CHAPTER 2

LITERATURE SURVEY

This chapter provides a detailed survey on various reasoning techniques and several DL reasoners. It also discusses a wide range of semantic relationships that help to provide complete knowledge about a domain and the various application areas in which semantic relationships are exploited advantageously. This chapter elaborates the role of semantic relationships in reasoning, the techniques used to extract the semantic relationships, the algorithms which are used to classify the semantic relationships and the methods used to achieve the reasoning with semantic relationships.

Most of the information retrieval systems retrieve the relevant documents which match with the given keyword(s). The retrieved relevant documents are ranked with the use of number of occurrences of the given keyword(s) in the set of documents. Little effort has been made to retrieve and display the most relevant documents with the use of semantic relationships. This chapter also exploits the role of semantic relationships in information retrieval and ranking. The prospect of applying semantic relationships to the question generation system can revolutionize the learning experience. It is an important component in intelligent tutoring systems. This chapter also provides detailed survey on question generation systems.

2.1 DL REASONING

Reasoning is accomplished through logic. Predicate Logic and Description Logic are widely used to achieve reasoning process. DL is more expressive for representing knowledge with decidability than predicate logic (Baader et al., 2008). The various DL based reasoning techniques have been used to provide the reasoning services such as satisfiability checking, subsumption, consistency checking, instance checking, equivalence and disjointness. Based on the functionalities of reasoning techniques, it could be classified into direct and indirect reasoning. Various DL based reasoning techniques (Baader, 2003) are shown in Figure 2.1. In indirect reasoning process, the given problem is translated into equivalent problem and solves it by using appropriate technology. But in direct reasoning process, it applies some approaches/algorithms on
logics to derive inferences. Generally, direct reasoning is achieved by constructing a model or using proof derivation method. Resolution based approaches and structural approaches are based on proof derivation method. Automata based approaches and tableau based approaches are used to perform reasoning which is based on model construction.

2.1.1 DL based Reasoning Techniques

2.1.1.1 Resolution based Approaches

Resolution based approaches is applied to modal logic which include operators for expressing modality. In (De Nivelle et al., 2000), three resolution design procedures such as ordered resolution design procedure, ordering resolution for guarded fragment and selection based resolution were described. Ordered resolution design procedure
uses clause fragment i.e. DL* (Hustadt and Schmidt, 2000) and guarded fragment (Andréka et al., 1995). In this, decision procedures were constructed for the fragments and generic decision procedures were obtained for modal logic (Li et al., 2007). The effective design procedure was constructed and decidability was achieved in ordered resolution guarded fragment (De Nivelle, 1998). The decidability was achieved by selecting negative literals and then more generalized models were constructed automatically in selection based resolution method (Hustadt and Schmidt, 2002).

### 2.1.1.2 Structural Approaches

Structural approaches were used to test the subsumption among concepts (Baader et al., 2008). It could be used to compare the syntactic structure of the concepts. In this approach, subsumption can be decided within polynomial time. The disadvantage is that it can’t detect all the subsumption since it can’t allow all boolean operations, constructor conjunction and value restrictions. The logic $\mathcal{FL}_o$ (Frame Language without limited existential quantification) supports concept conjunction, value restriction and universal concepts and $\mathcal{EL}$ (DL with existential quantification) which supports concept conjunction, value restriction and universal concepts. Subsumption for $\mathcal{FL}_o$ and $\mathcal{EL}$ is proved without TBox and with TBox respectively. The subsumption for $\mathcal{FL}_o$ is in coNP (Nebel, 1990). The subsumption in $\mathcal{EL}$ with TBox is polynomial (Brandt, 2004). In order to prove it, TBox is translated into normalized graph with completion rules and subsumption relationships are extracted.

### 2.1.1.3 Automata based Approaches

In automata based approaches, the design procedures were obtained by reducing satisfiability into the emptiness problem (Vardi and Wolper, 1986). It was derived with the use of Streett’s techniques (Streett, 1982). But it took doubly exponential time for constructing the automata. Buchi automata (Büchi, 1990) was used to construct the infinite trees within polynomial time. Subtree automata (Vardi and Wolper, 1986) was introduced to check if there exists a finite subtrees for every node in the tree. Looping automata (Vardi and Wolper, 1994) dealt with fixed arity infinite trees. It took polynomial time for testing emptiness. Amorphous tree automata (Bernholtz and Grumberg, 1993) is a complicated model which dealt with arbitrary
branch. Generally the automata based approaches are used to check modal satisfiability.

