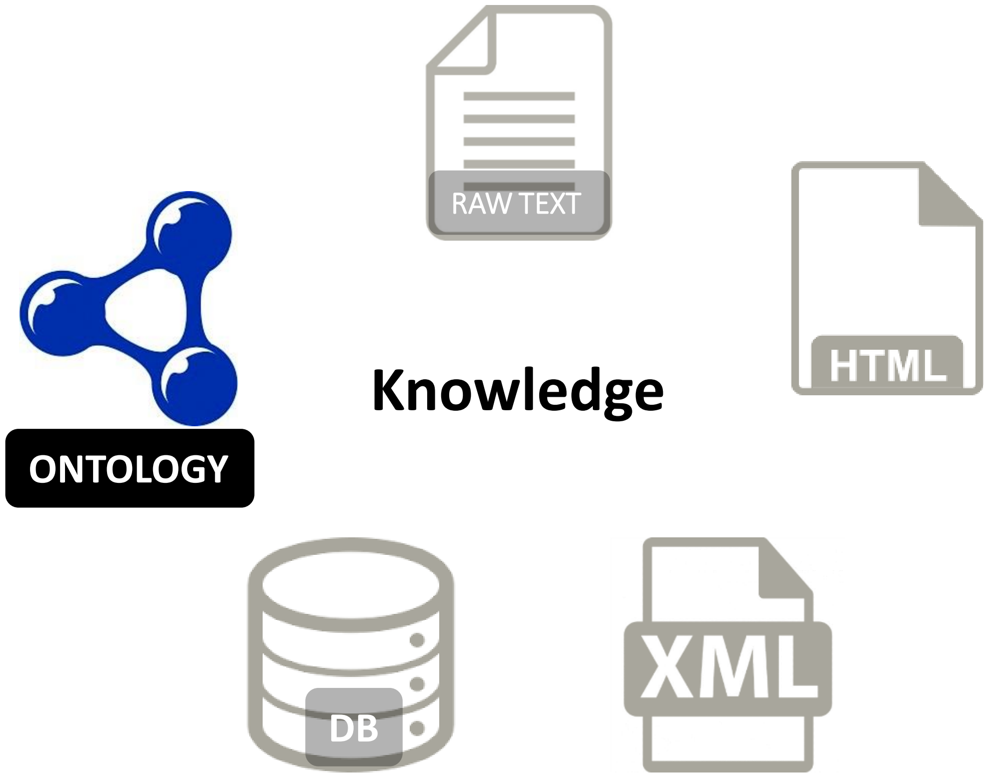
Automation



Automation

Knowledge represented **logically**

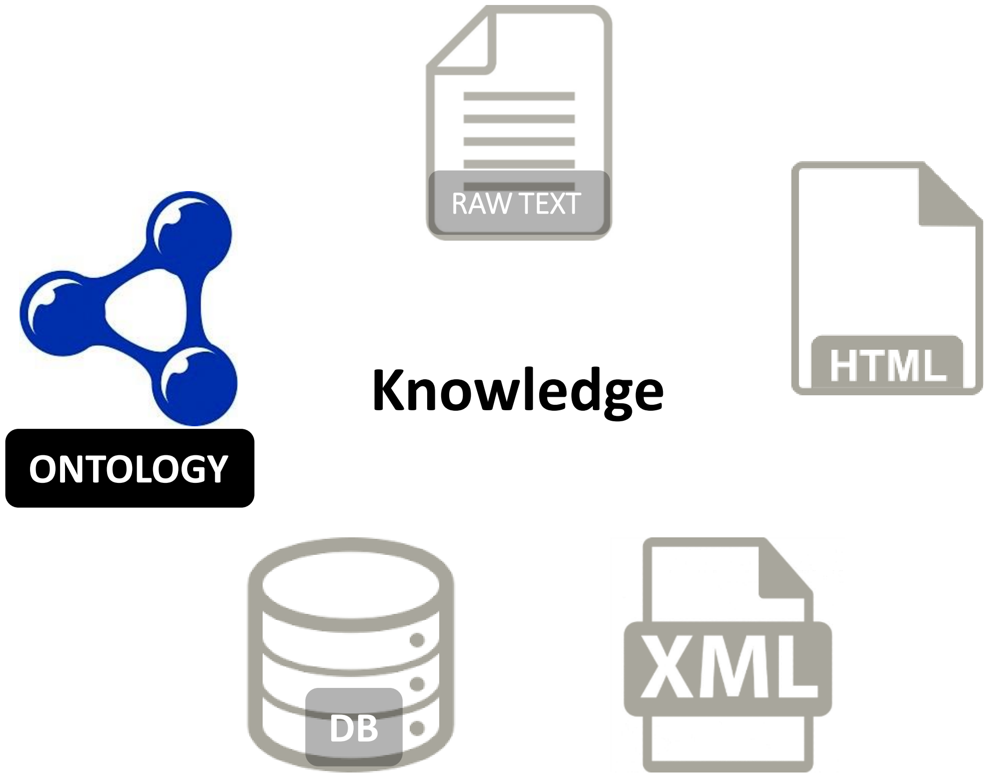
Introduction



Automation

Knowledge represented **logically**

Introduction



Automation

Knowledge represented **logically**

Description Logics(DL) based Ontologies

Fundamentals of DL Ontology

Description Logics (DLs) are **decidable fragments** of **first order logic**

Fundamentals of DL Ontology

Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C)	**Roles (R)**	**Individuals (I)**	**Literals (L)**
Movie	hasDirector	argo	“May 21st 2011”
Actor	wonAward	titanic	“2314”

Fundamentals of DL Ontology

Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C)   Roles (R)   Individuals (I)   Literals (L)

DL-based Ontologies have **two parts**

Fundamentals of DL Ontology

Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C) Roles (R) Individuals (I) Literals (L)

DL-based Ontologies have two parts

Knowledge Graph
(ABox)

$G(V,E)$ $V = I \cup C \cup L$ $E = R$

Directed Graph

Schema
(TBox)

Concept Inclusion ($C \sqsubseteq D$)
Concept Equality ($C \equiv D$)
Role Hierarchy ($R \sqsubseteq S$)
Role Transitivity ($Tran(R)$) etc.

Fundamentals of DL Ontology

Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C) Roles (R) Individuals (I) Literals (L)

DL-based Ontologies have two parts

Based on **Constructors** used in the TBox, DLs can be classified into:

Knowledge Graph
(ABox)

$G(V,E)$ $V = I \cup C \cup L$ $E = R$

Directed Graph

Schema
(TBox)

Concept Inclusion ($C \sqsubseteq D$)

Concept Equality ($C \equiv D$)

Role Hierarchy ($R \sqsubseteq S$)

Role Transitivity (Tran(R)) etc.

Fundamentals of DL Ontology

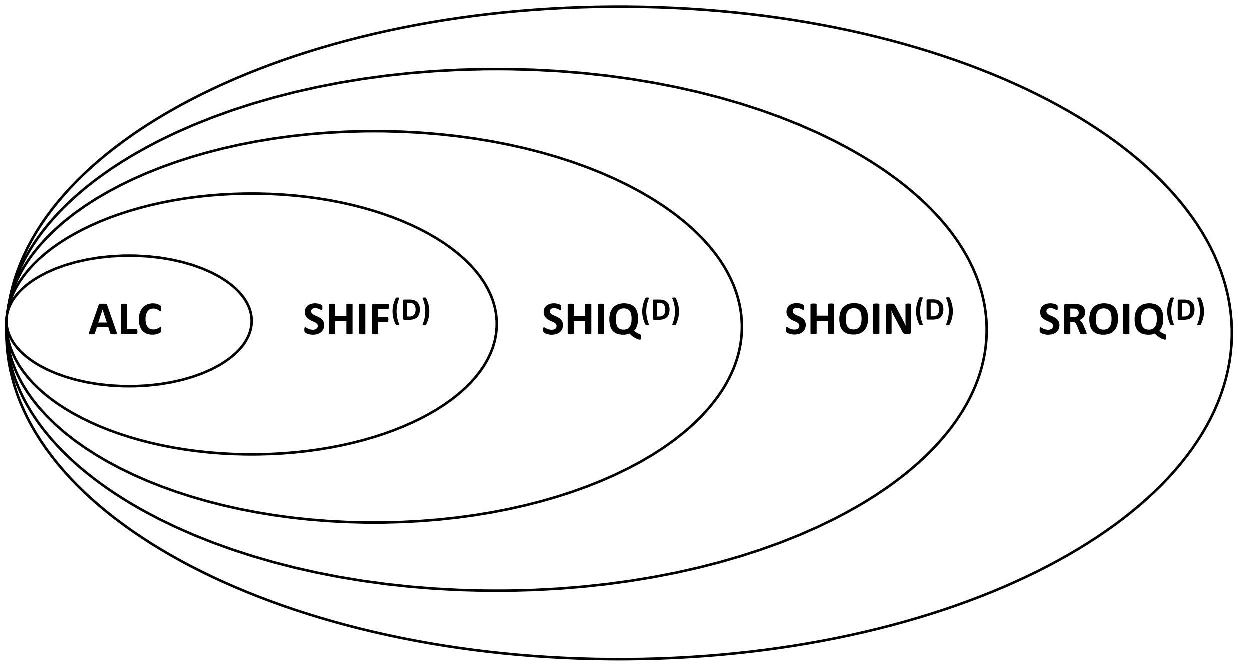
Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C)   Roles (R)   Individuals (I)   Literals (L)

DL-based Ontologies have two parts

Based on Constructors used in the TBox, DLs can be classified into:



Knowledge Graph
(ABox)

G(V,E) $V = I \cup C \cup L$ E= R

Directed Graph

Schema
(TBox)

Concept Inclusion ($C \sqsubseteq D$)

Concept Equality ($C \equiv D$)

Role Hierarchy ($R \sqsubseteq S$)

Role Transitivity (Tran(R)) etc.

Fundamentals of DL Ontology

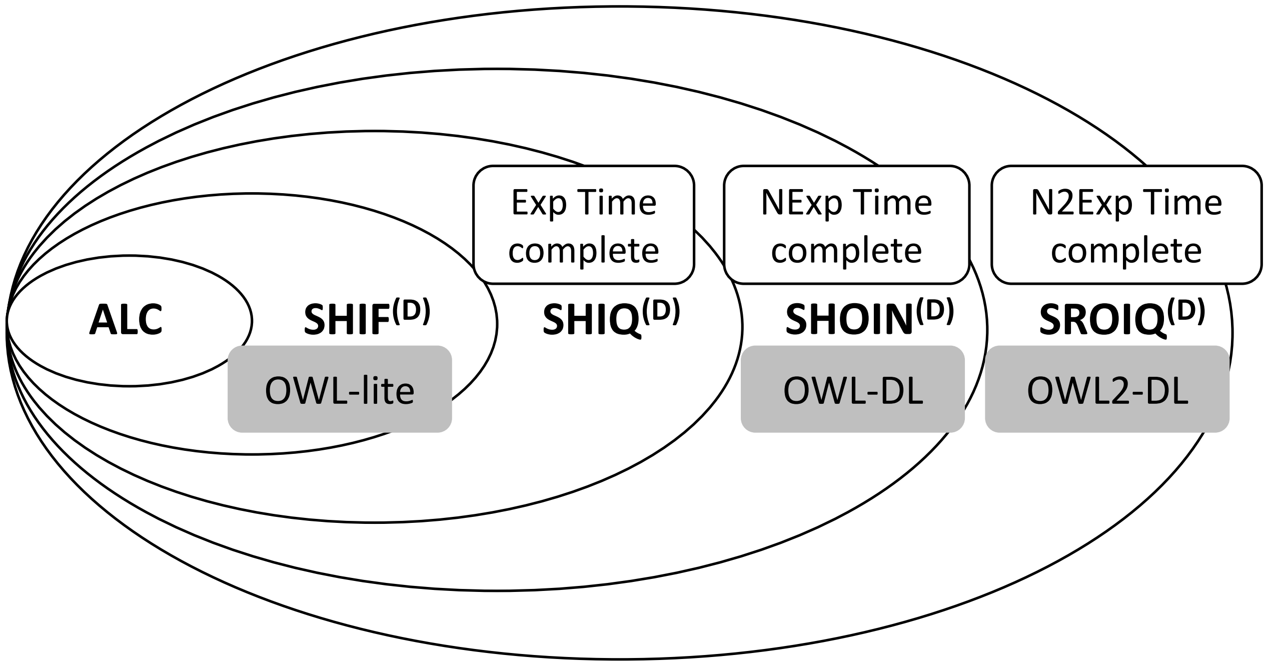
Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C)   Roles (R)   Individuals (I)   Literals (L)

DL-based Ontologies have two parts

Based on Constructors used in the TBox, DLs can be classified into:



Knowledge Graph
(ABox)

G(V,E) V = I ∪ C ∪ L E= R

Directed Graph

Schema
(TBox)

Concept Inclusion ($C \sqsubseteq D$)
Concept Equality ($C \equiv D$)
Role Hierarchy ($R \sqsubseteq S$)
Role Transitivity (Tran(R)) etc.

Fundamentals of DL Ontology

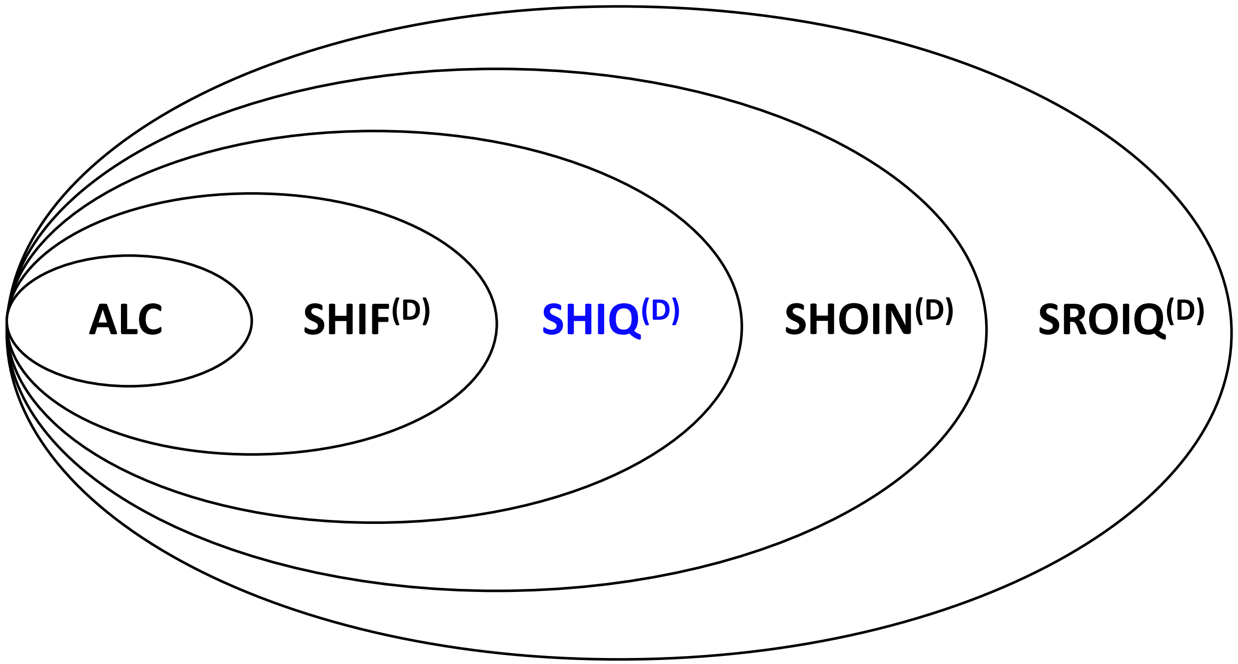
Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C) Roles (R) Individuals (I) Literals (L)

DL-based Ontologies have two parts

Based on Constructors used in the TBox, DLs can be classified into:



Knowledge Graph
(ABox)

$G(V,E)$ $V = I \cup C \cup L$ $E = R$

Directed Graph

Schema
(TBox)

Concept Inclusion ($C \sqsubseteq D$)
Concept Equality ($C \equiv D$)
Role Hierarchy ($R \sqsubseteq S$)
Role Transitivity (Tran(R)) etc.

Fundamentals of DL Ontology

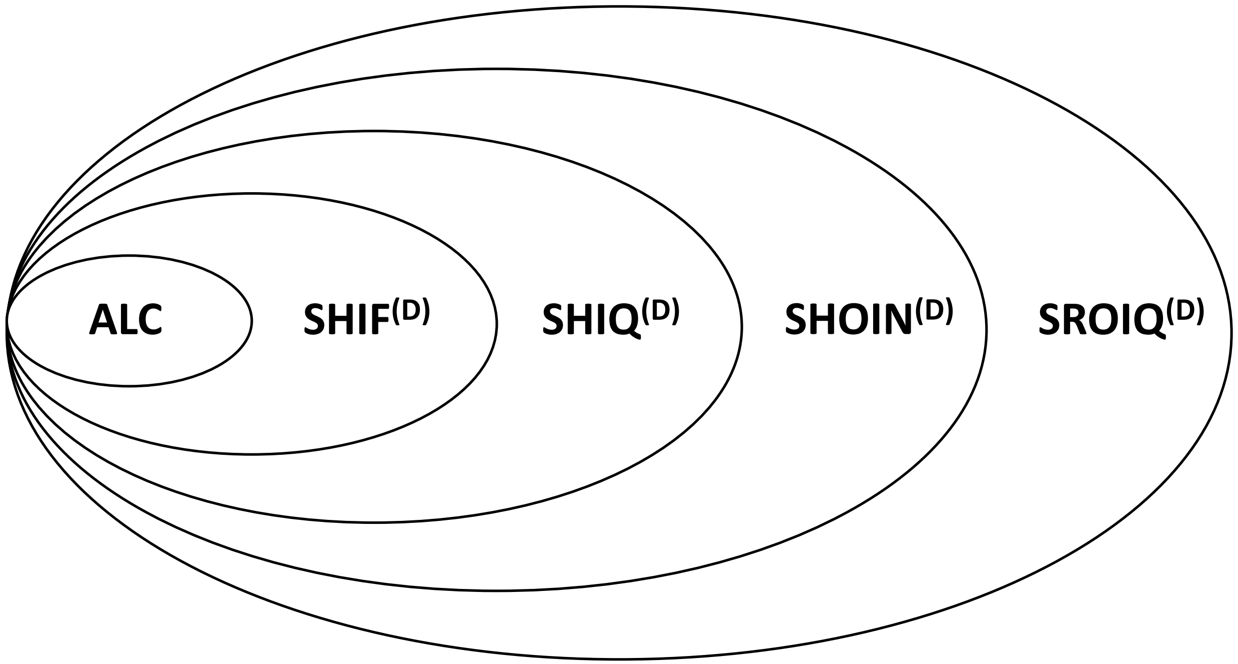
Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C)   Roles (R)   Individuals (I)   Literals (L)

DL-based Ontologies have two parts

Based on Constructors used in the TBox, DLs can be classified into:



Knowledge Graph
(ABox)

$G(V,E)$ $V = I \cup C \cup L$ $E = R$

Directed Graph

Schema
(TBox)

Concept Inclusion ($C \sqsubseteq D$)
Concept Equality ($C \equiv D$)
Role Hierarchy ($R \sqsubseteq S$)
Role Transitivity (Tran(R)) etc.

Why $SHIQ^{(D)}$?

Fundamentals of DL Ontology

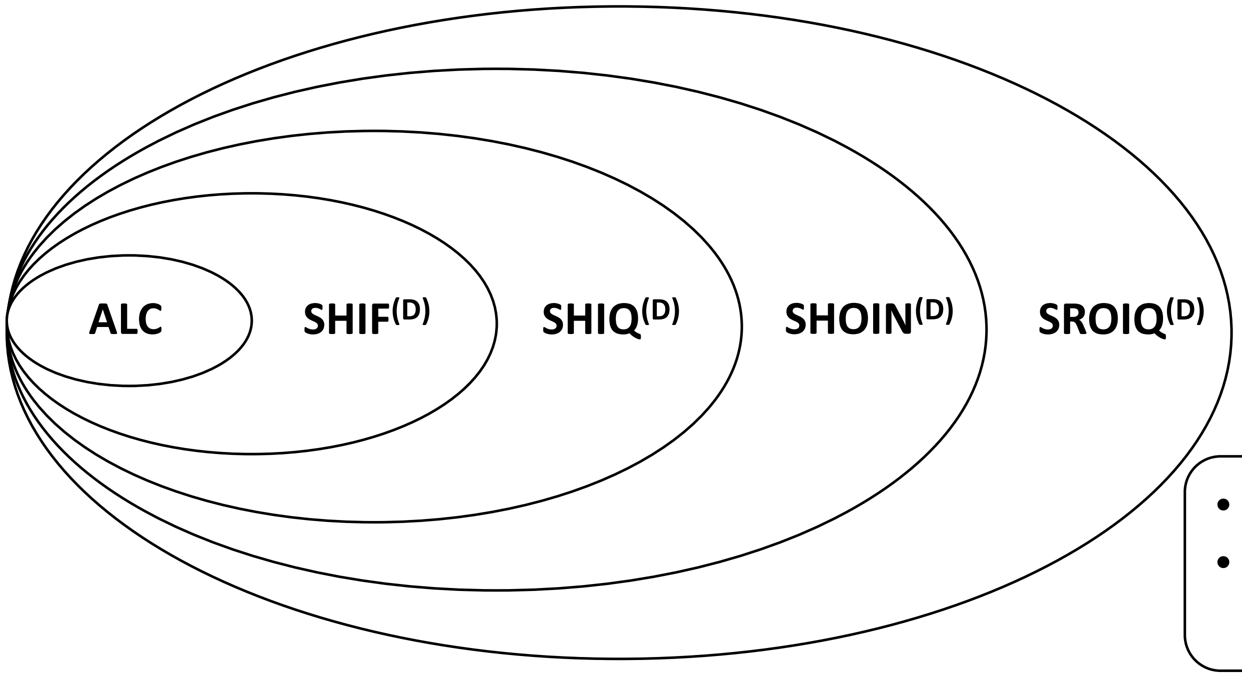
Description Logics (DLs) are decidable fragments of first order logic

DL-based Ontologies have basic building blocks:

Classes (C)   Roles (R)   Individuals (I)   Literals (L)

DL-based Ontologies have two parts

Based on Constructors used in the TBox, DLs can be classified into:



Knowledge Graph
(ABox)

$G(V,E)$ $V = I \cup C \cup L$ $E = R$

Directed Graph

Schema
(TBox)

Concept Inclusion ($C \sqsubseteq D$)
Concept Equality ($C \equiv D$)
Role Hierarchy ($R \sqsubseteq S$)
Role Transitivity (Tran(R)) etc.