2.1.1.4 Tableau based Approaches

Tableau method is introduced by (Beth, 1955). In this method, a tree is constructed with initial set of formulae and tried to find the contradiction among the given formulae. It expands the formulae with conjunction and disjunction and applies reduction; finally it contains only atomic formulae and their negations. It can be used to determine the satisfiability of the formulas for various logics (Girle, 2000). Tableau based approaches are also used to solve the knowledge base consistency problem (Baader et al., 2008). Two algorithms were proposed using tableau based design procedures such as Tableau Algorithm and Hypertableau algorithm in order to perform consistency checking.

Tableau algorithm and Hypertableau algorithm help to construct the models and extended its reasoning services to satisfiability, subsumption, disjointness, equivalence and instance checking. The algorithms are applied to DL to infer the conclusions. In this work, the tableau based approaches are used for the realization of the proposed SemPER reasoning algorithm. Next section describes some of the basic definitions of Description Logic which are used throughout this thesis.

2.2 BASIC DEFINITIONS

The domain knowledge is represented using Description Logic which helps to reason out the knowledge and manipulate the knowledge. It is stored in the knowledge base which consists of TBox and ABox. TBox includes the terminology which describes set of concepts, properties and relationships between the concepts and ABox includes the assertions which describe the instances associated with the concepts. The concepts, properties and relations are represented using DL. The expressiveness of DL is indicated by the use of constructors (Baader and Nutt, 2003) such as negation, union, intersection, universal restriction etc. Some of the expressiveness of DL is given below:

\[ \mathcal{AL} \quad - \quad \text{Attributive Language} \]
\[ \mathcal{FL} \quad - \quad \text{Frame based DL} \]
\( \mathcal{EL} \) - DL with existential quantification
\( \mathcal{F} \) - Functional properties
\( \mathcal{E} \) - Full existential quantification
\( \mathcal{U} \) - Concept Union
\( \mathcal{C} \) - Concept Negation
\( \mathcal{H} \) - Role Hierarchy
\( \mathcal{O} \) - Nominals
\( \mathcal{Q} \) - Qualified cardinality restrictions
\( \mathcal{N} \) - Cardinality Restrictions
\( \mathcal{I} \) - Inverse properties
\( \mathcal{R} \) - Limited Role Restrictions
\( \mathcal{Q} \) - Data type properties, data type values and types
\( \mathcal{S} \) - \( \mathcal{ALC} \) with transitive roles
\( \mathcal{FL} \) - \( \mathcal{FL} \) without role restriction
\( \mathcal{FL}_0 \) - \( \mathcal{FL} \) without limited existential quantification

\( \mathcal{ALC} \) is an important description logic which was introduced by (Schmidt-Schauß and Smolka, 1991). The Ontology Web Language (OWL) provides expressiveness for \( \mathcal{SHOIN}(\mathcal{Q}) \). The domain knowledge is represented using ontology and its corresponding DL notations (Baader and Nutt, 2003) as follows:

\( \top \) - Universal concepts
\( \bot \) - Empty Concept
\( \cup \) - Union / Disjunction of concepts
\( \cap \) - Intersection / Conjunction of concepts
\( \neg \) - Negation/ Complement of concept
\( \forall \) - Universal restriction
\( \exists \) - Existential restriction
\( \equiv \) - Equivalent Concepts
\( : \) - Concept / Role assertion
\( \subseteq \) - Concept inclusion
In order to define semantics of \textit{ALCNU} logic, an interpretation \( I \) which consists of non-empty set \( \Delta' \) and an interpretation function which assigns to every concept \( C \) such that \( C \subseteq \Delta' \) and to every relation \( R \) such that \( R \subseteq \Delta' \times \Delta' \) are considered. The interpretation function is extended to domain knowledge by the following:

Consider ‘C’ and ‘D’ be the concepts, ‘a’ and ‘b’ be the instances and ‘R’ be the relations.

\[
\begin{align*}
T' & \rightarrow \Delta' \\
\bot^I & \rightarrow \phi \\
\neg C^I & \rightarrow \Delta' \setminus C^I \\
(C \cup D)^I & \rightarrow C^I \cup D^I \\
(C \cap D)^I & \rightarrow C^I \cap D^I \\
(\forall R. C)^I & \rightarrow \{ a \in \Delta' \mid \forall b. (a, b) \in R^I \rightarrow b \in C^I \} \\
(\exists R. C)^I & \rightarrow \{ a \in \Delta' \mid \exists b. (a, b) \in R^I \land b \in C^I \} \\
(\geq n R)^I & \rightarrow \{ a \in \Delta' \mid \{ b \mid (a, b) \in R^I \} \geq n \} \\
(\leq n R)^I & \rightarrow \{ a \in \Delta' \mid \{ b \mid (a, b) \in R^I \} \leq n \} \\
(R : a)^I & \rightarrow \{ b \in \Delta' \mid (b, a^I) \in R^I \}
\end{align*}
\]

The above mentioned notations are used in the representation and interpretation of this research work’s proposal of reasoning methodology. Next section explains various existing reasoning algorithms in detail.

\textbf{2.3 \textsc{Reasoning Algorithms}}

The reasoning algorithms infer implicit knowledge and provide the reasoning services with TBox and ABox as described in Chapter 1. Checking satisfiability of concepts is important in TBox reasoning and consistency checking is accomplished with ABox (Baader, 2003). The tableau based approaches are widely used for reasoning (Baader, 2003).
2.3.1 Tableau algorithm

Tableau algorithm is applied by (Schmidt-Schauß and Smolka, 1991) for checking satisfiability in $\mathcal{ALC}$. This algorithm is extended by (Hollunder and Baader, 2011) (Baader and Sattler, 1999) (Buchheit and Hollunder, 1993) for $\mathcal{ALCN}$. The satisfiability algorithm using tableau approach is applied to $\mathcal{S}$ i.e. $\mathcal{ALC}$ with transitive roles by (Baader, 2011) (Haarslev et al., 2001). Tableau algorithms are also extended to check consistency for ABoxes (Haarslev and Möller, 2000) (Donini and Massacci, 2000). In this algorithm, negation has been used to reduce subsumption to (un)satisfiability. It applies various transformation rules to check satisfiability which is shown in Figure 2.2 (Baader, 2003).

$$
\lor \text{Rule : If } A \text{ contains } C_i(x) \cap C_j(x), \text{ but it does not contain } C_i(x) \text{ and } C_j(x) \\
\text{then } A = A \cup \{C_i(x), C_j(x)\}
$$

$$
\land \text{Rule : If } A \text{ contains } C_i(x) \cap C_j(x), \text{ but neither } C_i(x) \text{ nor } C_j(x) \\
\text{then } A = A \cup \{C_i(x), C_j(x)\}
$$

$$
\forall \text{Rule : If } A \text{ contains } (\forall R.C)(x) \text{ and } R(x, y), \text{ but it does not contain } C(y) \\
\text{then } A = A \cup \{C(y)\}
$$

$$
\exists \text{Rule : If } A \text{ contains } (\exists R.C)(x), \text{ but there is an individual } y \text{ such that } C(z) \text{ and} \\
R(x, z) \text{ are in } A \\
\text{then } A = A \cup \{C(y), R(x, y)\}
$$

$$
\geq \text{Rule : } A \text{ contains } (\geq n R)(x), \text{ and there are no individual names } z_1, ..., z_n \text{ such that} \\
R(x, z_i) \text{ for } (1 \leq i \leq n) \text{ and } z_i \neq z_j \text{ for } (1 \leq i \neq j \leq n) \text{ are in } A \\
\text{then } A = A \cup \{R(x, y_i) \mid 1 \leq i \leq n\} \cup \{y_i \neq y_j \mid 1 \leq i \neq j \leq n\},
$$

where $y_1, ..., y_n$ are distinct individual names not occurring in $A$.

$$
\leq \text{Rule : } A \text{ contains distinct individual names } y_1, ..., y_n \text{ such that } (\leq n R)(x), \text{ and } R(x, y_i), \\
..., R(x, y_n) \text{ are in } A \text{ and } y_i \neq y_j \text{ is not in } A \text{ for some } i, j \leq n \text{ and } y_i \neq y_j \text{ not in } A, \text{ the} \\
A\text{Box } A_{y_i} := [y_i/y_j] \text{ A is obtained from } A \text{ by replacing each occurrence of } y_j \text{ by } y_i.
$$