- It is a highly expressive DL
- It is implemented in fast reasoning systems such as FaCT and RACER [IJCAI '01]

Fundamentals of DL Ontology

SHIQ is an extension of ALC DL

S	**H**	**I**	**Q**
ALC with Transitive roles	Role Hierarchy	Inverse Role	Qualified Number Restriction

Fundamentals of DL Ontology

SHIQ is an extension of ALC DL

S	**H**	**I**	**Q**
ALC with Transitive roles	Role Hierarchy	Inverse Role	Qualified Number Restriction

Name	**Syntax**	**Semantics**
atomic concept	A	$A^{\mathcal{I}}$
top concept	$\top$	$\Delta^{\mathcal{I}}$
bottom concept	$\bot$	ϕ
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
existential restriction	$\exists R.C$	$\{ x \in \Delta^{\mathcal{I}} \mid \exists y .\ \langle x, y \rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}} \}$
universal restriction	$\forall R.C$	$\{ x \in \Delta^{\mathcal{I}} \mid \forall y .\ \langle x, y \rangle \in R^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}} \}$

Fundamentals of DL Ontology

SHIQ is an extension of ALC DL

S	**H**	**I**	**Q**
ALC with Transitive roles	Role Hierarchy	Inverse Role	Qualified Number Restriction

Name	**Syntax**	**Semantics**
atomic concept	A	$A^{\mathcal{I}}$
top concept	$\top$	$\Delta^{\mathcal{I}}$
bottom concept	$\bot$	ϕ
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
existential restriction	$\exists R.C$	$\{ x \in \Delta^{\mathcal{I}} \mid \exists y . \langle x, y \rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}} \}$
universal restriction	$\forall R.C$	$\{ x \in \Delta^{\mathcal{I}} \mid \forall y . \langle x, y \rangle \in R^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}} \}$
min cardinality*	$\geq nR.C$	$\{ x \in \Delta^{\mathcal{I}} \mid \# \{ y . \langle x, y \rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}} \} \geq n \}$
max cardinality*	$\leq mR.C$	$\{ x \in \Delta^{\mathcal{I}} \mid \# \{ y . \langle x, y \rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}} \} \leq m \}$

*Roles R of ≥ nR.C and ≤ mR.C should be simple roles (not a composition of other roles)

Motivation

Framing competent question is **challenging** and **time consuming task**

Question Generation from existing ontologies for educational & professional applications:

- Test generation to **measure the learning success** of students
- Question generation systems can be deployed as **Chatbot components**
- **Training of employees** about products, customers or an organization
- **Fraud detection** in crowd-sourcing platforms
- Question-driven **knowledge learning** is less explored

Existing Approaches

Knowledge Representation	System Name	Method	Question-type
Raw Text	**EVALING** [AENLP ’99]	Using linguistic patterns	WH-questions
Raw Text	[EdAppsNLP ‘05, AIED ‘09]	NLP, Sentence Transformation	Language Learning
Raw Text	[EMNLP ’05]	Using WordNet	Lexical Relationships
Raw Text	[ACL ‘17]	Neural sequence learning based models	Reading comprehension materials
KG + Raw Text	[ACL ‘16]	Neural machine translation models	Triple-based, Question-answer corpus for training
Ontology (Knowledge Graphs)	**ASSESS** [ISWC ‘15], [WWW ’15] **Sherlock** [CEUR ‘14]	SPARQL template	Triple-based questions
Ontology (Schema)	[OWLED ’14]	Using Class Hierarchy	Analogy type

Existing Approaches

Knowledge Representation	System Name	Method	Question-type
Raw Text	**EVALING** [AENLP ’99]	Using linguistic patterns	WH-questions
Raw Text	[EdAppsNLP ‘05, AIED ‘09]	NLP, Sentence Transformation	Language Learning
Raw Text	[EMNLP ’05]	Using WordNet	Lexical Relationships
Raw Text	[ACL ‘17]	Neural sequence learning based models	Reading comprehension materials
KG + Raw Text	[ACL ‘16]	Neural machine translation models	Triple-based, Question-answer corpus for training
Ontology (Knowledge Graphs)	**ASSESS** [ISWC ‘15], [WWW ’15] **Sherlock** [CEUR ‘14]	SPARQL template	Triple-based questions
Ontology (Schema)	[OWLED ’14]	Using Class Hierarchy	Analogy type

Our approaches also focus on **Knowledge Graphs** & **Schema**

We could generate questions that were close to human generated ones, considering **Patterns** & **Role Restrictions**

Thesis Statement

Utilizing a well structured **DL Ontology** is highly effective in **generating meaningful questions** (mainly, multiple-choice questions).

It provides ease in **automating various functionalities** of the applications (especially, educational applications) that employ the generated questions.

Thesis Overview

Question Generation from Knowledge Graphs

Question Generation from Schema using Semantic-Refinement

Applications of Semantic-Refinement

Thesis Overview

Question Generation from Knowledge Graphs

1. To **reduce manual intervention** in the **Question Authoring** Module of the existing e-Learning systems
2. To develop system(s) that uses the **Patterns in the Knowledge Graph** to frame Factual Questions
3. To develop heuristics to select the most **relevant set of questions** for conducting domain related assessments
4. To measure the **difficulty level of question** statements for learners with various proficiency levels

Question Generation from Schema using Semantic-Refinement

Applications of Semantic-Refinement

Thesis Overview

Question Generation from Knowledge Graphs

1. To reduce manual intervention in the Question Authoring Module of the existing e-Learning systems
2. To develop system(s) that uses the Patterns in the Knowledge Graph to frame Factual Questions
3. To develop heuristics to select the most relevant set of questions for conducting domain related assessments
4. To measure the difficulty level of question statements for learners with various proficiency levels

Question Generation from Schema using Semantic-Refinement

1. To frame competent question statements (Stems) by **making use of Role Restrictions** in a DL-based Ontology
2. To generate **valid incorrect answers** (Distractors) for the framed stems
3. To remove **Redundant Restrictions** from Question Statements using **Semantic-Refinement**

Applications of Semantic-Refinement

Thesis Overview

Question Generation from Knowledge Graphs

1. To reduce manual intervention in the Question Authoring Module of the existing e-Learning systems
2. To develop system(s) that uses the Patterns in the Knowledge Graph to frame Factual Questions
3. To develop heuristics to select the most relevant set of questions for conducting domain related assessments
4. To measure the difficulty level of question statements for learners with various proficiency levels

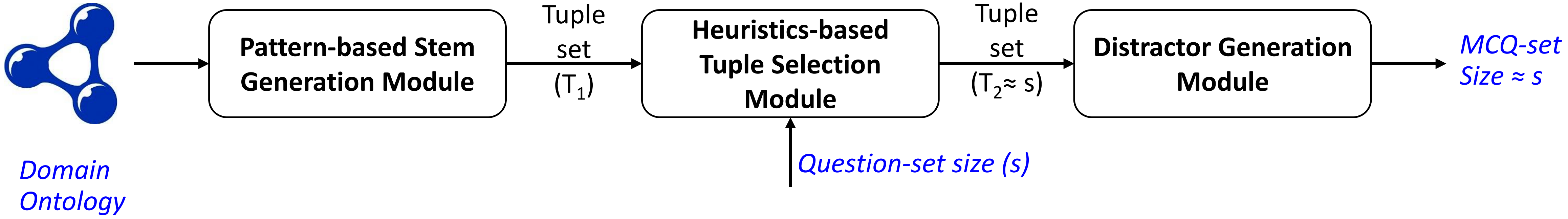
Question Generation from Schema using Semantic-Refinement

1. To frame competent question statements (Stems) by making use of Role Restrictions in a DL-based Ontology
2. To generate valid incorrect answers (Distractors) for the framed stems
3. To remove Redundant Restrictions from Question Statements using Semantic-Refinement

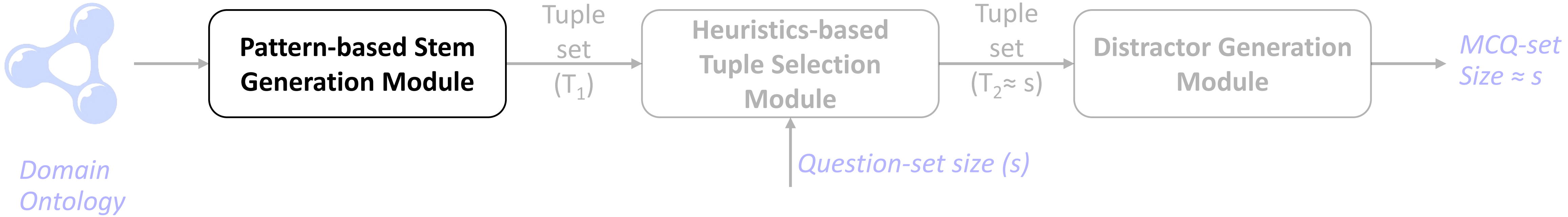
Applications of Semantic-Refinement

1. To generate **refined descriptions** of ontology entities for knowledge validation

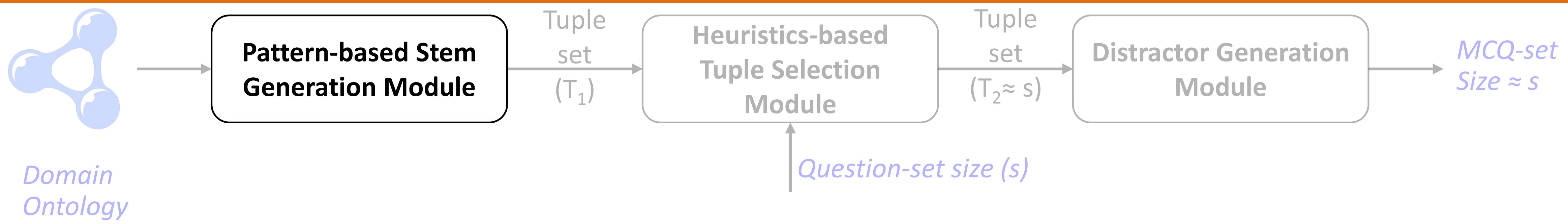
Proposed System – QG from Knowledge Graph: ATG System

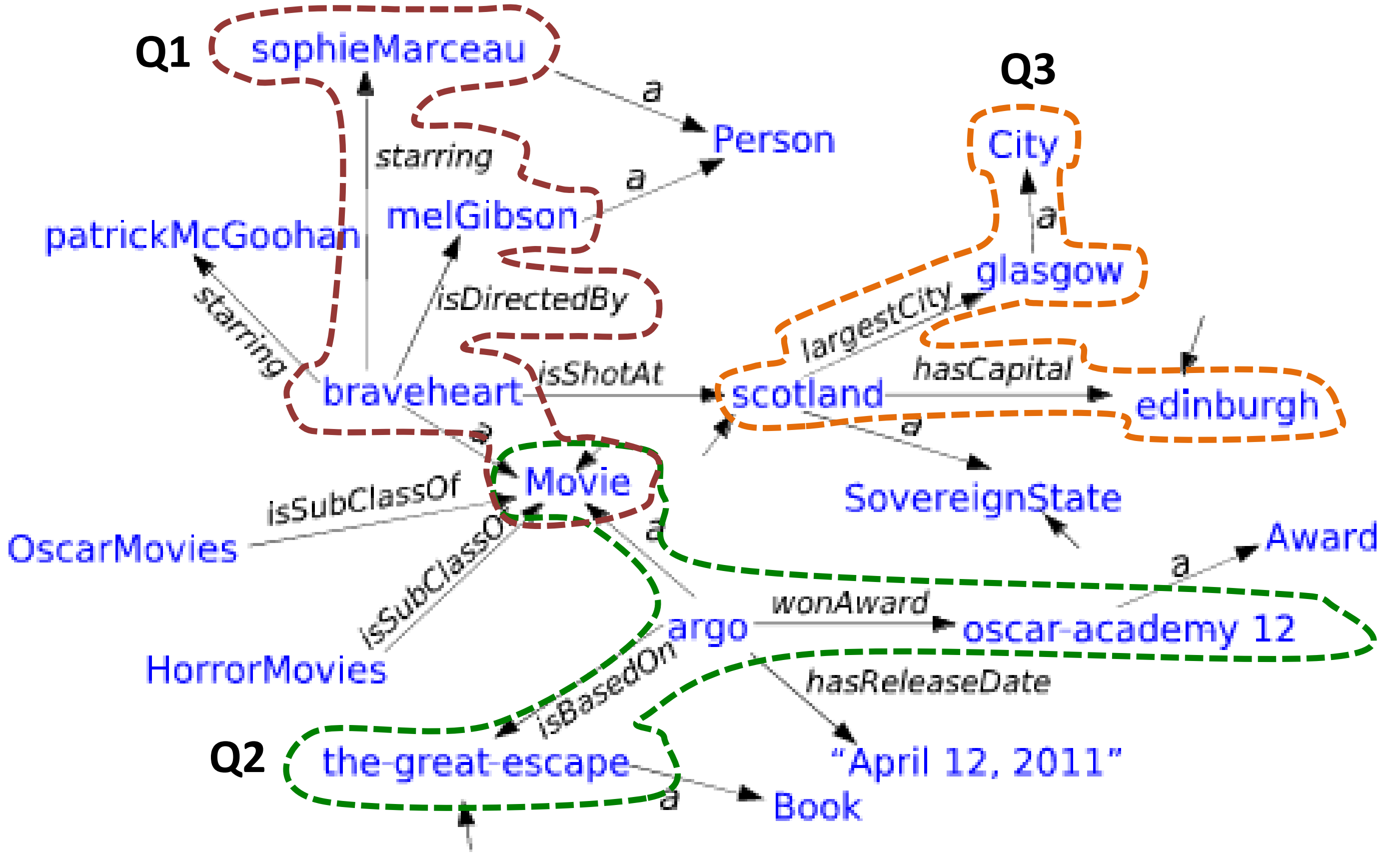


Proposed System – QG from Knowledge Graph: ATG System



Proposed System – QG from Knowledge Graph: ATG System



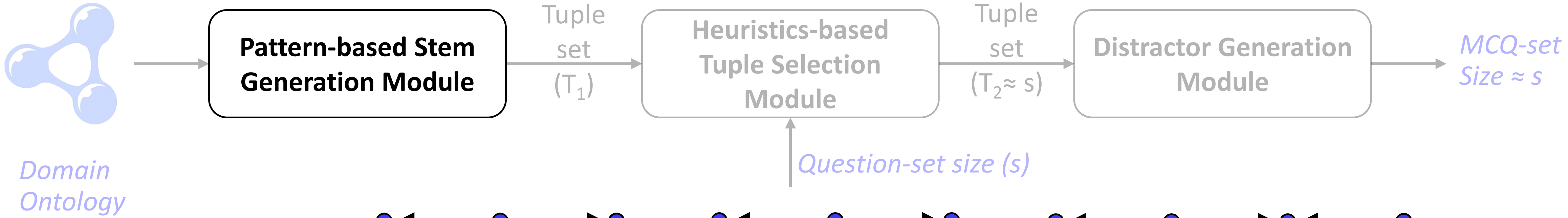


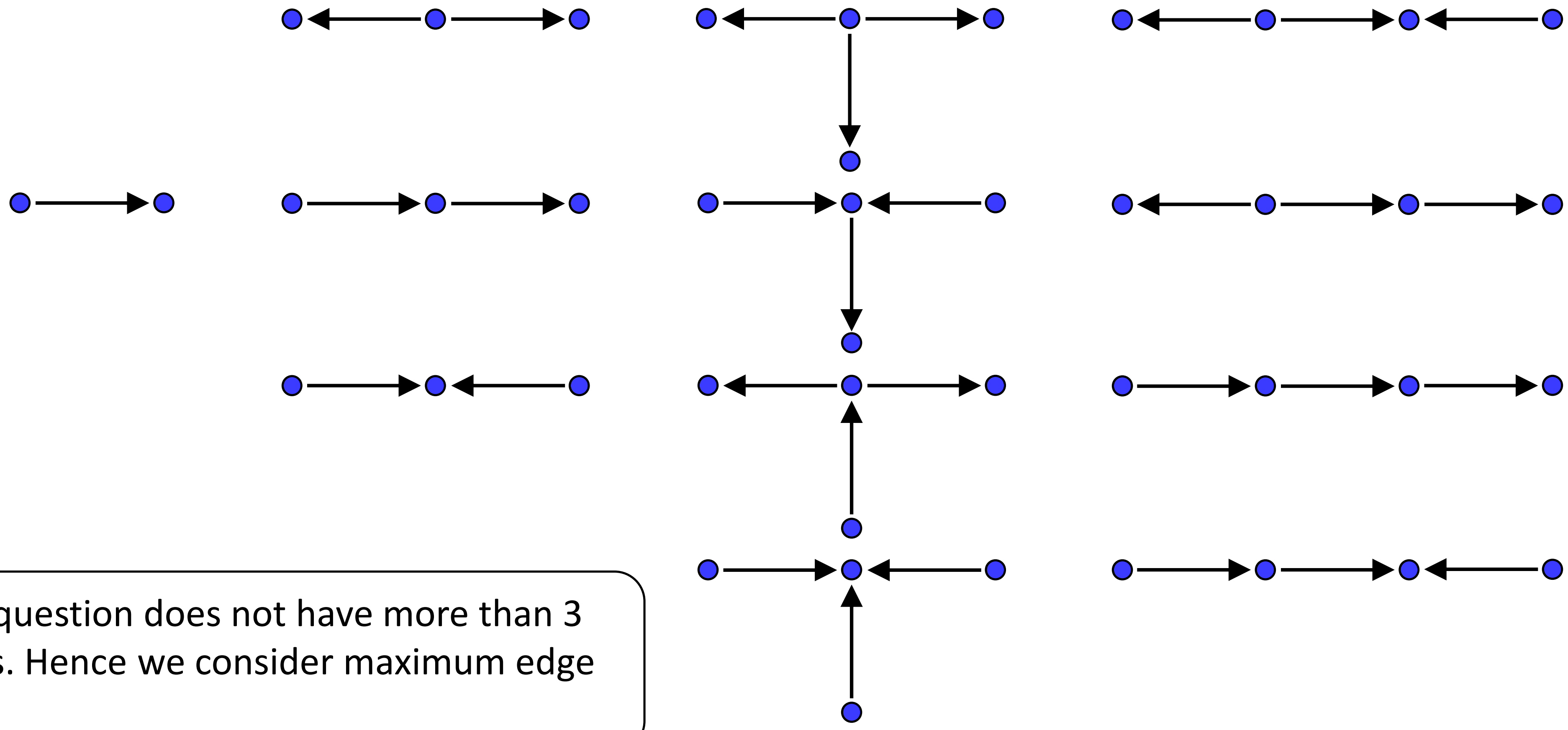
Q1. Choose the movie that is directed by Mel Gibson and starring Sophie Marceau.

Q2. Choose the movie that won Oscar award and was based on the great escape.

Q3. Choose the country that has capital Edinburgh and has Glasgow as the largest city.

Proposed System – QG from Knowledge Graph: ATG System

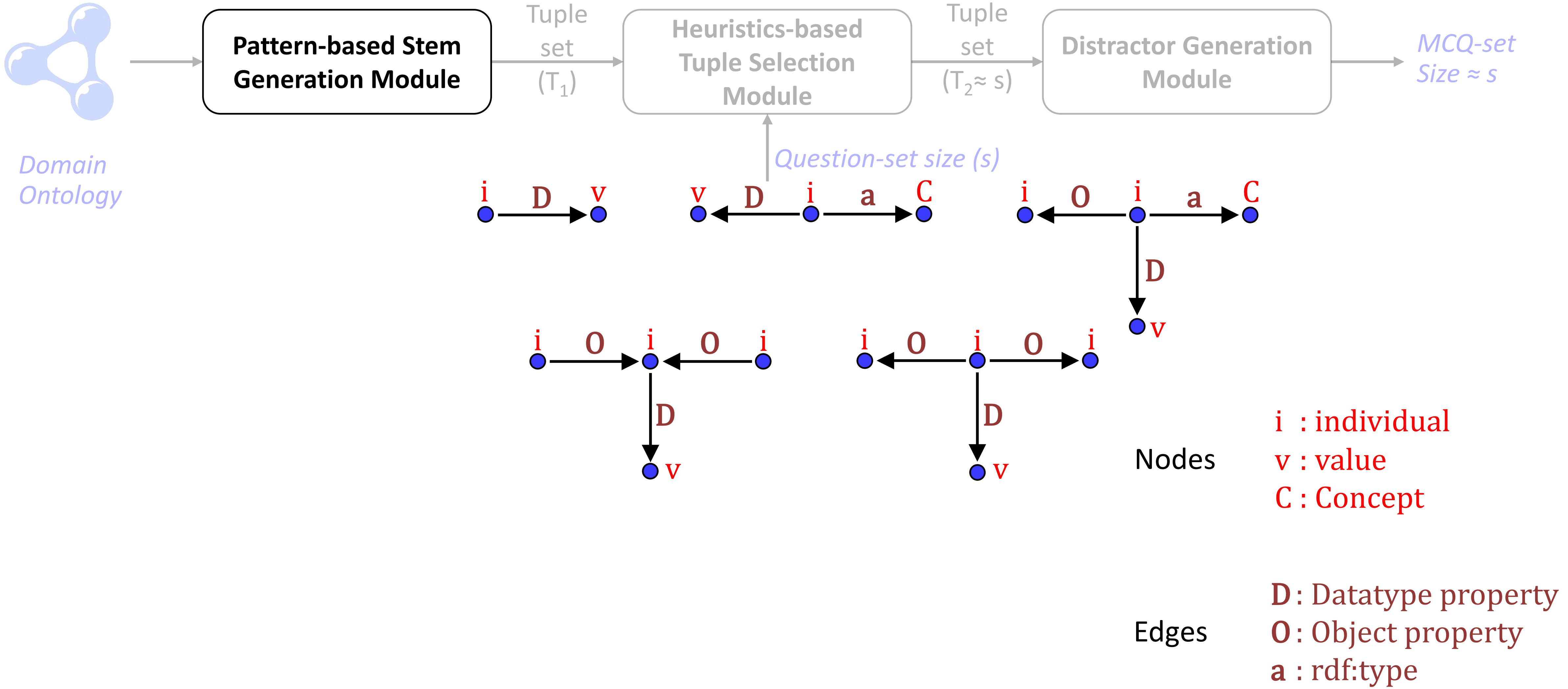




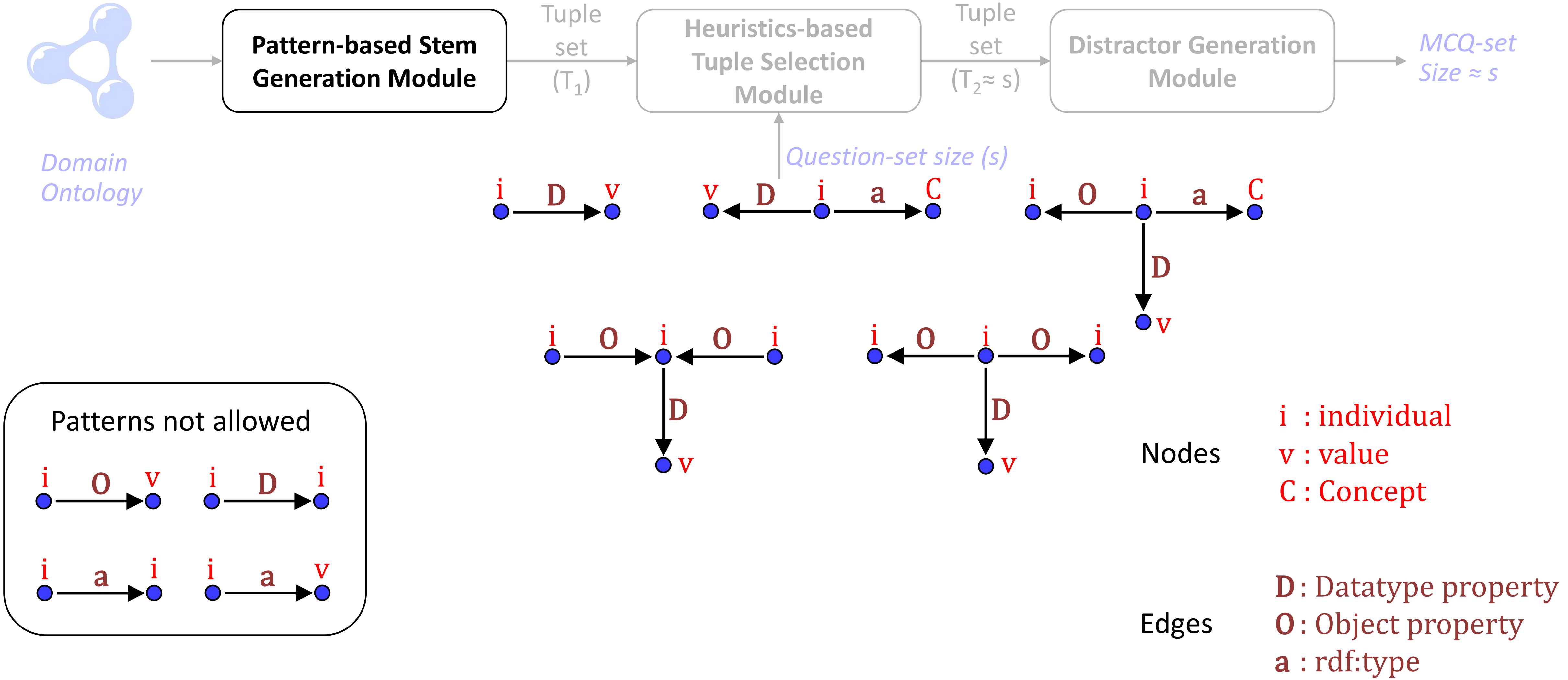
Usually a question does not have more than 3 predicates. Hence we consider maximum edge size of 3.

Considering the directionality of the edges alone, we can have 12 patterns

Proposed System – QG from Knowledge Graph: ATG System

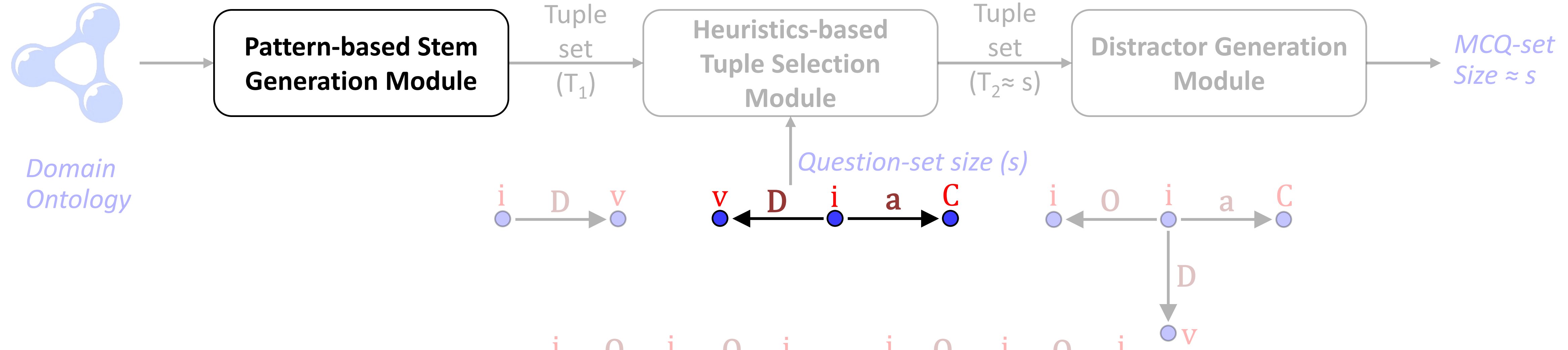


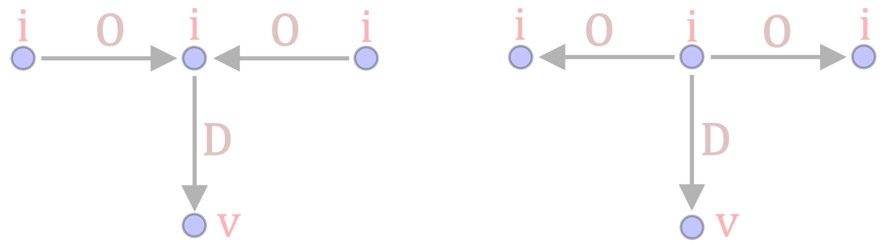
Proposed System – QG from Knowledge Graph: ATG System



With all the allowed combinations, **40 patterns** are possible

Proposed System – QG from Knowledge Graph: ATG System



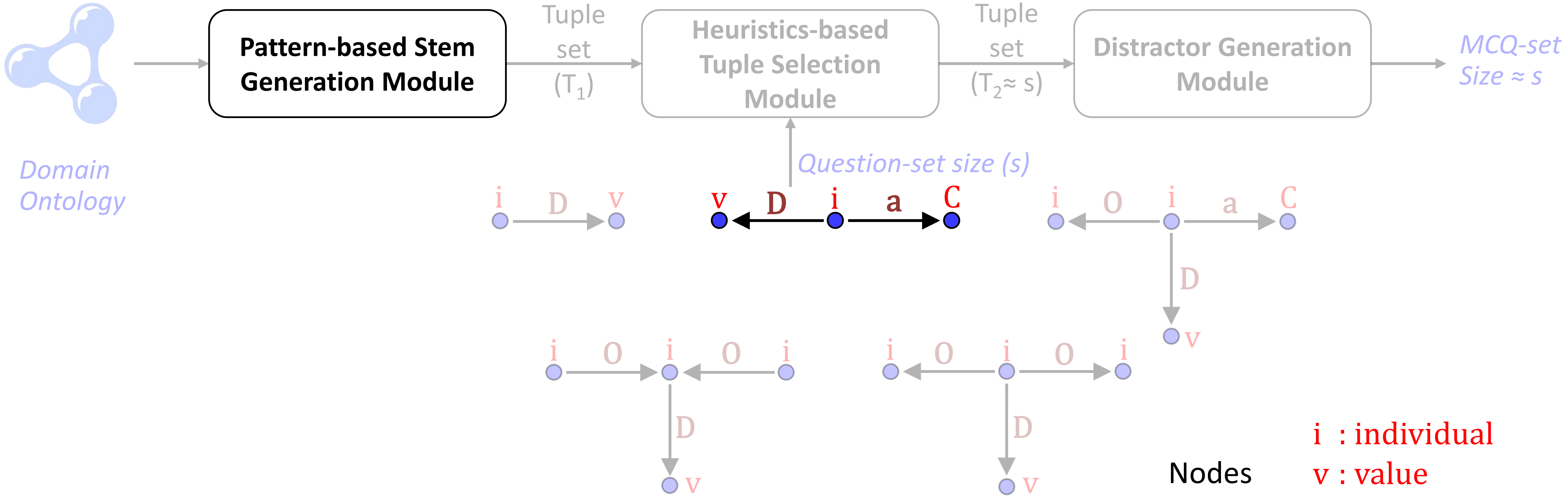


Nodes
- i : individual
- v : value
- C : Concept

Pattern: $C \; \overleftarrow{a} \; i \; \overrightarrow{D} \; v$

Edges
- **D** : Datatype property
- **O** : Object property
- **a** : rdf:type

Proposed System – QG from Knowledge Graph: ATG System



Nodes
- i : individual
- v : value
- C : Concept

Edges
- **D** : Datatype property
- **O** : Object property
- **a** : rdf:type

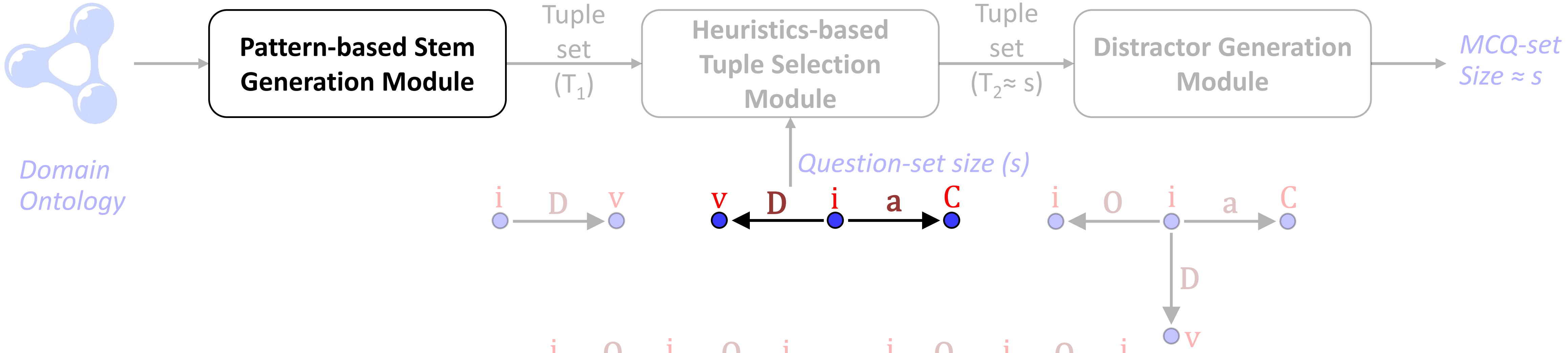
Pattern: $C\ \overleftarrow{a}\ i\ \overrightarrow{D}\ v$

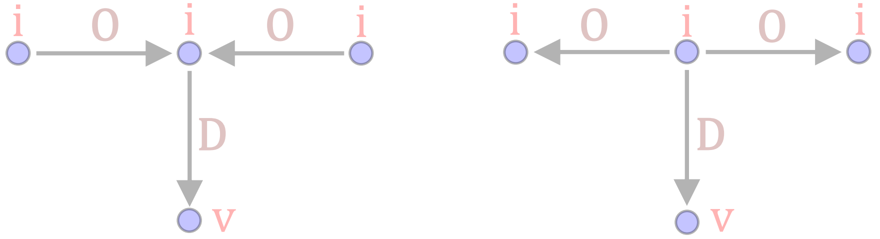
Stem-template: Choose a [C] with [D] [v]

SPARQL-template:

```
SELECT ?C ?i ?D ?v
WHERE {?i a ?C . ?i ?O ?v . ?O a owl:DatatypeProperty.}
```

Proposed System – QG from Knowledge Graph: ATG System





Nodes
- i : individual
- v : value
- C : Concept

Edges
- D : Datatype property
- O : Object property
- a : rdf:type

Pattern: C $\overleftarrow{a}$ i $\overrightarrow{D}$ v

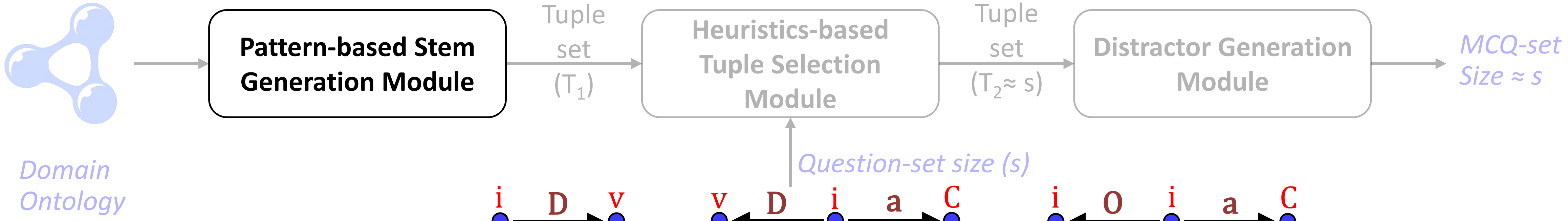
Stem-template: Choose a [C] with [D] [v]

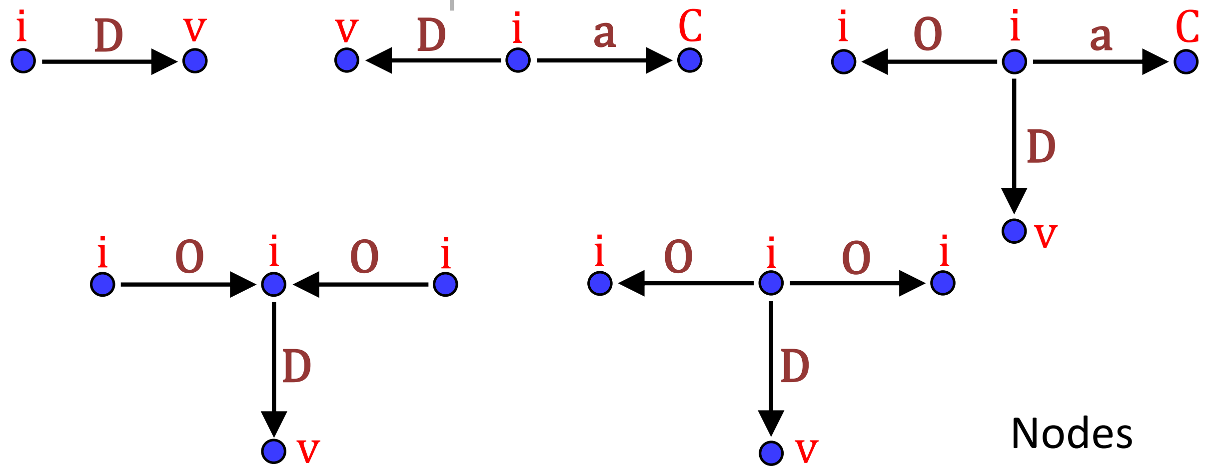
SPARQL-template:

```
SELECT ?C ?i ?D ?v
WHERE {?i a ?C . ?i ?O ?v . ?O a owl:DatatypeProperty.}
```

Example: *Choose a Movie with release-date “May 21st 2011”.*

Proposed System – QG from Knowledge Graph: ATG System





Nodes
- i : individual
- v : value
- C : Concept

Edges
- D : Datatype property
- O : Object property
- a : rdf:type

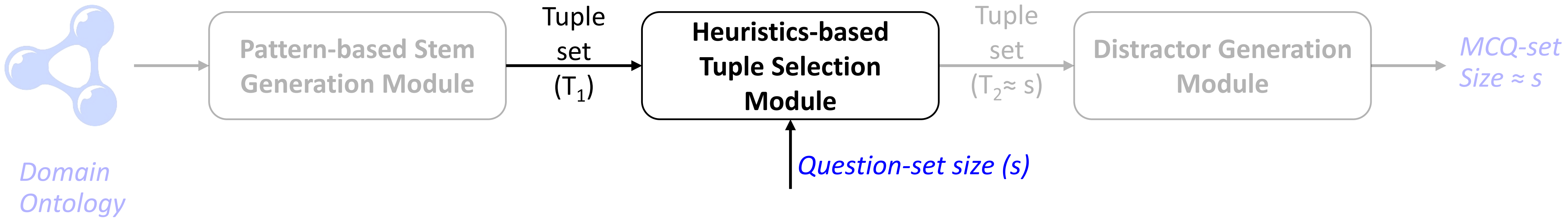
Significance of studying these patterns

These patterns could be used to identify the possible set of answers known as the **Potential-Set**

Pattern: $C\ \overleftarrow{a}\ i\ \overrightarrow{D}\ v$

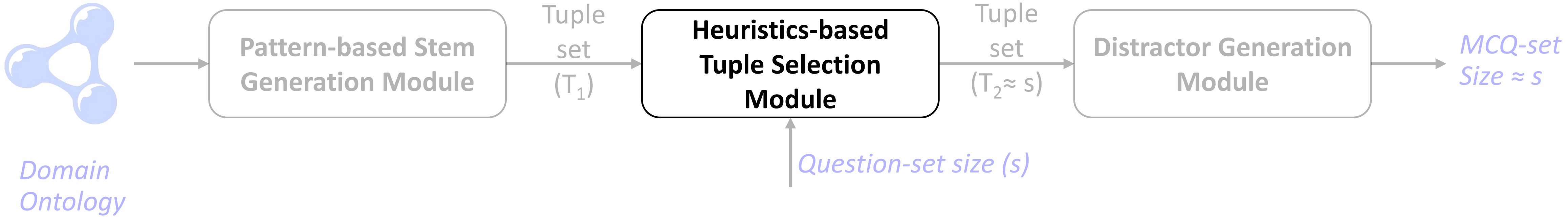
Potential-Set: $C \sqcap \text{Domain}(D)$

Proposed System – QG from Knowledge Graph: ATG System



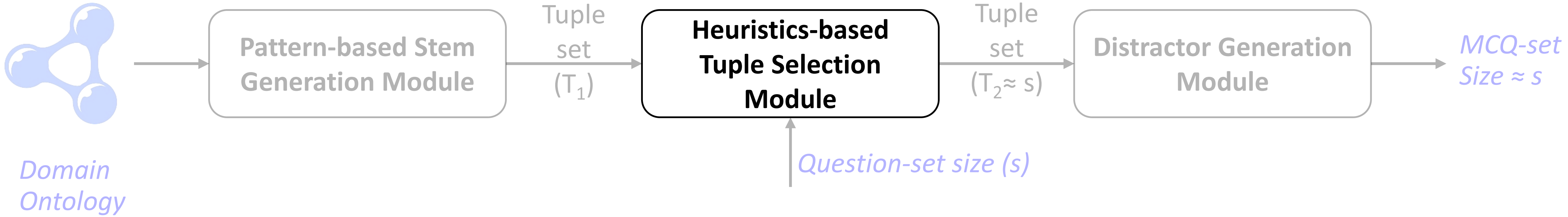
Screening: Property-based
Concept-based
Similarity-based

Proposed System – QG from Knowledge Graph: ATG System



Property-based: We assign low triviality score (PSTS) to rare role-combination.

Proposed System – QG from Knowledge Graph: ATG System

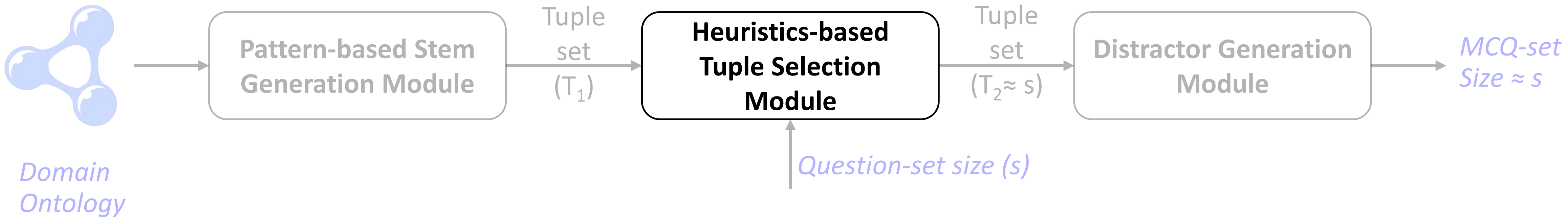


Property-based: We assign low triviality score (PSTS) to rare role-combination.

low triviality score ⇒ high significance

$$\text{Property Sequence Triviality Score (t)} = \frac{\text{No. of individuals satisfying all the properties in t}}{\text{No. of individuals in the Potential-Set of t}}$$

Proposed System – QG from Knowledge Graph: ATG System



Property-based: We assign low triviality score (PSTS) to rare role-combination.

low triviality score ⇒ high significance

$$\text{Property Sequence Triviality Score (t)} = \frac{\text{No. of individuals satisfying all the properties in t}}{\text{No. of individuals in the Potential-Set of t}}$$

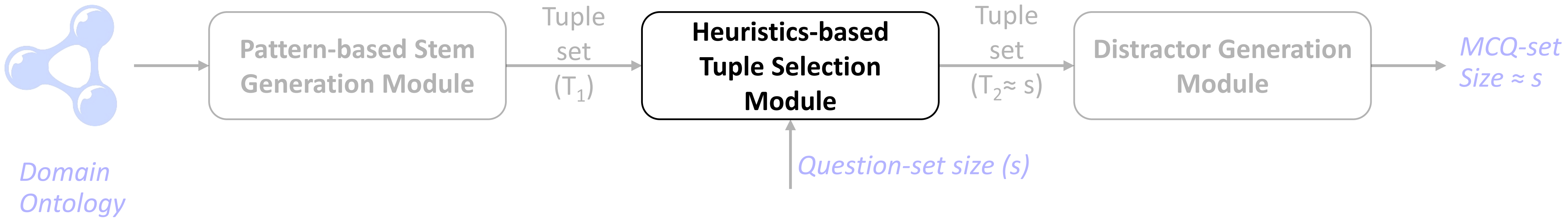
Example

"Choose a movie which is *based-on* "the great escape", and *won-award* "Oscar'12".

"Choose a movie which is *directed-by* Clint-Eastwood and *starring* Hilary Swank.

PSTS({*won-award, based-on*}) << PSTS({*is-directed-by, starring*})

Proposed System – QG from Knowledge Graph: ATG System

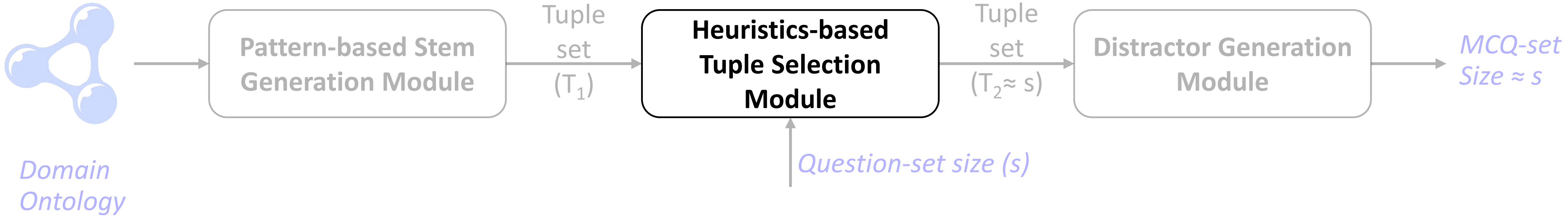


Property-based: We assign low triviality score (PSTS) to rare role-combination.

Concept-based: We select only those tuples/questions whose individual belongs to the relevant concepts in the domain

Key Concept Extraction – API [SW ‘08]

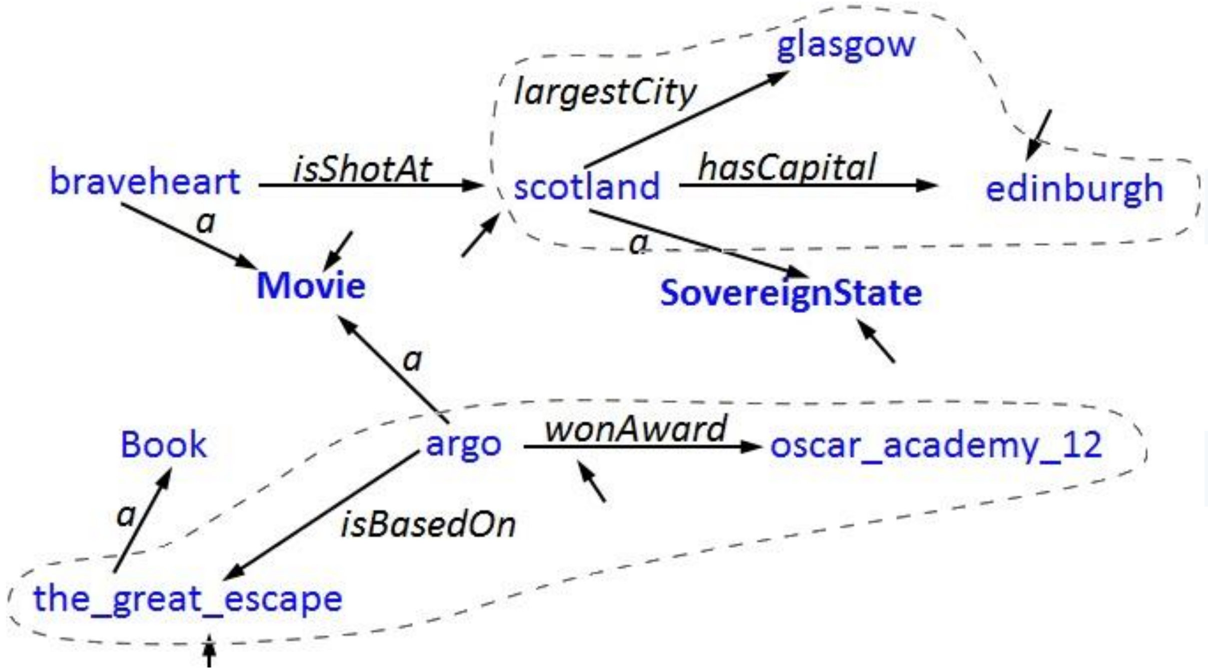
Proposed System – QG from Knowledge Graph: ATG System



Property-based: We assign low triviality score (PSTS) to rare role-combination.

Concept-based: We select only those tuples/questions whose individual belongs to the relevant concepts in the domain

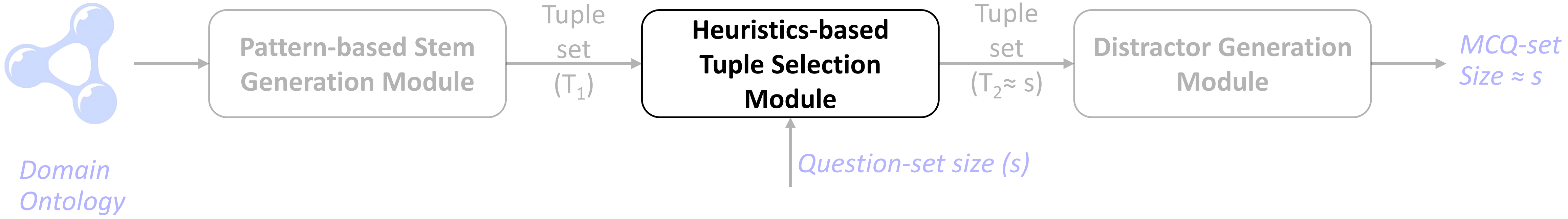
Key Concept Extraction – API [SW '08]



Choose the Sovereign state with capital Edinburgh and having largest city Glasgow.

Choose a movie which is based-on "the great escape", and won-award "Oscar'12".

Proposed System – QG from Knowledge Graph: ATG System



Property-based: We assign low triviality score (PSTS) to rare role-combination.

Concept-based: We select only those tuples/questions whose individual belongs to the relevant concepts in the domain

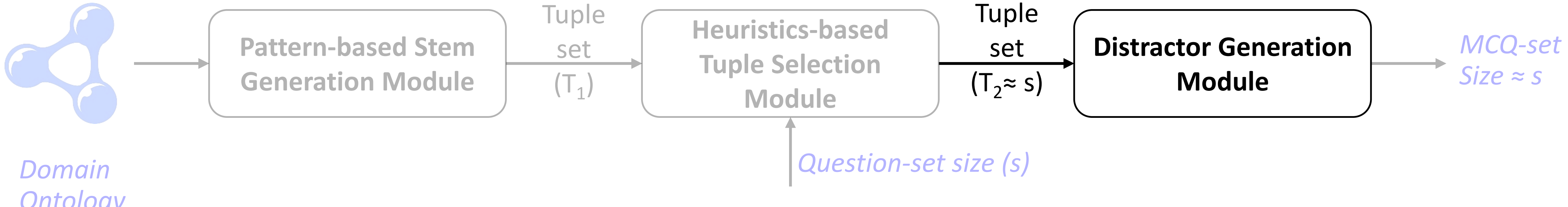
Similarity-based: We group the tuples based on their semantic similarity and select a representative tuple from each group.

$$\mathrm{SemSimilarity}(t_1, t_2) = \frac{1}{2}\left(\frac{|\,\mathrm{Ind}(\mathrm{P}(t_1)) \cap \mathrm{Ind}(\mathrm{P}(t_2))\,|}{|\,\mathrm{Ind}(\mathrm{P}(t_1)) \cup \mathrm{Ind}(\mathrm{P}(t_2))\,|} + \frac{\text{No. of semantically similar triples in } t_1 \text{ and } t_2}{\mathrm{Max}(\text{no. of triples in } t_1, \text{no. of triples in } t_2)}\right)$$

P(t): Properties in the tuple t

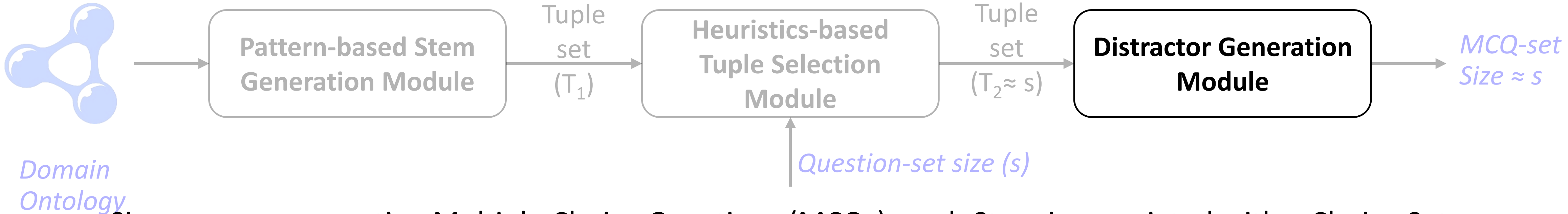
Ind(P(t)): Individuals which satisfy the properties in P(t)

Proposed System – QG from Knowledge Graph: ATG System

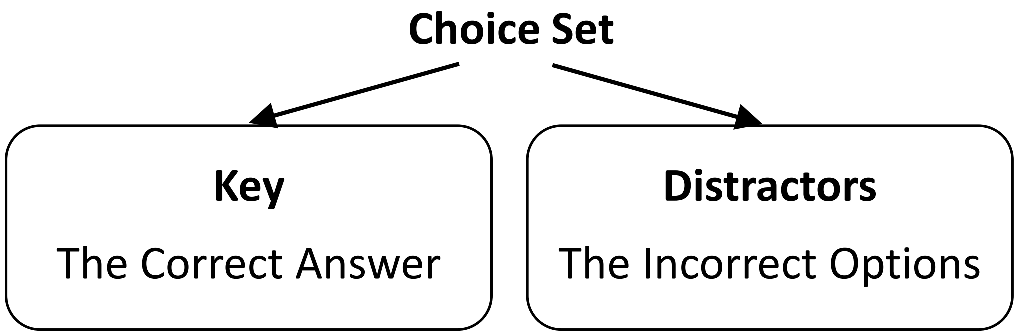


Since we are generating Multiple Choice Questions (MCQs), each Stem is associated with a Choice Set.

Proposed System – QG from Knowledge Graph: ATG System



Since we are generating Multiple Choice Questions (MCQs), each Stem is associated with a Choice Set.

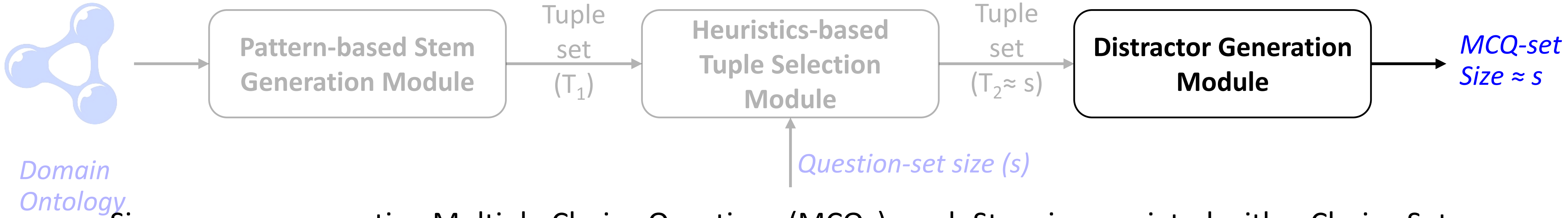


The **Key** can be obtained directly from the tuple

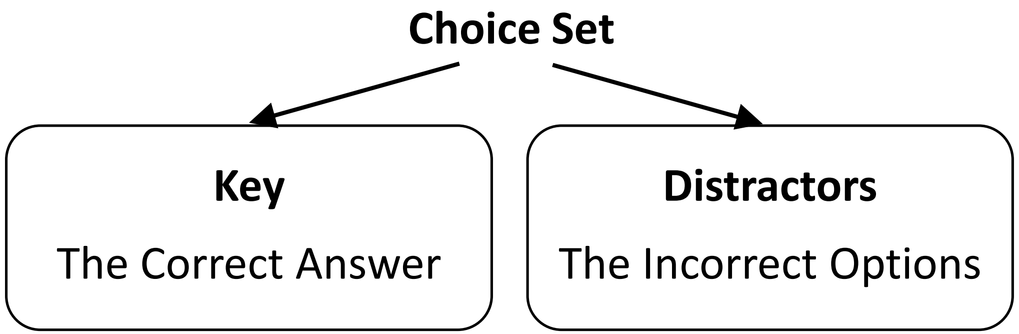
Good **distractors** can be selected using the Potential-Set of the tuple

$$\text{Distractors} = \text{Potential-Set} - \text{Correct Answer}$$

Proposed System – QG from Knowledge Graph: ATG System



Since we are generating Multiple Choice Questions (MCQs), each Stem is associated with a Choice Set.



The **Key** can be obtained directly from the tuple

Good **distractors** can be selected using the Potential-Set of the tuple

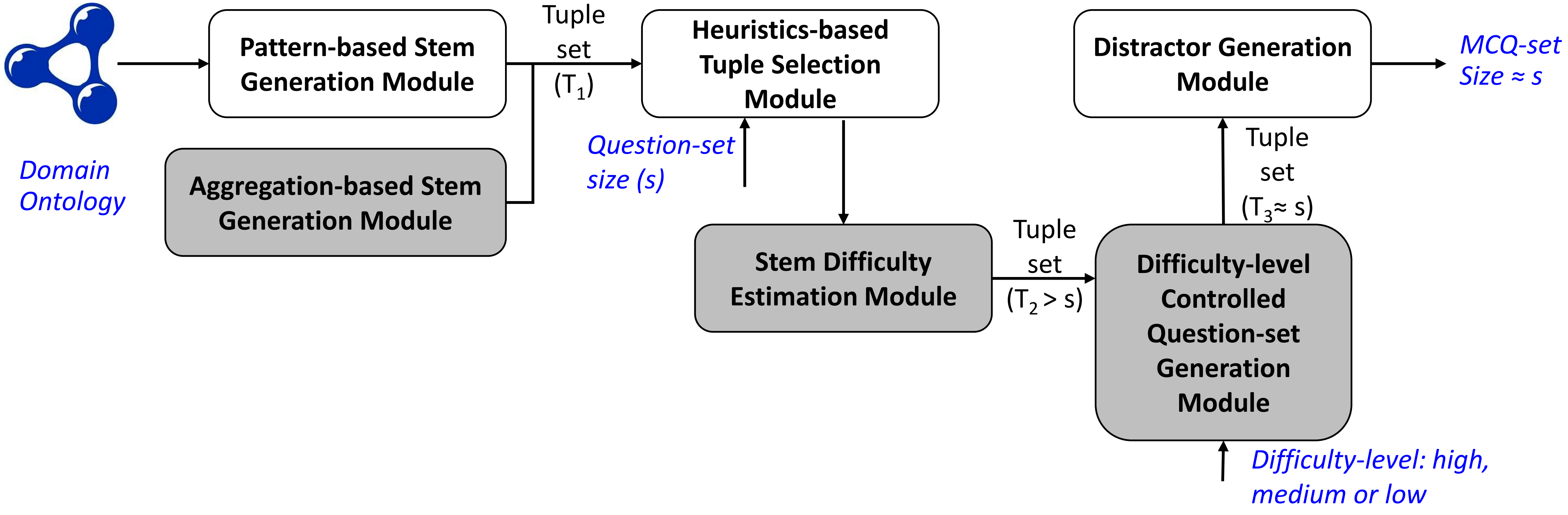
$$\text{Distractors} = \text{Potential-Set} - \text{Correct Answer}$$

Also, the difficulty-level of the question depends on the Distractors.

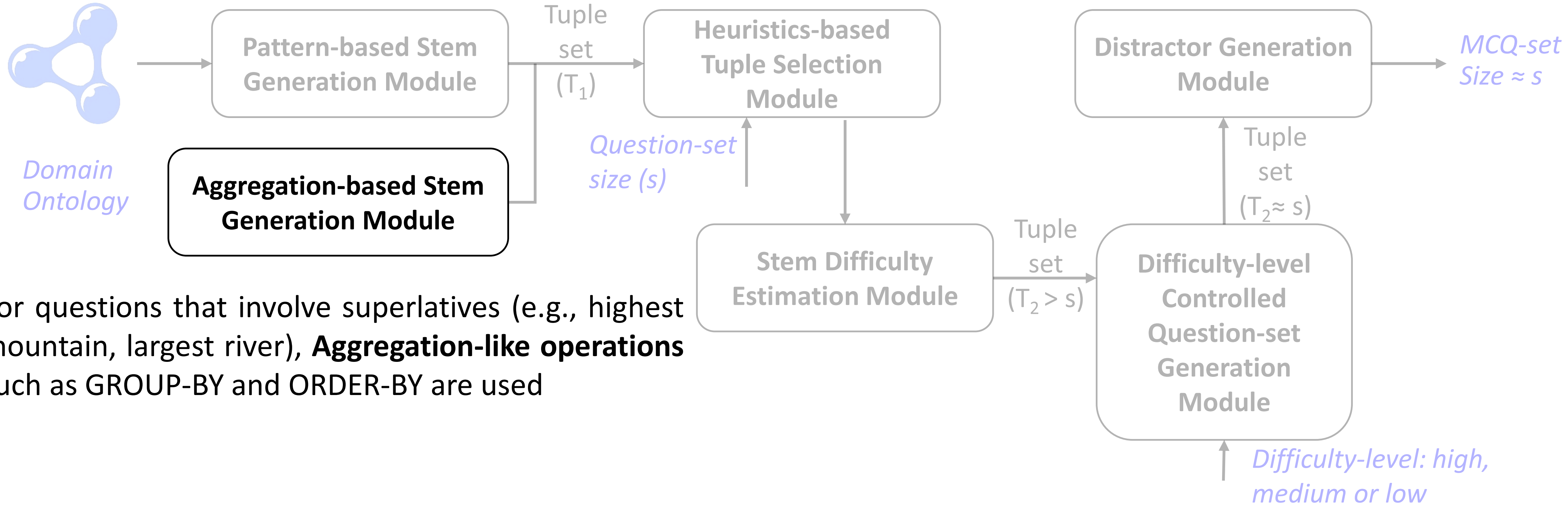
Distractors from Potential-Set $\Rightarrow$ High Difficult MCQs

Distractors from outside Potential-Set $\Rightarrow$ Low Difficult MCQs

Proposed System – QG from Knowledge Graph: Extended-ATG System

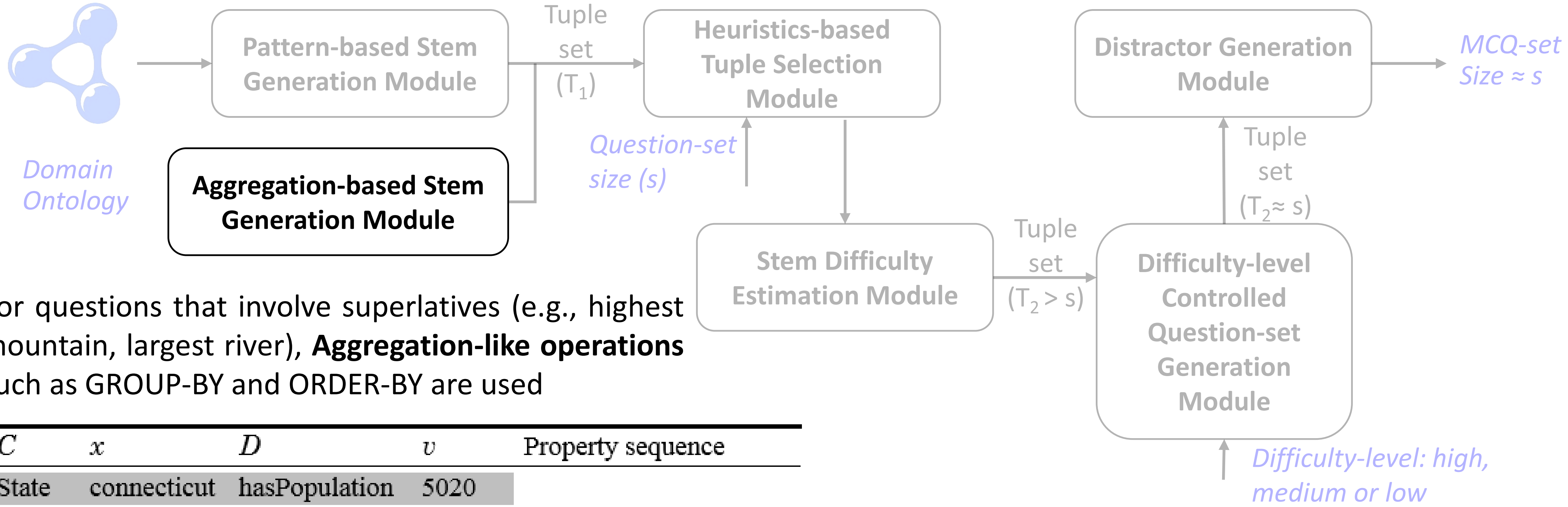


Proposed System – QG from Knowledge Graph: Extended-ATG System



For questions that involve superlatives (e.g., highest mountain, largest river), **Aggregation-like operations** such as GROUP-BY and ORDER-BY are used

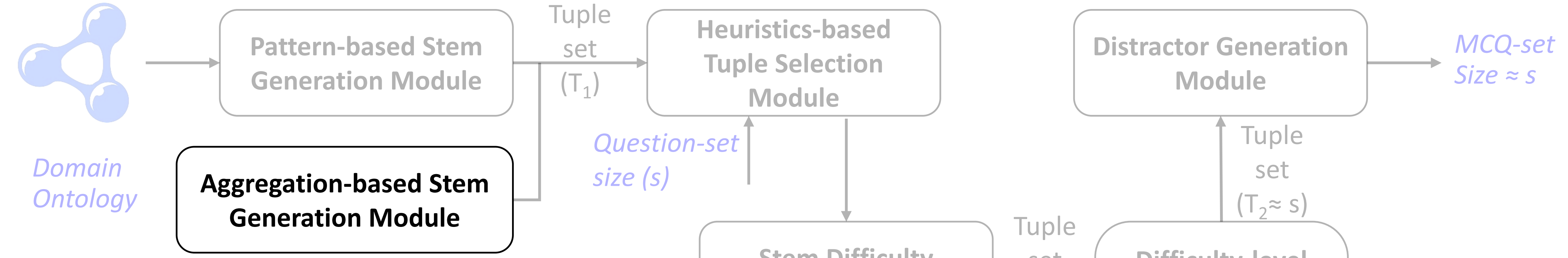
Proposed System – QG from Knowledge Graph: Extended-ATG System

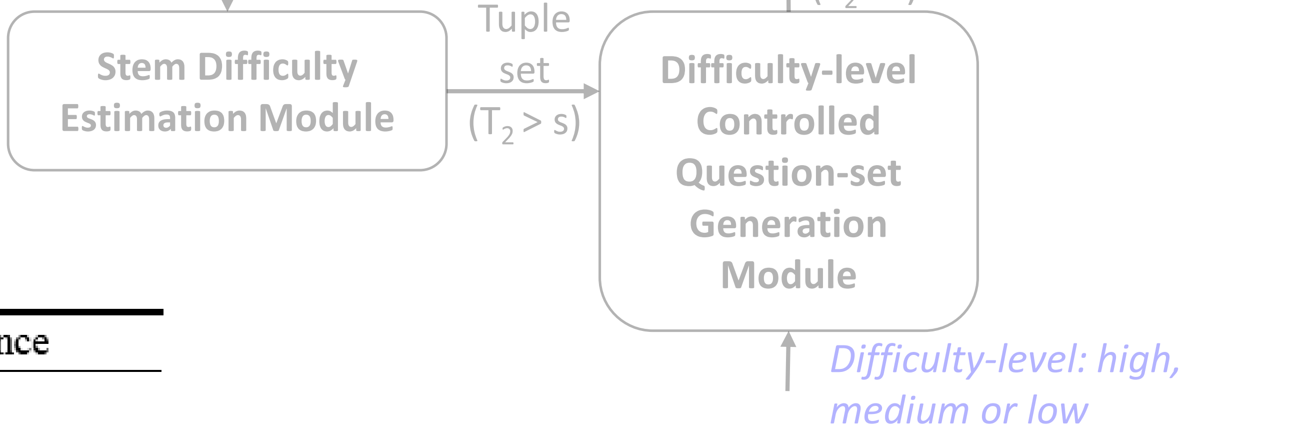


For questions that involve superlatives (e.g., highest mountain, largest river), **Aggregation-like operations** such as GROUP-BY and ORDER-BY are used

C	x	D	v	Property sequence
State	connecticut	hasPopulation	5020	
State	florida	hasPopulation	68664	
State	colorado	hasPopulation	104000	{State, hasPopulation}
…	…	…	…	
State	arizona	hasPopulation	114000	
River	neosho	length	740	
River	wasbash	length	764	
River	pecos	length	805	{River, length}
…	…	…	…	
River	ouachita	length	973	

Proposed System – QG from Knowledge Graph: Extended-ATG System





For questions that involve superlatives (e.g., highest mountain, largest river), **Aggregation-like operations** such as GROUP-BY and ORDER-BY are used

C	x	D	v	Property sequence
State	connecticut	hasPopulation	5020	
State	florida	hasPopulation	68664	
State	colorado	hasPopulation	104000	{State, hasPopulation}
…	…	…	…	
State	arizona	hasPopulation	114000	
River	neosho	length	740	
River	wasbash	length	764	
River	pecos	length	805	{River, length}
…	…	…	…	
River	ouachita	length	973	

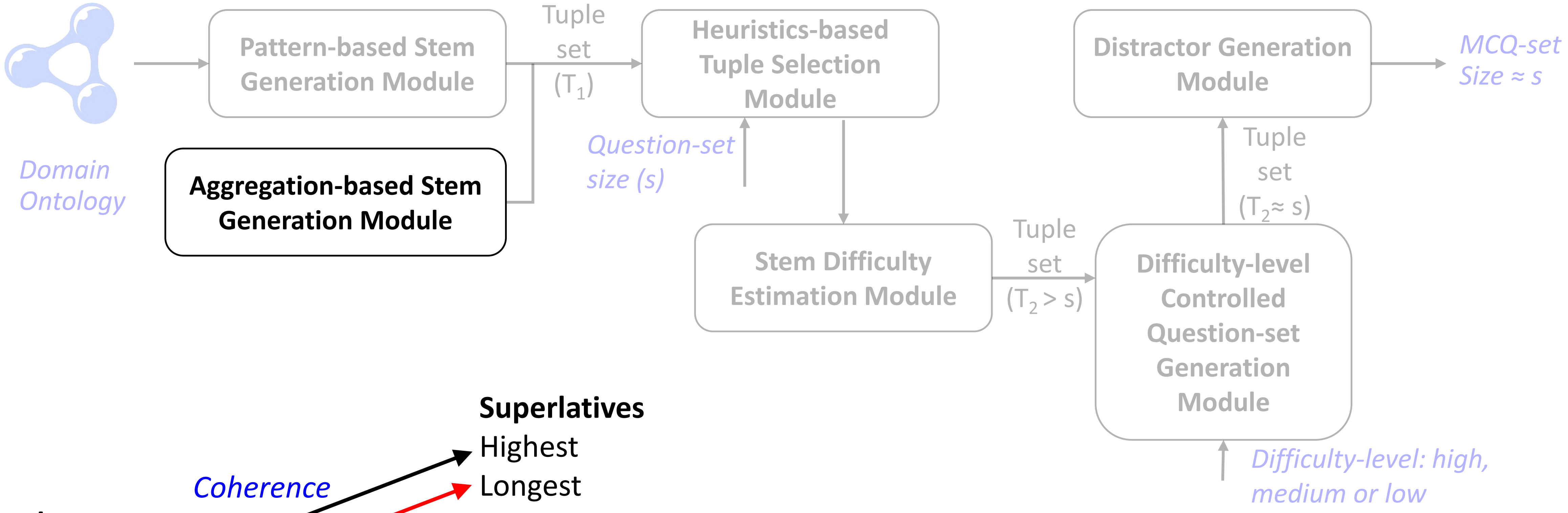
Choose the State with the *highest* population. (Key: Arizona)

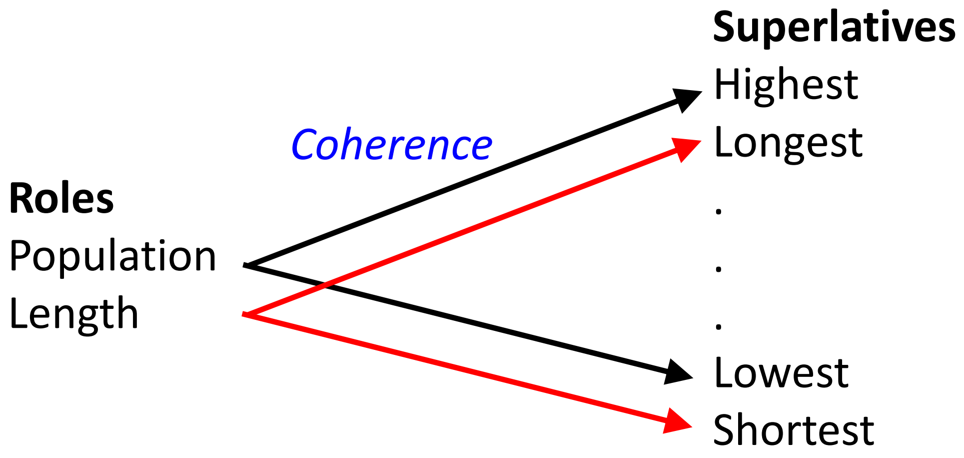
Choose the State with the *lowest* population. (Key: Connecticut)

Choose the River with the *longest* length. (Key: Ouachita)

Choose the River with the *shortest* length. (Key: Neosho)

Proposed System – QG from Knowledge Graph: Extended-ATG System





Choose the State with the *highest* population. (Key: Arizona)

Choose the State with the *lowest* population. (Key: Connecticut)

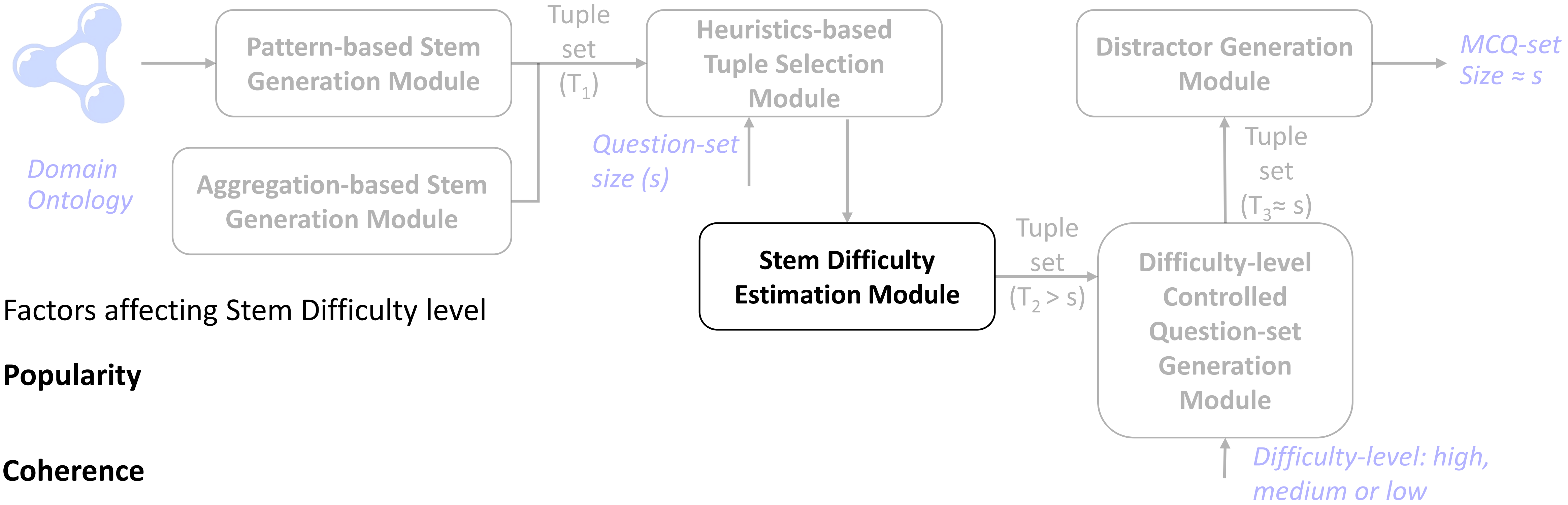
Choose the River with the *longest* length. (Key: Ouachita)

Choose the River with the *shortest* length. (Key: Neosho)

Coherence: Using Explicit-Semantic-Analysis relatedness score

EasyESA API [ISWC ‘14]

Proposed System – QG from Knowledge Graph: Extended-ATG System



Factors affecting Stem Difficulty level

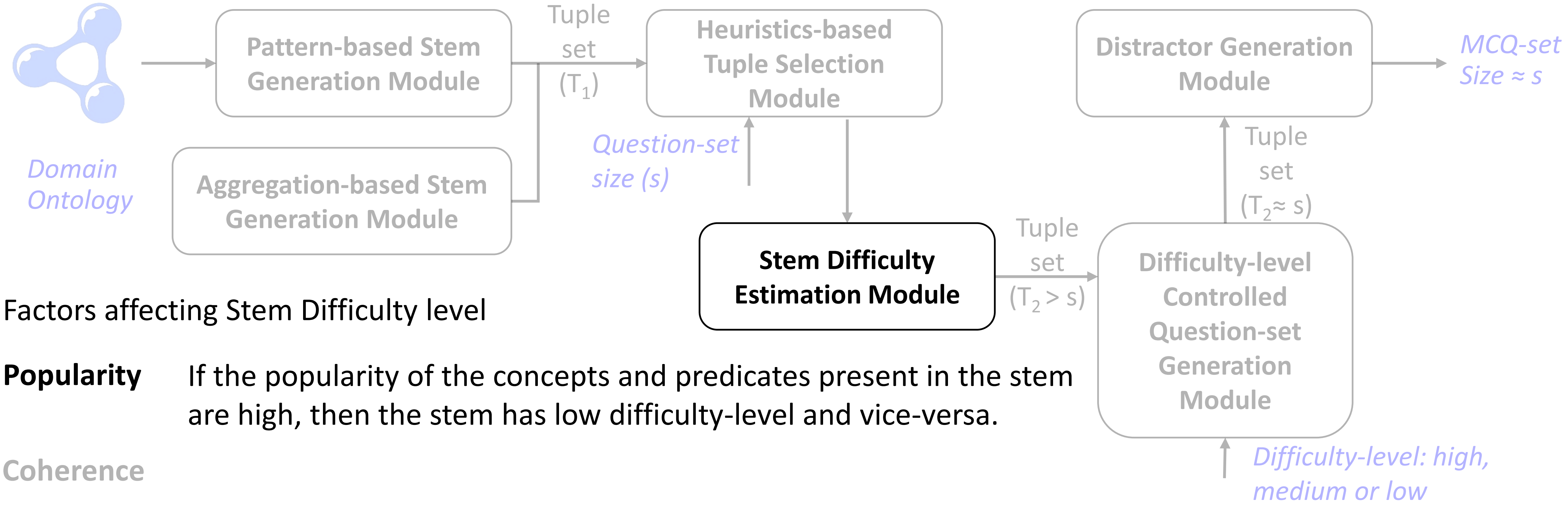
Popularity

Coherence

Specificity

Selectivity

Proposed System – QG from Knowledge Graph: Extended-ATG System



Factors affecting Stem Difficulty level

Popularity If the popularity of the concepts and predicates present in the stem are high, then the stem has low difficulty-level and vice-versa.

Coherence

Specificity

Selectivity

Proposed System – QG from Knowledge Graph: Extended-ATG System

Domain Ontology

Pattern-based Stem Generation Module

Aggregation-based Stem Generation Module

Tuple set (T_1)

Heuristics-based Tuple Selection Module

Question-set size (s)

Stem Difficulty Estimation Module

Tuple set ($T_2 > s$)

Difficulty-level Controlled Question-set Generation Module

Difficulty-level: high, medium or low

Tuple set ($T_2 \approx s$)

Distractor Generation Module

MCQ-set Size ≈ s

Factors affecting Stem Difficulty level

Popularity If the popularity of the concepts and predicates present in the stem are high, then the stem has low difficulty-level and vice-versa.

$$\text{Popularity}(C) = \frac{1}{n}\sum_{j=1}^{n} \text{Indegree}(i_j)$$

where, C is the Concept

n is the no. of individuals satisfying C

i is the individual

Coherence

Specificity

Selectivity

Example
Choose an *Oscar-movie*.
Choose a *Thriller-movie*.

Popularity(*Oscar-movie*) > Popularity(*Thriller-movie*)

Proposed System – QG from Knowledge Graph: Extended-ATG System

Domain Ontology

Pattern-based Stem Generation Module

Aggregation-based Stem Generation Module

Tuple set (T_1)

Heuristics-based Tuple Selection Module

Question-set size (s)

Stem Difficulty Estimation Module

Tuple set ($T_2 > s$)

Difficulty-level Controlled Question-set Generation Module

Difficulty-level: high, medium or low

Tuple set ($T_2 \approx s$)

Distractor Generation Module

MCQ-set Size ≈ s

Factors affecting Stem Difficulty level

Popularity

If the coherence of individuals/concepts in the stem is high, then the stem has low difficulty-level and vice-versa.

Coherence

Specificity

Selectivity

Proposed System – QG from Knowledge Graph: Extended-ATG System

Domain Ontology

Pattern-based Stem Generation Module

Aggregation-based Stem Generation Module

Tuple set (T_1)

Heuristics-based Tuple Selection Module

Question-set size (s)

Stem Difficulty Estimation Module

Tuple set ($T_2 > s$)

Difficulty-level Controlled Question-set Generation Module

Difficulty-level: high, medium or low

Tuple set ($T_2 \approx s$)

Distractor Generation Module

MCQ-set Size ≈ s

Factors affecting Stem Difficulty level

Popularity

Coherence

Specificity

Selectivity

If the coherence of individuals/concepts in the stem is high, then the stem has low difficulty-level and vice-versa.

$$\text{Coherence}(i,j) = \frac{I_i \cap I_j}{I_i \cup I_j} + \frac{O_i \cap O_j}{O_i \cup O_j}$$

where, i,j are Individuals/Concepts

I_i is the set of entities having incoming relation to i

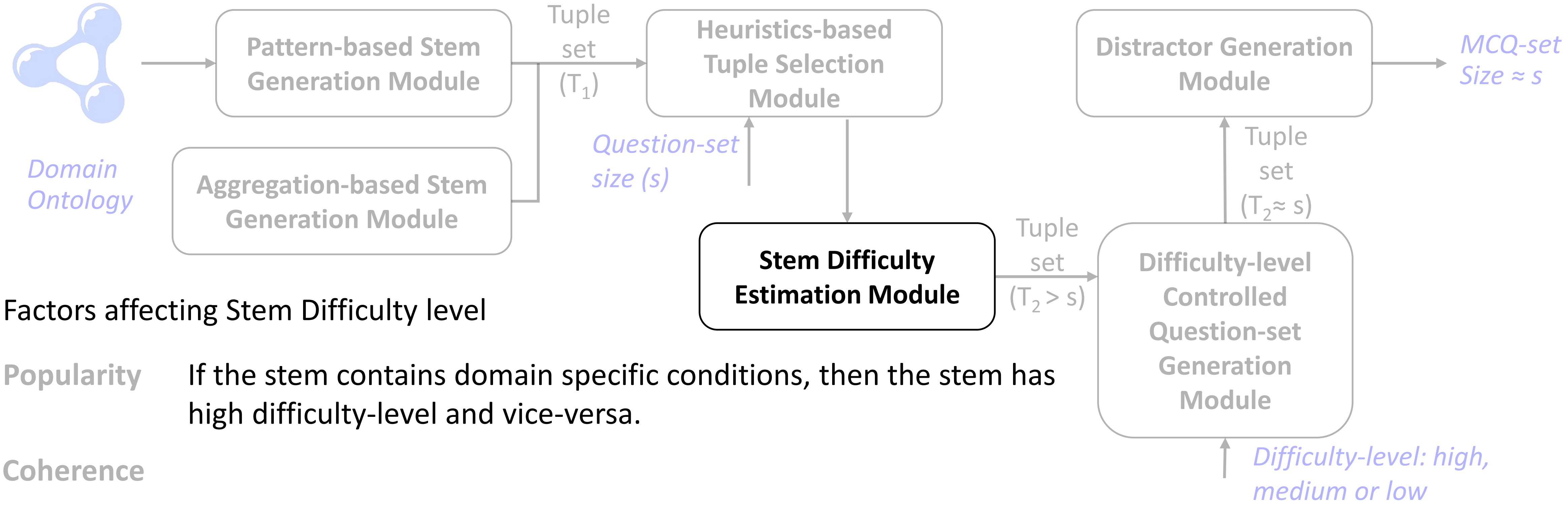
O_i is the set of entities having outgoing relation to i

Example

Choose a *hollywood-movie* starring *Anil Kapoor* and *Tom Cruise*.
Choose a *hollywood-movie* starring *Tom Cruise* and *Tim Robbins*.

Coherence(*hollywood-movie*, *Anil Kapoor*, *Tom Cruise*) < Coherence(*hollywood-movie*, *Tom Cruise*, *Tim Robbins*)

Proposed System – QG from Knowledge Graph: Extended-ATG System



Factors affecting Stem Difficulty level

Popularity

If the stem contains domain specific conditions, then the stem has high difficulty-level and vice-versa.

Coherence

Specificity

Selectivity

Proposed System – QG from Knowledge Graph: Extended-ATG System

Domain Ontology

Pattern-based Stem Generation Module

Aggregation-based Stem Generation Module

Tuple set (T_1)

Heuristics-based Tuple Selection Module

Question-set size (s)

Stem Difficulty Estimation Module

Tuple set ($T_2 > s$)

Difficulty-level Controlled Question-set Generation Module

Difficulty-level: high, medium or low

Tuple set ($T_2 \approx s$)

Distractor Generation Module

MCQ-set Size ≈ s

Factors affecting Stem Difficulty level

Popularity

Coherence

Specificity

Selectivity

If the stem contains domain specific conditions, then the stem has high difficulty-level and vice-versa.

$$\text{Specificity}(p) = \frac{\text{Distance of p from the root of the predicate hierarchy}}{\text{Maximum length of the path contacting p}}$$

where, p is the predicate

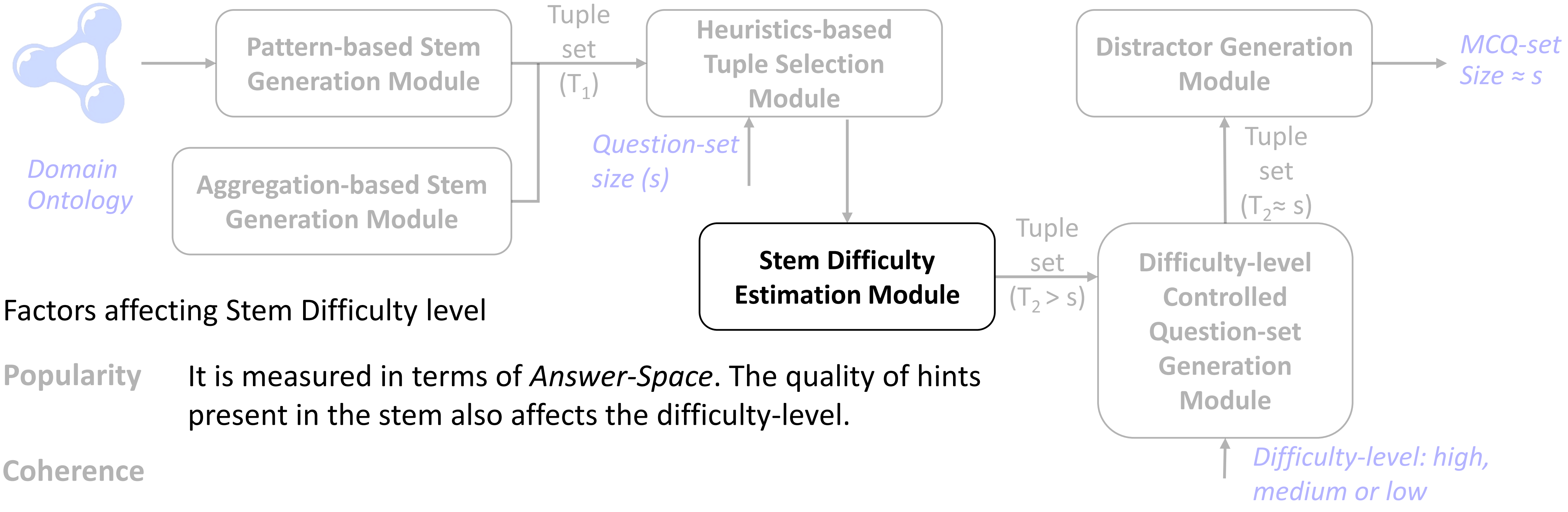
Example

Choose an *Oscar-movie* which was *directed-by* Clint Eastwood.
Choose a *movie* which *is-related-to* Clint Eastwood.

Specificity(*Oscar-movie*) > Specificity(*movie*)

Specificity(*directed-by*) > Specificity(*is-related-to*)

Proposed System – QG from Knowledge Graph: Extended-ATG System



Factors affecting Stem Difficulty level

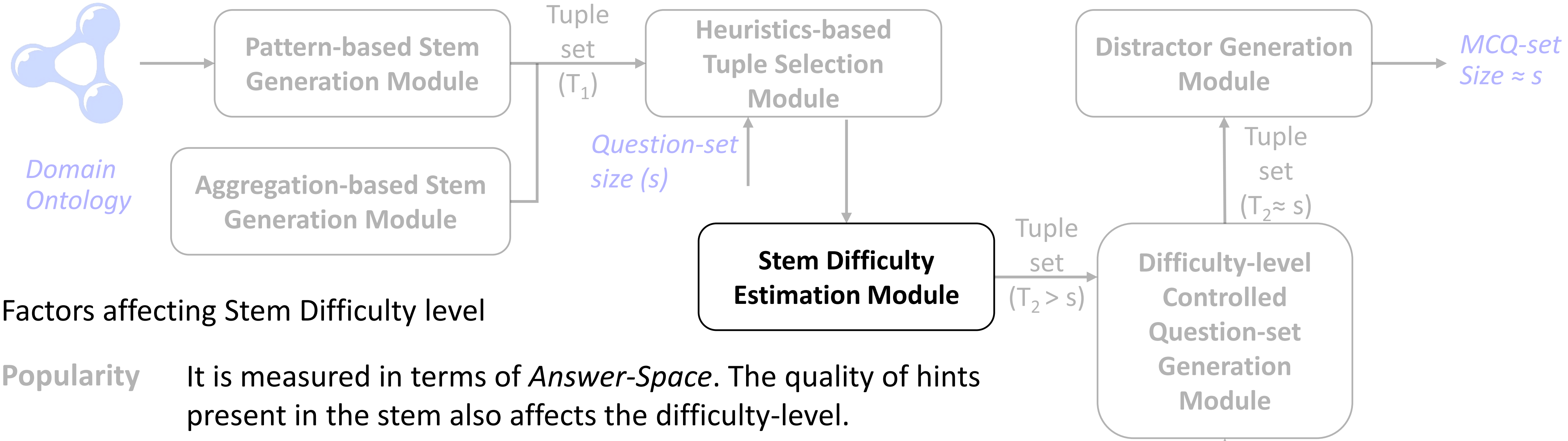
Popularity It is measured in terms of *Answer-Space*. The quality of hints present in the stem also affects the difficulty-level.

Coherence

Specificity

Selectivity

Proposed System – QG from Knowledge Graph: Extended-ATG System



Factors affecting Stem Difficulty level

Popularity

Coherence

Specificity

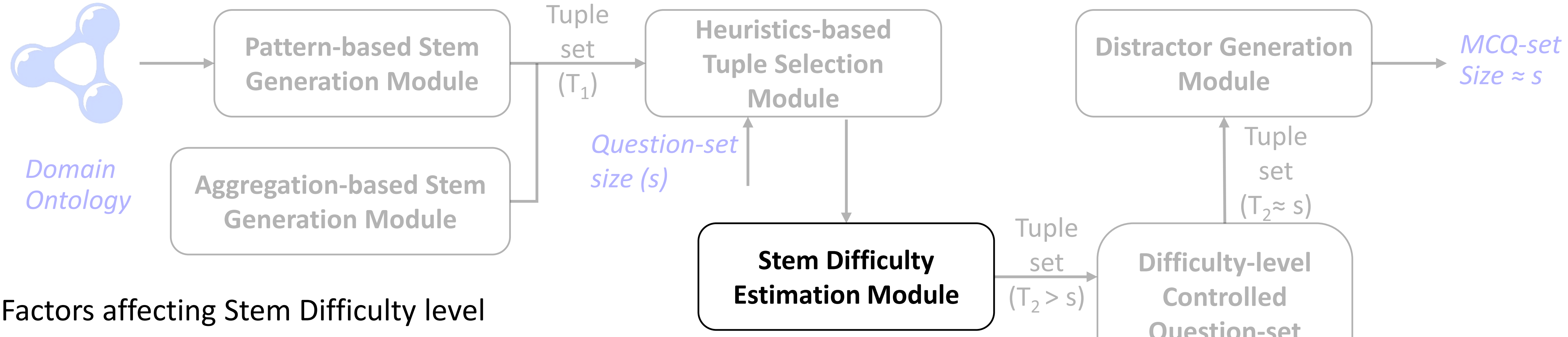
Selectivity It is measured in terms of *Answer-Space*. The quality of hints present in the stem also affects the difficulty-level.

Example

1. Choose a movie acted by Tom Hanks and Denzel Washington.
2. Choose a movie in which Tom Hanks plays the role of Prof. Robert Langdon.
3. Choose a movie for which Tom Hanks won award for best actor.
4. Choose a Tom Hanks movie.

Q. No.	Answer-Space	$\text{Difficulty}_{\text{EXPERT}}$	$\text{Difficulty}_{\text{BEGINNER}}$
1	1	Very Difficult	Very Difficult
2	3	Easy	Difficult
3	6	Difficult	Difficult
4	67	Very Easy	Easy

Proposed System – QG from Knowledge Graph: Extended-ATG System



Factors affecting Stem Difficulty level

Popularity — It is measured in terms of *Answer-Space*. The quality of hints present in the stem also affects the difficulty-level.

Coherence

Specificity

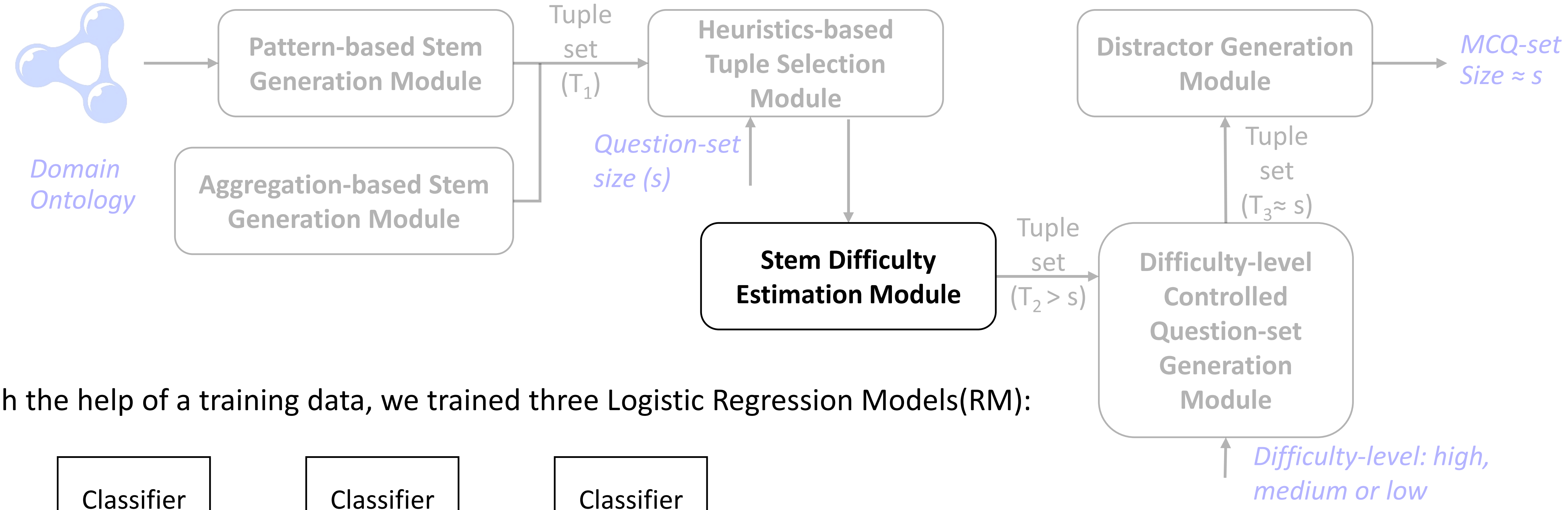
Selectivity

Example

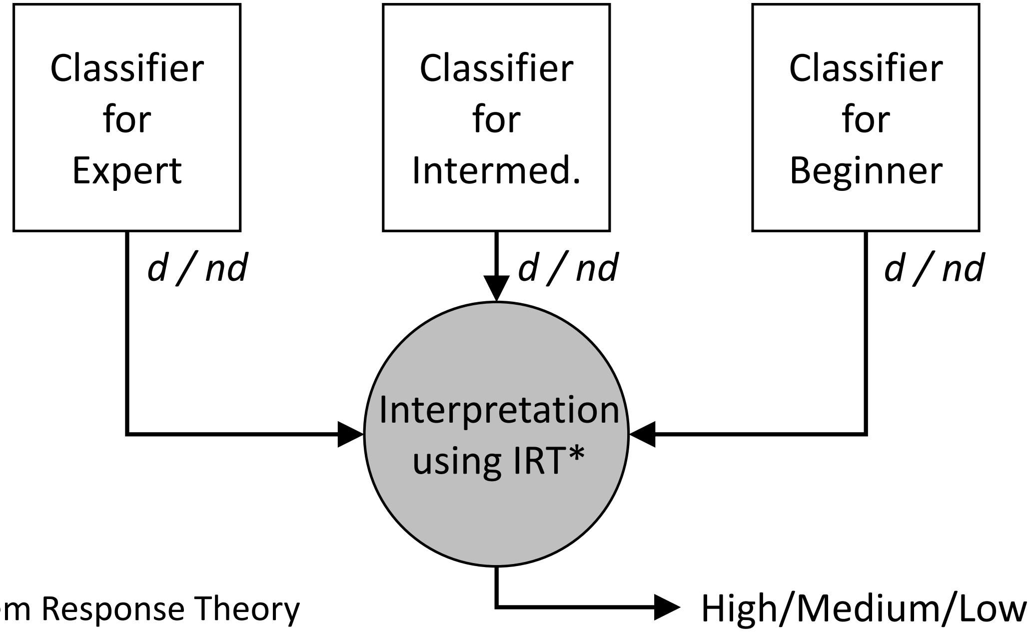
1. Choose a movie acted by Tom Hanks and Denzel Washington. **Rare Question**
2. Choose a movie in which Tom Hanks *plays the role of Prof. Robert Langdon*.
3. Choose a movie for which Tom Hanks *won award for best actor*.
4. Choose a **Tom Hanks movie**.

Q. No.	Answer-Space	$\text{Difficulty}_{\text{EXPERT}}$	$\text{Difficulty}_{\text{BEGINNER}}$
1	1	Very Difficult	Very Difficult
2	3	Easy	Difficult
3	6	Difficult	Difficult
4	67	Very Easy	Easy

Proposed System – QG from Knowledge Graph: Extended-ATG System



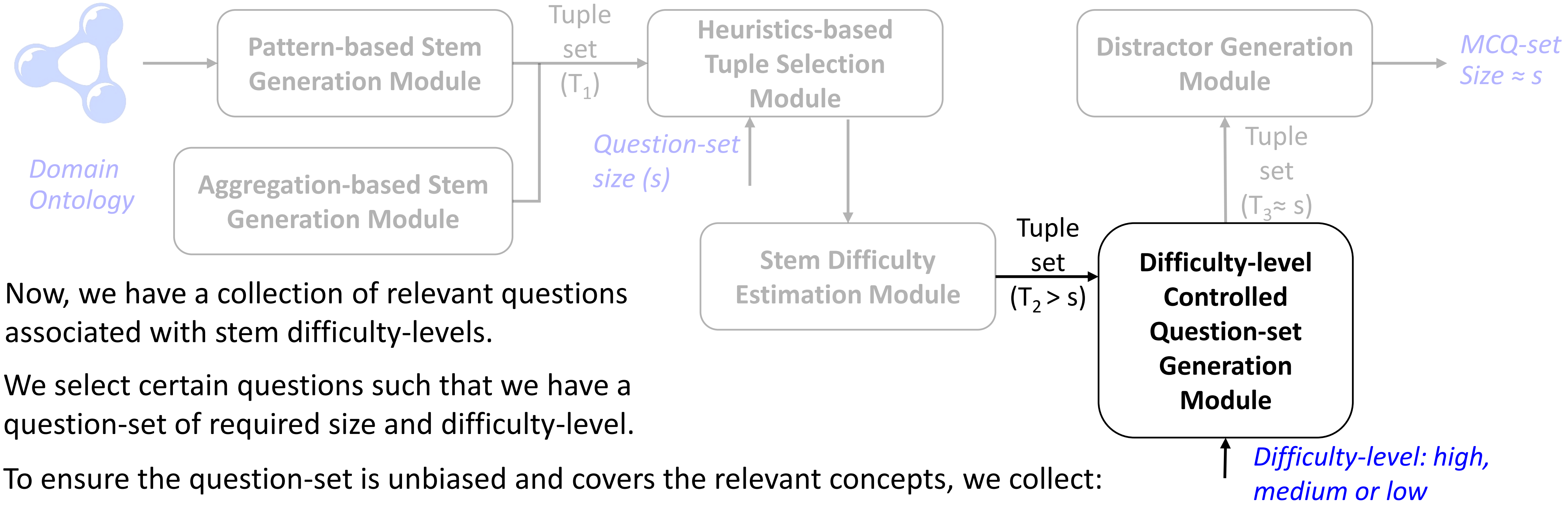
With the help of a training data, we trained three Logistic Regression Models(RM):



* Item Response Theory

RM for Expert	**RM for Intermediate**	**RM for Beginner**	**Difficulty Level**
d	*d*	*d*	High
nd	*d*	*d*	Medium
nd	*nd*	*d*	Low

Proposed System – QG from Knowledge Graph: Extended-ATG System



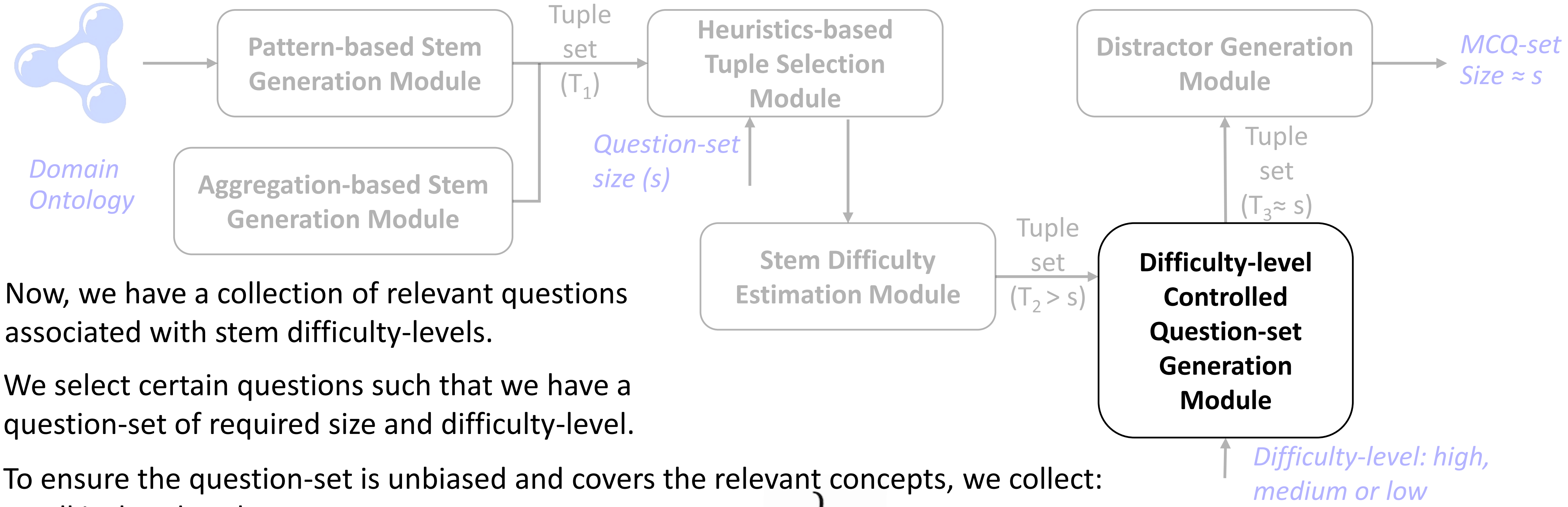
Now, we have a collection of relevant questions associated with stem difficulty-levels.

We select certain questions such that we have a question-set of required size and difficulty-level.

To ensure the question-set is unbiased and covers the relevant concepts, we collect:

- all isolated tuples
- one representative tuple each from group of similar tuples

Proposed System – QG from Knowledge Graph: Extended-ATG System



Now, we have a collection of relevant questions associated with stem difficulty-levels.

We select certain questions such that we have a question-set of required size and difficulty-level.

To ensure the question-set is unbiased and covers the relevant concepts, we collect:

- all isolated tuples
- one representative tuple each from group of similar tuples

≡ **Maximal Independent-Set** (MIS) **of Vertices**

$T_2 = \{t_1, t_2, \ldots, t_n\}$

$G = (T_2, E)$ Tuple Graph

$E = \{(t_1, t_2) \mid t_1, t_2 \in T_2 \text{ and } SemSimiliarity(t_1, t_2) \geq mgs\}$

mgs refers to Minimum Global Similarity threshold

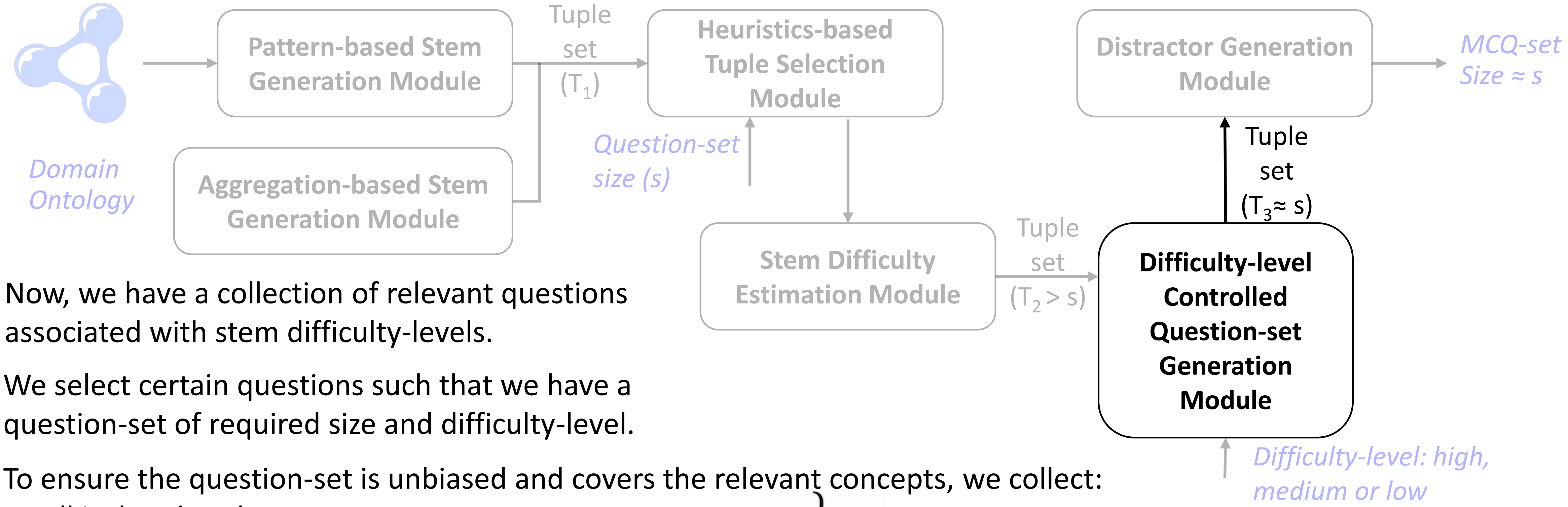
$T_3 \subseteq T_2$, is an Independent-Set

For each $t \in T_2$ 1. $t \in T_3$

or 2. $N(t) \cap T_3 \neq \phi$

where $N(t)$ denotes Neighbours of t

Proposed System – QG from Knowledge Graph: Extended-ATG System



Now, we have a collection of relevant questions associated with stem difficulty-levels.

We select certain questions such that we have a question-set of required size and difficulty-level.

To ensure the question-set is unbiased and covers the relevant concepts, we collect:

- all isolated tuples
- one representative tuple each from group of similar tuples

≡ **Maximal Independent-Set** (MIS) **of Vertices**

$T_2 = \{t_1, t_2, \ldots, t_n\}$

$G = (T_2, E)$ Tuple Graph

$E = \{(t_1, t_2) \mid t_1, t_2 \in T_2 \text{ and } SemSimiliarity(t_1, t_2) \geq mgs\}$

mgs refers to Minimum Global Similarity threshold

$T_3 \subseteq T_2$, is an Independent-Set

For each $t \in T_2$ 1. $t \in T_3$

or 2. $N(t) \cap T_3 \neq \phi$

where N(t) denotes Neighbours of t

In the MIS algorithm, the tuples are prioritized to generate question-sets of High, Medium and Low difficulty-levels.

Results – QG from Knowledge Graph

Experiment-1: Evaluation with the Benchmark Question-Sets

Generated question-sets of sizes: 25, 45 and 75
from Data Structures & Algorithms (DSA), Mahabharata (MAHA) and Geographical Entities (GEO)

Compared with respective Benchmark Question-Sets.

Evaluation Metrics: Precision and Recall

Precision: on an avg. 81%

Recall: on an avg. 50%

Results – QG from Knowledge Graph

Experiment-1: Evaluation with the Benchmark Question-Sets

Generated question-sets of sizes: 25, 45 and 75
from Data Structures & Algorithms (DSA), Mahabharata (MAHA) and Geographical Entities (GEO)

Compared with respective Benchmark Question-Sets.

Evaluation Metrics: Precision and Recall

Precision: on an avg. 81%

Recall: on an avg. 50%

Experiment-2: Evaluation of Stem Difficulty

Question-set of 24 test MCQs from DSA ontology
8 MCQs each of high, medium and low (predicted) difficulty-levels.

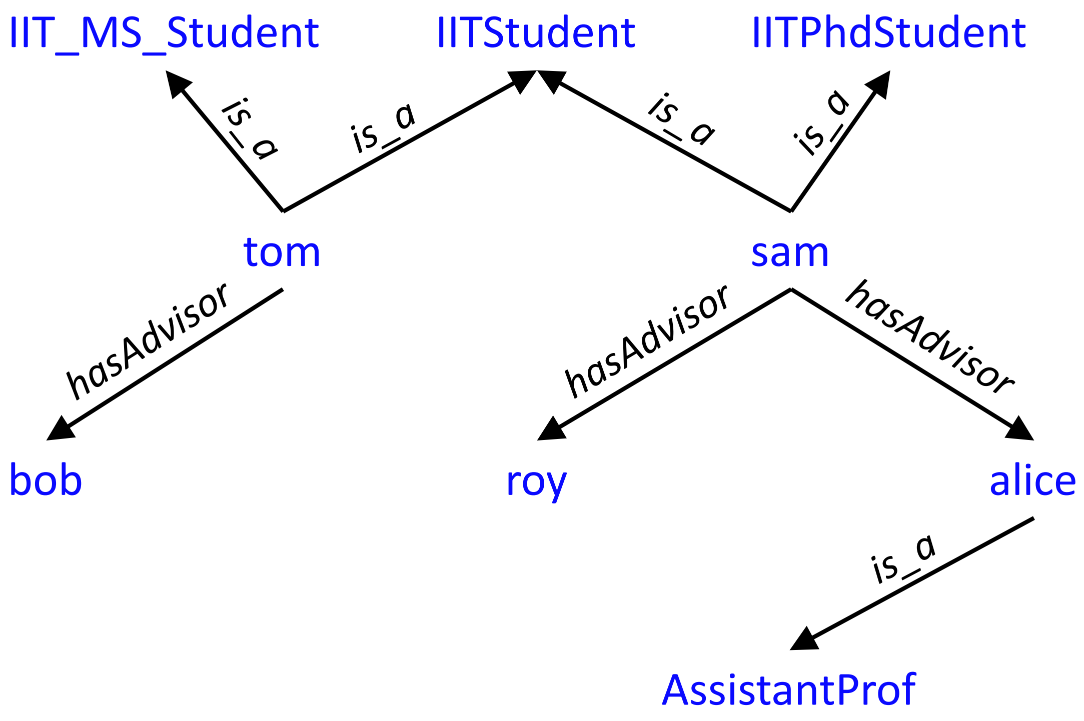
Evaluated in a classroom setup including 54 students
Three groups with 18 students each – with knowledge proficiencies Expert, Intermediate and Beginner

The actual difficulty-levels calculated had a correlation of 71% with the predicted difficulty-levels

Proposed System – QG from Schema

ABox

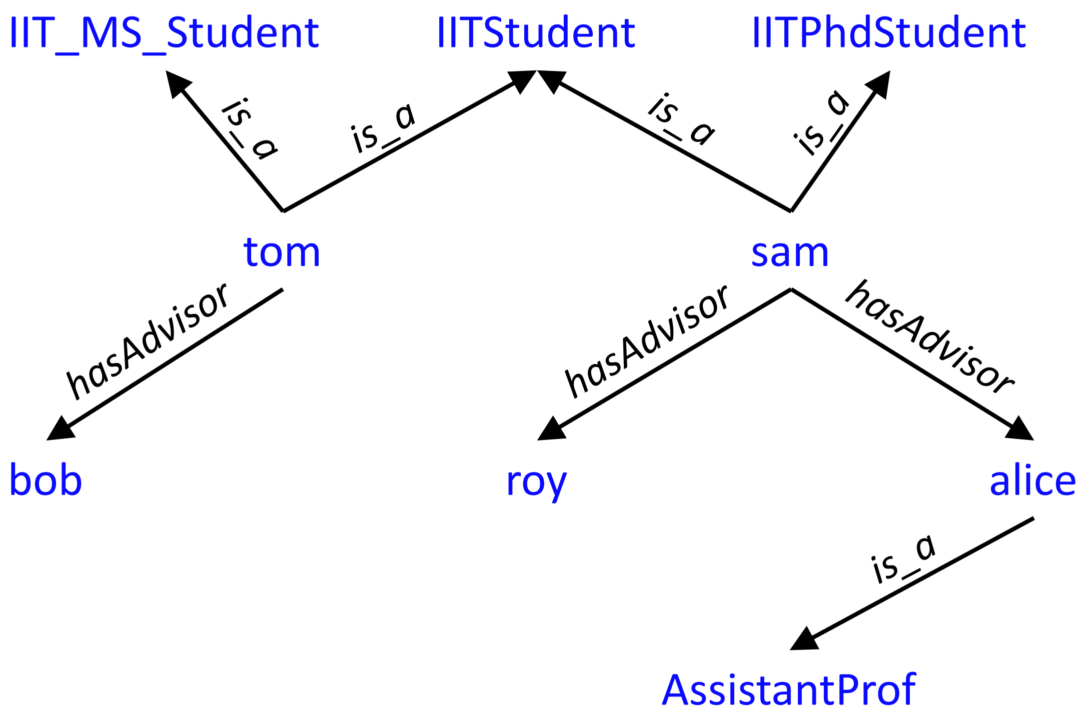
IITStudent(tom)
IIT_MS_Student(tom)
hasAdvisor(tom, bob)
IITPhdStudent(sam)
hasAdvisor(sam, alice)
hasAdvisor(sam, roy)
AssistantProf(alice)



Proposed System – QG from Schema

ABox

IITStudent(tom)
IIT_MS_Student(tom)
hasAdvisor(tom, bob)
IITPhdStudent(sam)
hasAdvisor(sam, alice)
hasAdvisor(sam, roy)
AssistantProf(alice)



Stem: Choose an IIT student whose advisor is Bob.

Key: Tom

Stem: Choose an assistant professor.

Key: Alice

Proposed System – QG from Schema

TBox

$$
\begin{aligned}
\mathsf{IITStudent} &\equiv \mathsf{Student} \sqcap \forall \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \exists \mathsf{hasAdvisor.Professor} \sqcap \exists \mathsf{enrolledIn.IITProgramme} \\
\mathsf{IIT\_MS\_Student} &\equiv \mathsf{IITStudent} \sqcap \leq 1\, \mathsf{hasAdvisor.TeachingStaff} \\
\mathsf{IITPhdStudent} &\equiv \mathsf{IITStudent} \sqcap \geq 2\, \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \leq 1\, \mathsf{hasAdvisor.Professor} \\
\mathsf{Professor} &\sqsubseteq \mathsf{TeachingStaff} \\
\mathsf{AssistantProf} &\sqsubseteq \mathsf{TeachingStaff} \\
\bot &\equiv \mathsf{Professor} \sqcap \mathsf{AssistantProf}
\end{aligned}
$$

Proposed System – QG from Schema

TBox

$$
\begin{aligned}
\mathsf{IITStudent} &\equiv \mathsf{Student} \sqcap \forall \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \exists \mathsf{hasAdvisor.Professor} \sqcap \exists \mathsf{enrolledIn.IITProgramme} \\
\mathsf{IIT\_MS\_Student} &\equiv \mathsf{IITStudent} \sqcap \leq 1\, \mathsf{hasAdvisor.TeachingStaff} \\
\mathsf{IITPhdStudent} &\equiv \mathsf{IITStudent} \sqcap \geq 2\, \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \leq 1\, \mathsf{hasAdvisor.Professor} \\
\mathsf{Professor} &\sqsubseteq \mathsf{TeachingStaff} \\
\mathsf{AssistantProf} &\sqsubseteq \mathsf{TeachingStaff} \\
\bot &\equiv \mathsf{Professor} \sqcap \mathsf{AssistantProf}
\end{aligned}
$$

Label Set

The set of all the (concept & role) restrictions corresponding to an individual is called its Label Set

Proposed System – QG from Schema

TBox

$$
\begin{aligned}
\mathsf{IITStudent} &\equiv \mathsf{Student} \sqcap \forall \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \exists \mathsf{hasAdvisor.Professor} \sqcap \exists \mathsf{enrolledIn.IITProgramme} \\
\mathsf{IIT\_MS\_Student} &\equiv \mathsf{IITStudent} \sqcap \leq 1\, \mathsf{hasAdvisor.TeachingStaff} \\
\mathsf{IITPhdStudent} &\equiv \mathsf{IITStudent} \sqcap \geq 2\, \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \leq 1\, \mathsf{hasAdvisor.Professor} \\
\mathsf{Professor} &\sqsubseteq \mathsf{TeachingStaff} \\
\mathsf{AssistantProf} &\sqsubseteq \mathsf{TeachingStaff} \\
\perp &\equiv \mathsf{Professor} \sqcap \mathsf{AssistantProf}
\end{aligned}
$$

Label Set

The set of all the (concept & role) restrictions corresponding to an individual is called its Label Set

For a SHIQ ontology O, the label-sets of a individual x can be defined as:

$$L_O(x) = \{c_1 , c_2 , c_3 , \ldots , c_n \}$$

$c_i = A \mid \exists R.C \mid \forall R.C \mid \leq nR.C \mid \geq nR.C$

$C, D = A \mid C \sqcup D \mid C \sqcap D \mid \exists R.C \mid \forall R.C \mid \leq nR.C \mid \geq nR.C$

for all $c_i \in L_O(x)$, $O \models c_i(x)$

Proposed System – QG from Schema

TBox

$$
\begin{array}{rcl}
\mathsf{IITStudent} & \equiv & \mathsf{Student} \sqcap \forall \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
 & & \exists \mathsf{hasAdvisor.Professor} \sqcap \exists \mathsf{enrolledIn.IITProgramme} \\
\mathsf{IIT\_MS\_Student} & \equiv & \mathsf{IITStudent} \sqcap \leq 1\, \mathsf{hasAdvisor.TeachingStaff} \\
\mathsf{IITPhdStudent} & \equiv & \mathsf{IITStudent} \sqcap \geq 2\, \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
 & & \leq 1\, \mathsf{hasAdvisor.Professor} \\
\mathsf{Professor} & \sqsubseteq & \mathsf{TeachingStaff} \\
\mathsf{AssistantProf} & \sqsubseteq & \mathsf{TeachingStaff} \\
\perp & \equiv & \mathsf{Professor} \sqcap \mathsf{AssistantProf}
\end{array}
$$

Label Set

The set of all the (concept & role) restrictions corresponding to an individual is called its Label Set

For a SHIQ ontology O, the label-sets of a individual x can be defined as:

$$L_O(x) = \{c_1, c_2, c_3, \ldots, c_n\}$$

$c_i = A \mid \exists R.C \mid \forall R.C \mid \leq nR.C \mid \geq nR.C$

$C, D = A \mid C \sqcup D \mid C \sqcap D \mid \exists R.C \mid \forall R.C \mid \leq nR.C \mid \geq nR.C$

for all $c_i \in L_O(x)$, $O \vDash c_i(x)$

Property of a label-set

Conjunction of all the elements in the label-set of an individual is entailed by the ontology.

$$O \vDash \prod_{i=1}^{n} c_i(x)$$

Proposed System – QG from Schema

TBox

$$
\begin{aligned}
\mathsf{IITStudent} &\equiv \mathsf{Student} \sqcap \forall \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \exists \mathsf{hasAdvisor.Professor} \sqcap \exists \mathsf{enrolledIn.IITProgramme} \\
\mathsf{IIT\_MS\_Student} &\equiv \mathsf{IITStudent} \sqcap \leq 1\, \mathsf{hasAdvisor.TeachingStaff} \\
\mathsf{IITPhdStudent} &\equiv \mathsf{IITStudent} \sqcap \geq 2\, \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \leq 1\, \mathsf{hasAdvisor.Professor} \\
\mathsf{Professor} &\sqsubseteq \mathsf{TeachingStaff} \\
\mathsf{AssistantProf} &\sqsubseteq \mathsf{TeachingStaff} \\
\perp &\equiv \mathsf{Professor} \sqcap \mathsf{AssistantProf}
\end{aligned}
$$

Label Set

The set of all the (concept & role) restrictions corresponding to an individual is called its Label Set

For a SHIQ ontology O, the label-sets of a individual x can be defined as:

$$L_O(x) = \{c_1 , c_2 , c_3 , \ldots , c_n \}$$

$c_i = A \mid \exists R.C \mid \forall R.C \mid \leq nR.C \mid \geq nR.C$

$C, D = A \mid C \sqcup D \mid C \sqcap D \mid \exists R.C \mid \forall R.C \mid \leq nR.C \mid \geq nR.C$

for all $c_i \in L_O(x)$, $O \models c_i(x)$

Property of a label-set

Conjunction of all the elements in the label-set of an individual is entailed by the ontology.

$$O \models \prod_{i=1}^{n} c_i(x)$$

Generation of label-set

We find all the membership concepts of the individual and all the associated restrictions.

Proposed System – QG from Schema

TBox

$$
\begin{aligned}
\mathsf{IITStudent} &\equiv \mathsf{Student} \sqcap \forall \mathsf{hasAdvisor}.\mathsf{TeachingStaff} \sqcap \\
&\quad \exists \mathsf{hasAdvisor}.\mathsf{Professor} \sqcap \exists \mathsf{enrolledIn}.\mathsf{IITProgramme} \\
\mathsf{IIT\_MS\_Student} &\equiv \mathsf{IITStudent} \sqcap \leq 1\, \mathsf{hasAdvisor}.\mathsf{TeachingStaff} \\
\mathsf{IITPhdStudent} &\equiv \mathsf{IITStudent} \sqcap \geq 2\, \mathsf{hasAdvisor}.\mathsf{TeachingStaff} \sqcap \\
&\quad \leq 1\, \mathsf{hasAdvisor}.\mathsf{Professor} \\
\mathsf{Professor} &\sqsubseteq \mathsf{TeachingStaff} \\
\mathsf{AssistantProf} &\sqsubseteq \mathsf{TeachingStaff} \\
\bot &\equiv \mathsf{Professor} \sqcap \mathsf{AssistantProf}
\end{aligned}
$$

L_O(tom) = { Student, IITStudent, IIT_MS_Student, ∃enrolledIn.IITProgramme, ≤ 1 hasAdvisor.TeachingStaff, ∀hasAdvisor.TeachingStaff, ∃hasAdvisor.Professor }

L_O(sam) = { Student, IITStudent, IITPhdStudent, ∃isEnrolledIn.IITProgramme, ≥ 2 hasAdvisor.TeachingStaff, ≤ 1 hasAdvisor.Professor, ∀hasAdvisor.TeachingStaff, ∃hasAdvisor.Professor }

Proposed System – QG from Schema

TBox

$$
\begin{aligned}
\mathsf{IITStudent} &\equiv \mathsf{Student} \sqcap \forall \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \exists \mathsf{hasAdvisor.Professor} \sqcap \exists \mathsf{enrolledIn.IITProgramme} \\
\mathsf{IIT\_MS\_Student} &\equiv \mathsf{IITStudent} \sqcap \leq 1\, \mathsf{hasAdvisor.TeachingStaff} \\
\mathsf{IITPhdStudent} &\equiv \mathsf{IITStudent} \sqcap \geq 2\, \mathsf{hasAdvisor.TeachingStaff} \sqcap \\
&\quad \leq 1\, \mathsf{hasAdvisor.Professor} \\
\mathsf{Professor} &\sqsubseteq \mathsf{TeachingStaff} \\
\mathsf{AssistantProf} &\sqsubseteq \mathsf{TeachingStaff} \\
\bot &\equiv \mathsf{Professor} \sqcap \mathsf{AssistantProf}
\end{aligned}
$$

L_O(tom) = { Student, IITStudent, IIT_MS_Student, ∃enrolledIn.IITProgramme, ≤ 1 hasAdvisor.TeachingStaff, ∀hasAdvisor.TeachingStaff, ∃hasAdvisor.Professor }

L_O(sam) = { Student, IITStudent, IITPhdStudent, ∃isEnrolledIn.IITProgramme, ≥ 2 hasAdvisor.TeachingStaff, ≤ 1 hasAdvisor.Professor, ∀hasAdvisor.TeachingStaff, ∃hasAdvisor.Professor }

Stem: Choose a student who is enrolled in an IIT Programme, has one professor as advisor and all his advisors are teaching staffs.

Key: Tom

Stem: Choose a student who is enrolled in an IIT Programme, has exactly one advisor who is a professor and has at least one more advisor who is a teaching staff but not a professor.

Key: Sam

Proposed System – Semantic-Refinement of Label-Sets

Naive method

Consider all combinations of labels in the label-set and check whether they can be reduced.

Proposed System – Semantic-Refinement of Label-Sets

Naive method

Consider all combinations of labels in the label-set and check whether they can be reduced.

Proposed Method

We proposed 7 set of rules, where labels of a specific restriction types are handled in a pre-defined order to achieve complete reduction.

I Concept Refinement Rule → II Superclass Refinement Rule → III Existential Role Refinement Rule → IV Universal Role Refinement Rules → V III & IV Refinement Rules → VI Qualified Number Restriction Refinement Rules → VII Exactly-n Role Refinement Rules

Proposed System – Semantic-Refinement of Label-Sets

Naive method

Consider all combinations of labels in the label-set and check whether they can be reduced.

Proposed Method

We proposed 7 set of rules, where labels of a specific restriction types are handled in a pre-defined order to achieve complete reduction.

I. Concept Refinement Rule
II. Superclass Refinement Rule
III. Existential Role Refinement Rule
IV. Universal Role Refinement Rules
V. III & IV Refinement Rules
VI. Qualified Number Restriction Refinement Rules
VII. Exactly-n Role Refinement Rules

$\exists R.U \qquad \exists S.V$

$U \sqsubseteq V$

$R \sqsubseteq S$

$\exists R.U$

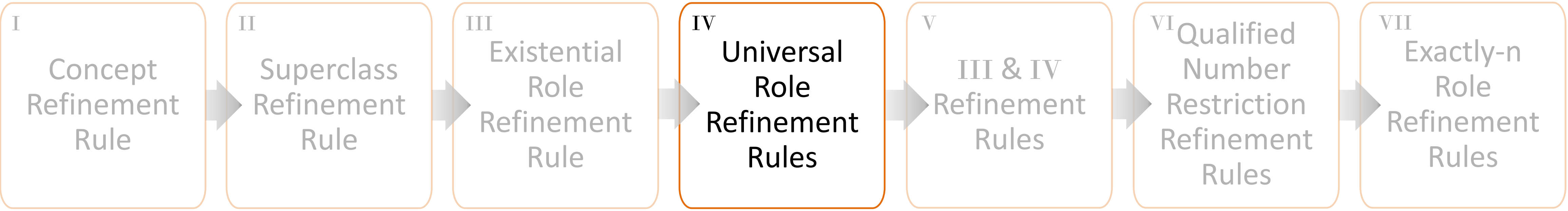
Proposed System – Semantic-Refinement of Label-Sets

Naive method

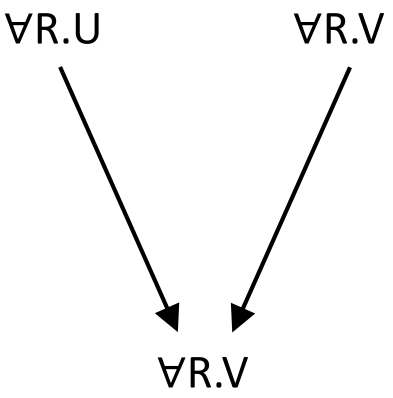
Consider all combinations of labels in the label-set and check whether they can be reduced.

Proposed Method

We proposed 7 set of rules, where labels of a specific restriction types are handled in a pre-defined order to achieve complete reduction.



$V \sqsubseteq U$



Proposed System – Semantic-Refinement of Label-Sets

Naive method

Consider all combinations of labels in the label-set and check whether they can be reduced.

Proposed Method

We proposed 7 set of rules, where labels of a specific restriction types are handled in a pre-defined order to achieve complete reduction.

I Concept Refinement Rule → II Superclass Refinement Rule → III Existential Role Refinement Rule → IV Universal Role Refinement Rules → V III & IV Refinement Rules → VI Qualified Number Restriction Refinement Rules → VII **Exactly-n Role Refinement Rules**

$\geq mR.V$ $\quad \exists_{=n}R.U$

$U \sqsubseteq V$

$m \geq n$

$\exists{=}nR.U,\ \geq(m-n)R.(V \sqcap \neg U)$

Proposed System – Semantic-Refinement of Label-Sets

Individual	Label-Set L_O	Semantically Refined Label-Set $R(L_O)$
tom	{ Student, IITStudent, IIT_MS_Student, ∃isEnrolledIn.IITProgramme, ≤1hasAdvisor.TeachingStaff, ∀hasAdvisor.TeachingStaff, ∃hasAdvisor.Professor }	{ Student, ∃isEnrolledIn.IITProgramme, $\exists_{=1}$hasAdvisor.Professor }
sam	{ Student, IITStudent, IITPhdStudent, ∃isEnrolledIn.IITProgramme, ≥2hasAdvisor.TeachingStaff, ≤1hasAdvisor.Professor, ∀hasAdvisor.TeachingStaff, ∃hasAdvisor.Professor }	{ Student, ∃isEnrolledIn.IITProgramme, $\exists_{=1}$hasAdvisor.Professor, ∀hasAdvisor.TeachingStaff, ≥1hasAdvisor.(TeachingStaff ⊓¬Professor)}

Proposed System – Semantic-Refinement of Label-Sets

Individual	Label-Set L_O	Semantically Refined Label-Set $R(L_O)$
tom	{ Student, IITStudent, IIT_MS_Student, $\exists$isEnrolledIn.IITProgramme, $\leq$1hasAdvisor.TeachingStaff, $\forall$hasAdvisor.TeachingStaff, $\exists$hasAdvisor.Professor }	{ Student, $\exists$isEnrolledIn.IITProgramme, $\exists_{=1}$hasAdvisor.Professor }
sam	{ Student, IITStudent, IITPhdStudent, $\exists$isEnrolledIn.IITProgramme, $\geq$2hasAdvisor.TeachingStaff, $\leq$1hasAdvisor.Professor, $\forall$hasAdvisor.TeachingStaff, $\exists$hasAdvisor.Professor }	{ Student, $\exists$isEnrolledIn.IITProgramme, $\exists_{=1}$hasAdvisor.Professor, [illegible]hasAdvisor.TeachingStaff, $\geq$1hasAdvisor.(TeachingStaff $\sqcap\neg$Professor)}

Question generated without Semantic-Refinement

Stem: Choose a student, who is an IIT student and an IIT MS student, who is enrolled in an IIT Programme, and have less than one advisor who is a teaching staff, and all his advisors are teaching staffs, and has an advisor who is a professor.

Key: Tom

Question generated after Semantic-Refinement

Stem: Choose a student who is enrolled in an IIT Programme, has exactly one advisor who is a professor.

Key: Tom

Proposed System & Results – QG from TBox

Distractor Generation

Under Open World Assumption, the statements, which are not a logical consequence of an ontology, are not considered false

i.e. $O \nvDash C(a)$ does not imply $C(a)$ is false.

but, $O \vDash \neg C(a)$ implies $C(a)$ is false.

Select the individual as a distractor if it is disjoint with at least one restriction in the label-set

Proposed System & Results – QG from TBox

Distractor Generation

Under Open World Assumption, the statements, which are not a logical consequence of an ontology, are not considered false

i.e. $O \nvDash C(a)$ does not imply $C(a)$ is false.

but, $O \vDash \neg C(a)$ implies $C(a)$ is false.

Select the individual as a distractor if it is disjoint with at least one restriction in the label-set

Experiment-1: Feasibility Study of the proposed methodology

Implemented a prototype system and generated questions from the ontologies:
Harry Potter(HP), People & Pets(PP), Plant Disease(PD), US Geography(GEO-US)

Proposed System & Results – QG from TBox

Distractor Generation

Under Open World Assumption, the statements, which are not a logical consequence of an ontology, are not considered false

i.e. $O \not\models C(a)$ does not imply $C(a)$ is false.

but, $O \models \neg C(a)$ implies $C(a)$ is false.

Select the individual as a distractor if it is disjoint with at least one restriction in the label-set

Experiment-1: Feasibility Study of the proposed methodology

Implemented a prototype system and generated questions from the ontologies:
Harry Potter(HP), People & Pets(PP), Plant Disease(PD), US Geography(GEO-US)

Experiment-2: Usefulness of the Generated Questions

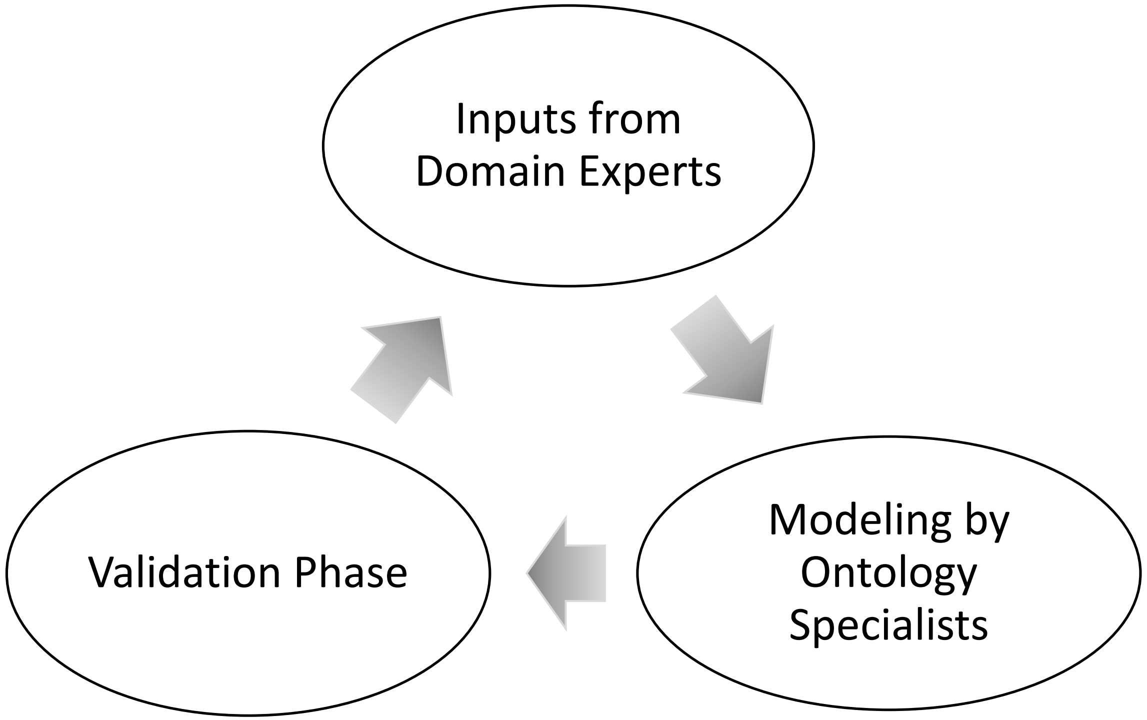
We considered Plant Disease (PD) ontology for this analysis, with the help of 7 domain experts

The results were as follows:

Useful	71%
Not useful, but domain related	3%
Not useful, not domain related	26%

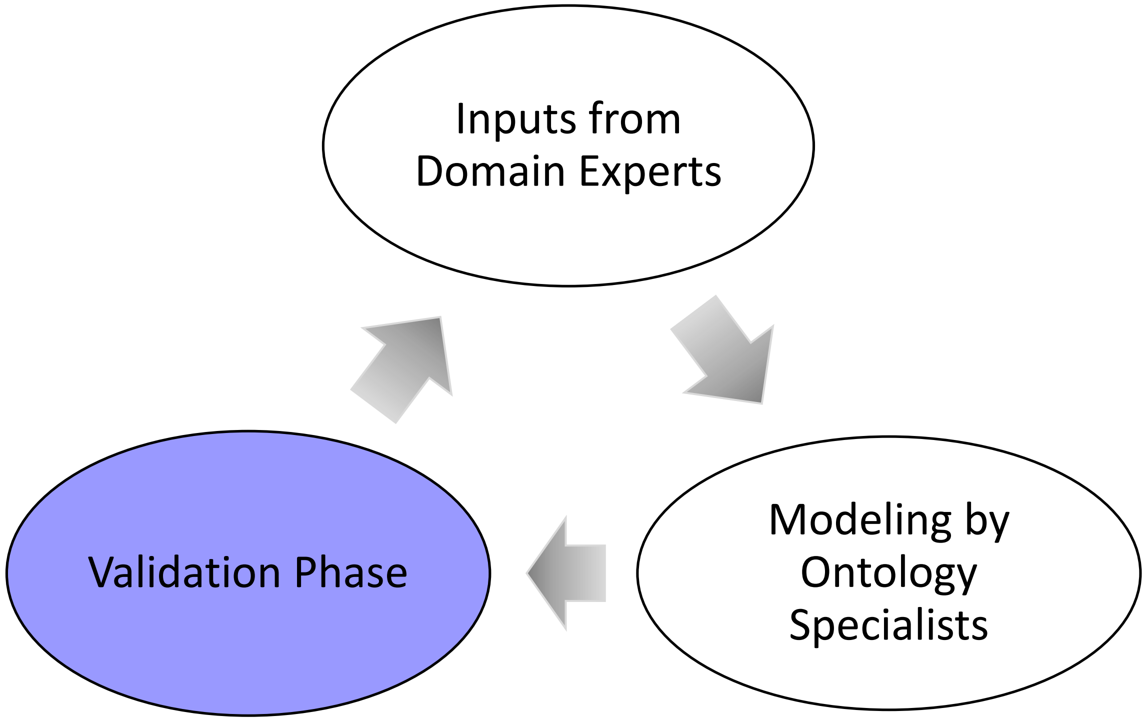
Proposed System – Applications of Semantic-Refinement

Ontology Development Cycle



Proposed System – Applications of Semantic-Refinement

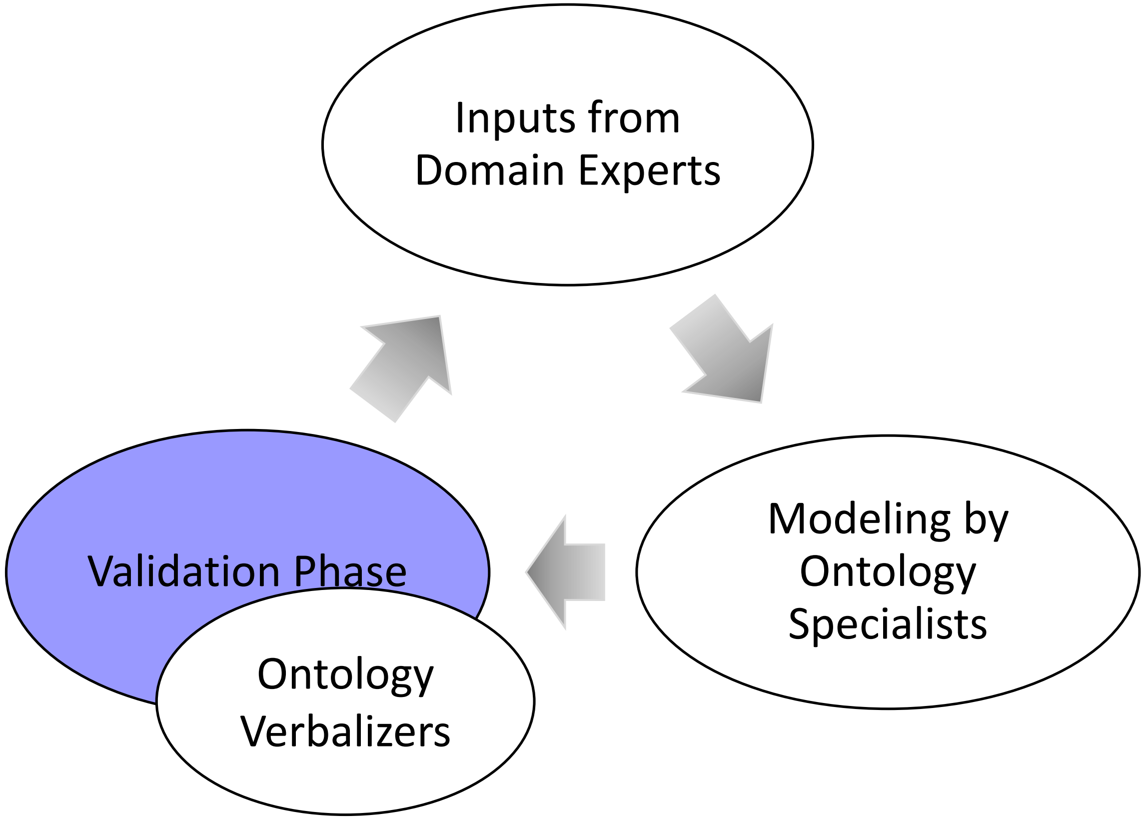
Ontology Development Cycle



At Validation Phase, Domain Experts come together with Ontology Specialists

Proposed System – Applications of Semantic-Refinement

Ontology Development Cycle



At Validation Phase, Domain Experts come together with Ontology Specialists

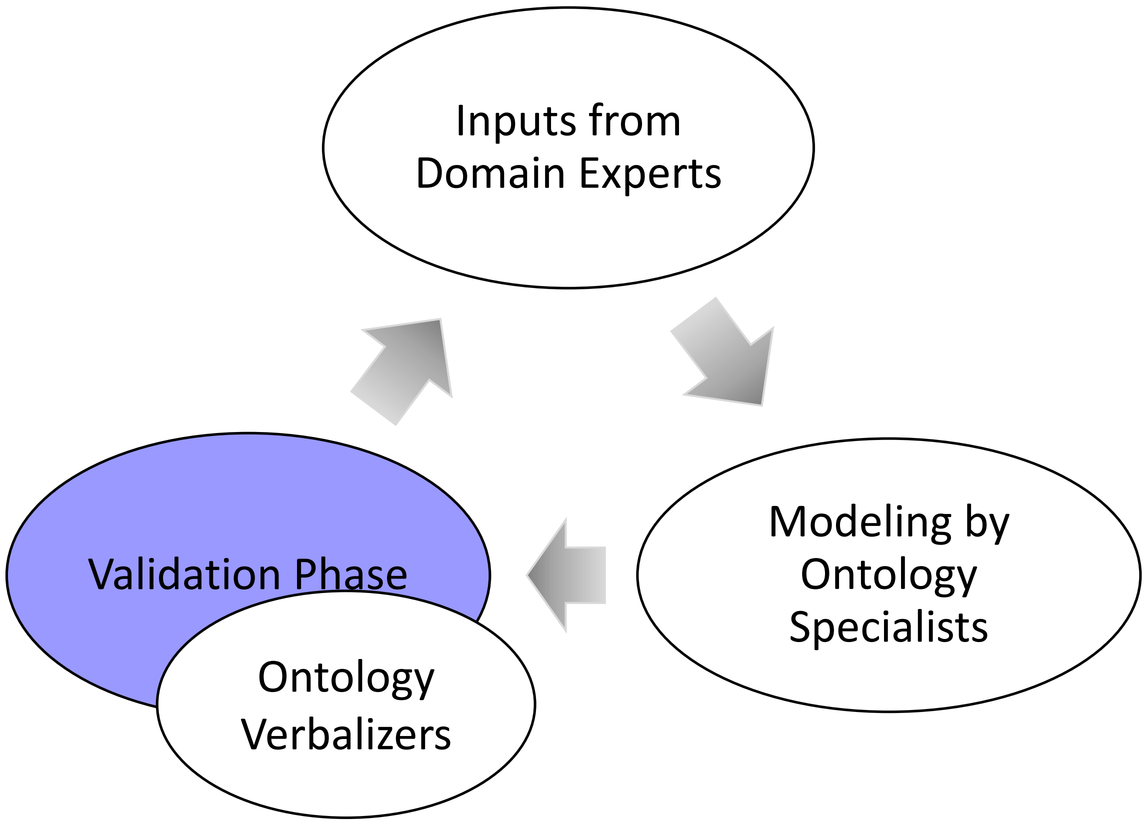
Ontology Verbalizers (OV) are of great use here

CNLs for OWL: ACE Rabbit SOS

Tool: SWAT

Proposed System – Applications of Semantic-Refinement

Ontology Development Cycle



At Validation Phase, Domain Experts come together with Ontology Specialists

Ontology Verbalizers (OV) are of great use here

CNLs for OWL: ACE Rabbit SOS

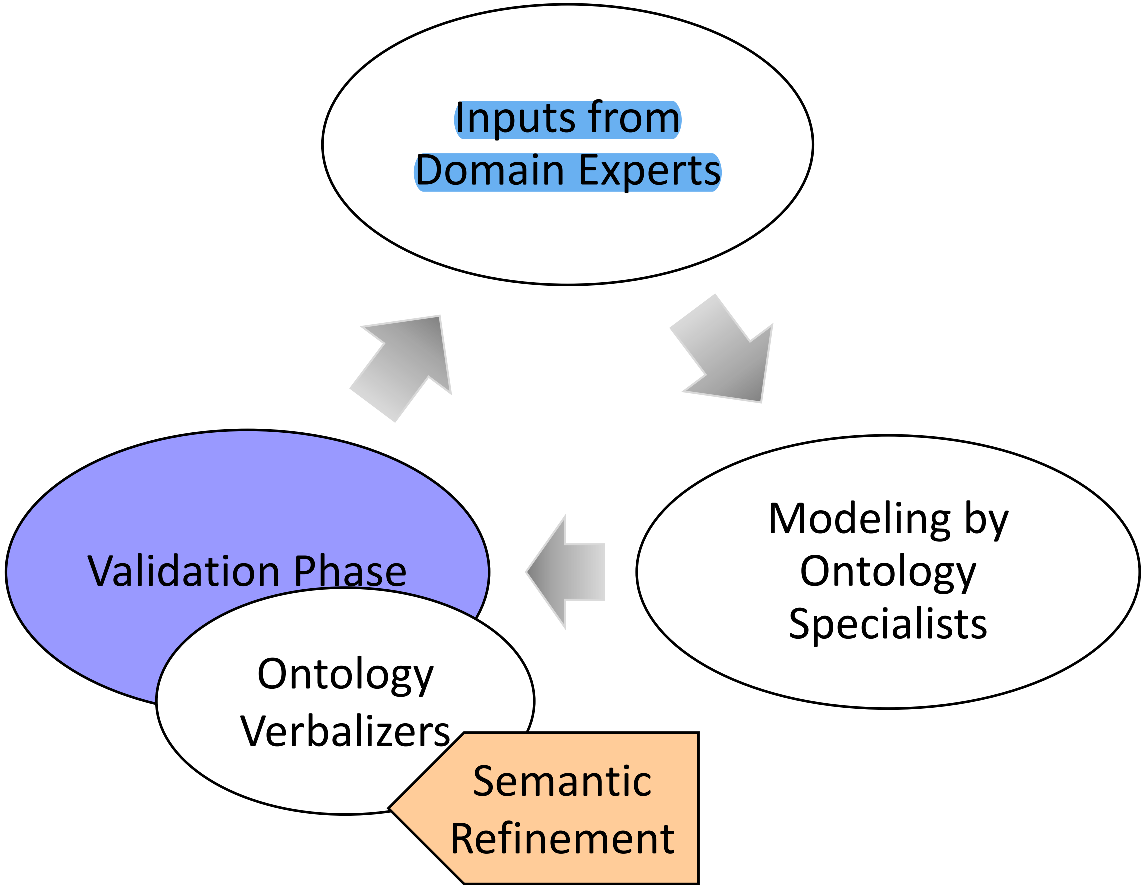
Tool: SWAT

The existing OVs generate texts that have verbatim fidelity to their logical counterparts

These texts tend to have redundancies and repetitions

Proposed System – Applications of Semantic-Refinement

Ontology Development Cycle



At Validation Phase, Domain Experts come together with Ontology Specialists

Ontology Verbalizers (OV) are of great use here

CNLs for OWL: ACE Rabbit SOS

Tool: SWAT

The existing OVs generate texts that have verbatim fidelity to their logical counterparts

These texts tend to have redundancies and repetitions

To remove these, we use Semantic-level Refinement

Results – Applications of Semantic-Refinement

Entity Type	Using Existing Verbalizers (SWAT)	Using Semantic-Refinement
Individual	Bird cherry Oat Aphid: is a disorder, bio-disorder, pest damage and insect damage. It is all the following: has as factor only pest-insect, has as factor only pest, has as factor only organism and has as factor something.	Bird cherry Oat Aphid: is a biotic- disorder, having at least one pest and all its factors are pests.
Concept	Mite Damage: is a disorder, a biotic-disorder and a pest damage. It is all the following: has as factor only organism, has as factor only pest, has as factor only mite pest, has as factor at least one thing.	Mite Damage: is a biotic-disorder, having at least one mite pest and all its factors are mite pests.

Results – Applications of Semantic-Refinement

Entity Type	Using Existing Verbalizers (SWAT)	Using Semantic-Refinement
Individual	Bird cherry Oat Aphid: is a disorder, bio-disorder, pest damage and insect damage. It is all the following: has as factor only pest-insect, has as factor only pest, has as factor only organism and has as factor something.	Bird cherry Oat Aphid: is a biotic- disorder, having at least one pest and all its factors are pests.
Concept	Mite Damage: is a disorder, a biotic-disorder and a pest damage. It is all the following: has as factor only organism, has as factor only pest, has as factor only mite pest, has as factor at least one thing.	Mite Damage: is a biotic-disorder, having at least one mite pest and all its factors are mite pests.

Experiment-1: Degree of understandability of the Verbalized Texts

7 domain experts were asked to rate 41 descriptions as: Poor, Medium and Good.

	Poor	Medium	Good
Using Existing Verbalizers	46%	39%	15%
Using Semantic-Refinement	7%	10%	83%

Results – Applications of Semantic-Refinement

Entity Type	Using Existing Verbalizers (SWAT)	Using Semantic-Refinement
Individual	Bird cherry Oat Aphid: is a disorder, bio-disorder, pest damage and insect damage. It is all the following: has as factor only pest-insect, has as factor only pest, has as factor only organism and has as factor something.	Bird cherry Oat Aphid: is a biotic- disorder, having at least one pest and all its factors are pests.
Concept	Mite Damage: is a disorder, a biotic-disorder and a pest damage. It is all the following: has as factor only organism, has as factor only pest, has as factor only mite pest, has as factor at least one thing.	Mite Damage: is a biotic-disorder, having at least one mite pest and all its factors are mite pests.

Experiment-1: Degree of understandability of the Verbalized Texts

7 domain experts were asked to rate 41 descriptions as: Poor, Medium and Good.

	Poor	Medium	Good
Using Existing Verbalizers	46%	39%	15%
Using Semantic-Refinement	7%	10%	83%

Experiment-2: Usefulness of the generated descriptions in validating Ontology

7 domain experts were asked to rate 41 descriptions as: Valid, Invalid, Don’t Know, Cannot be determined.

	Valid	Invalid	Don’t Know	Cannot be determined
Using Existing Verbalizers	29%	10%	5%	56%
Using Semantic-Refinement	71%	7%	5%	17%

Conclusion and Future works

1. Using Abox-based and Tbox-based methods, we have proposed several new ways to utilize DL Ontologies in developing QA systems

   Future Works

   - Expand the work to other advanced DLs
   - Employ the QG methods in other possible applications other than assessing learners

2. Introduced several semantics-based (ontology-based) measures which are employed for automating various functionalities of an e-learning system.

   Future Work

   - Utilizing these measures in other applications

3. Proposed a theory for controlling the difficulty level of ABox-based MCQs (distractor and stem difficulty).

   Future Work

   - Expand this model to TBox-based MCQs

4. Used Semantic-Refinement to improve Ontology Verbalizers.

   Future Work

   - Expand the technique to more expressive DLs

Publications

Journal Articles

E.V, Vinu; Kumar, P Sreenivasa. **A Novel Approach to Generate MCQs from Domain Ontology: Considering DL Semantics and Open-World Assumption.** Web Semantics: Science, Services and Agents on the World Wide Web, Volume 34(C), Pages 40-54, (May 2015).
URL: http://dx.doi.org/10.1016/j.websem.2015.05.005

E.V, Vinu; Kumar, P Sreenivasa. **Automated Generation of Assessment Tests from Domain Ontologies** Semantic Web: Interoperability, Usability, Applicability. Vol 8, Issue# 6, Pp. 1023-1047, 2017. DOI: 10.3233/SW-170252

E.V, Vinu; Kumar, P Sreenivasa. **Ontology Verbalization using Semantic-Refinement** Semantic Web: Interoperability, Usability, Applicability. ***Under review*** *(communicated on Nov 2016).*
URL: http://www.semantic-web-journal.net/system/files/swj1500.pdf

E.V, Vinu; Kumar, P Sreenivasa. **Difficulty-level Modeling of Ontology-based Factual Questions.** Semantic Web: Interoperability, Usability, Applicability. ***Under review*** *(communicated on Sep 2016).*
URL: http://www.semantic-web-journal.net/system/files/swj1712.pdf

Conference Papers

E.V, Vinu; Kumar, P Sreenivasa. Improving Large Scale Assessment Tests using Ontology-based Approaches. In the Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, FLAIRS-2015, Page 457, (2015).
URL: http://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS15/paper/view/10359

Doctoral Colloquiums

E.V, Vinu; Kumar, P Sreenivasa. Enhancing Assessment Systems Using Ontology Based Methods. Fourth IDRBT Doctoral Colloquium, Dec. 2014.

References

[**IJCAI ’01**] Haarslev, V. and R. Möller, High performance reasoning with very large knowledge bases: A practical case study. In Proceedings of the 17th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI’01. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001. ISBN 1-55860-812-5, 978-1-558-60812-2.
URL: http://dl.acm.org/citation.cfm?id=1642090.1642112

[**AENLP ’99**] Fairon, C., A web-based system for automatic language skill assessment: EVALING. In Proceedings of a Symposium on Computer Mediated Language Assessment and Evaluation in Natural Language Processing, AENLP ’99. Association for Computational Linguistics, Stroudsburg, PA, USA, 1999.
URL: http://dl.acm.org/citation.cfm?id=1598834.1598846

[**AIED ‘09**] Mostow, J. and W. Chen, Generating instruction automatically for the reading strategy of self-questioning. In Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling, Proceedings of the 14th International Conference on Artificial Intelligence in Education, AIED 2009, July 6-10, 2009, Brighton, UK. 2009.
URL: http://dx.doi.org/10.3233/978-1-60750-028-5-465

[**EdAppsNLP ‘05**] Sumita, E., F. Sugaya, and S. Yamamoto, Measuring non-native speakers’ proficiency of English by using a test with automatically-generated fill-in-the-blank questions. In Proceedings of the Second Workshop on Building Educational Applications Using NLP, EdAppsNLP 05. Association for Computational Linguistics, Stroudsburg, PA, USA, 2005.
URL: http://dl.acm.org/citation.cfm?id=1609829.1609839

[**ACL ‘16**] Iulian Vlad Serban, Alberto García-Durán, Caglar Gulcehre, Sungjin Ahn, Sarath Chandar, Aaron Courville, Yoshua Bengio, Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus, In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016.

[**OWLED ‘14**] Alsubait, T., Parsia, B., Sattler, U.: Generating multiple choice questions from Ontologies: Lessons learnt. In: Proceedings of the 11th International Workshop on OWL: Experiences and Directions (OWLED 2014). Volume 1265. (Oct 2014) 73-84

References

[**ISWC ‘15**] Bühmann, L., R. Usbeck, and A.-C. Ngonga Ngomo, ASSESS — Automatic Self-Assessment Using Linked Data. In International Semantic Web Conference (ISWC). 2015.
URL: http://svn.aksw.org/papers/2015/ISWC_ASSESS/Public.pdf.

[**EMNLP ‘05**] J. Brown, G. Firshko, and M. Eskenazi. Automatic question generation for vocabulary assessment. In Proceedings of HLT/EMNLP, pages 819–826, Vancouver, Canada, 2005.

[**CEUR ‘14**] Liu, D. and C. Lin, Sherlock: a semi-automatic quiz generation system using linked data. volume 1272 of CEUR Workshop Proceedings. CEUR-WS.org, 2014.

[**SW ‘08**] Peroni, S., E. Motta, and D’Aquin, Mathieu, Identifying key concepts in an ontology, through the integration of cognitive principles with statistical and topological measures. In the Semantic Web, Volume 5367 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-89703-3.
URL: http://dx.doi.org/10.1007/978-3-540-89704-0_17. 10.1007/978-3-540-89704-0_17.

[**ISWC ‘14**] Carvalho, D., C. Calli, A. Freitas, and E. Curry, EasyESA: A Low-effort Infrastructure for Explicit Semantic Analysis (Demo). In 13th International Semantic Web Conference (ISWC 2014). Springer, Rival del Garda, 2014.
URL: http://www.edwardcurry.org/publications/easyesa_demo_2014.pdf

[**ACL ’17**] Du, Xinya and Shao, Junru and Cardie, Claire, Learning to Ask: Neural Question Generation for Reading Comprehension, Association for Computational Linguistics (ACL). 2017

[**WWW ‘15**] Seyler, D., M. Yahya, and K. Berberich, Generating quiz questions from knowledge graphs. In Proceedings of the 24th International Conference on World Wide Web, WWW ’15 Companion. ACM, New York, NY, USA, 2015. ISBN 978-1-4503-3473-0.
URL: http://doi.acm.org/10.1145/2740908.2742722