Fig 2.2. Rules for Satisfiability Checking

The systems such as KRIS (Baader and Hollunder, 1991) and CRACK (Bresciani et al., 1995) have employed tableau algorithms and it is acceptable (Baader et al., 1994). The highly optimized reasoners such as FaCT (Horrocks, 1998), FaCT++ (Tsarkov and Horrocks, 2006), RacerPro (Haarslev and Müller, 2001) and Pellet (Sirin et al.,
2007) reasoner provide better results with tableau algorithms. It can also be applied to modal logic (Patel-Schneider, 1998) (Horrocks and Patel-Schneider, 1999) (Haarslev and Möller, 2000).

2.3.2 Hypertableau algorithm

Tableau algorithm performs consistency check by constructing model for the given knowledge base. The constructed model is extremely large even for small ontologies and also it constructs many models before concluding that no model is possible (Shearer et al., 2008). This leads to invention of Hypertableau algorithm. Hypertableau algorithm reduces the number of possible models by using anywhere blocking strategy (Motik et al., 2007). This algorithm is developed with the features of resolution and tableau (Motik et al., 2009) and it is related to hypertableau (Baumgartner et al., 1996) and hyperresolution (Robinson, 1965) methods. This algorithm eliminates or-branching and and-branching strategy which is used in tableau algorithm. Hermit (Shearer et al., 2008) reasoner employs anywhere blocking strategy which helps to limit the number of models. Hence, it classifies complex ontologies very fastly as compared to other reasoners such as Pellet and Fact++.

2.4 DL REASONERS

The reasoners are used to infer logical consequences from the set of known facts which is stored in knowledge base. The reasoners are performing inferences on the basis of description logic. The reasoners such as FaCT, FaCT++, Pellet and Hermit are freely available and RacerPro is a commercial reasoner.

2.4.1 FaCT

FaCT (Fast Classification of Terminologies) is used to check satisfiability for modal logic (Horrocks, 1998)(Horrocks, 1999) using tableau algorithm. This system includes two reasoners, one for the logic SHF and another for the logic SHIQ. This reasoner is also used to test subsumption. It is implemented in LISP.

2.4.2 FaCT++

FaCT++ is an advanced version of FaCT reasoner (Tsarkov and Horrocks, 2006). This reasoner employs tableau algorithm for SHOIQ logic. The optimizations such
as normalization, simplification, absorption, semantic branching, cycle elimination and model merging are employed for enhancing the performance of the reasoners. FaCT++ can be used to provide reasoning services for OWL ontologies with DIG (DL Implementation Group) interface. It is implemented in C++. This reasoner makes an effort for satisfiability checking, subsumption testing and consistency testing.

2.4.3 Pellet

Pellet reasoner (Sirin et al., 2007) is developed to meet the requirements of Web Ontology Language (OWL). The reasoner which supports all OWL-DL is Pellet. It widely supports $\text{SROIQ}(\Delta)$ logic. It is widely acceptable for its good performance, middleware support and various unique features such as qualified cardinality restrictions, user defined datatypes, object properties, etc. Pellet provides various reasoning services with the use of DIG and API bindings for RDF/OWL. It integrates tableau algorithm and includes the features of conjugative query answering, $\varepsilon$-connection reasoning, rule support and axiom pinpointing. It also implements various optimizations such as normalization, absorption, semantic branching, model merging, absorption and backjumping. It is implemented in Java.

2.4.4 Hermit

Hermit reasoner (Shearer et al., 2008) applies Hypertableau algorithm to classify the complex ontologies very fastly as compared to other reasoners. It supports OWL-DL. It incorporates anywhere blocking strategy to limit the model. It also integrates many optimization techniques to improve the performance. It is implemented in Java.

2.4.5 RacerPro

Racer (Renamed ABox and Concept Expression Reasoner) was introduced by (Haarslev and Müller, 2001) to provide reasoning services for OWL-DL. Later it was extended and named as RacerPro (Haarslev et al., 2012). It can be used for managing OWL ontologies. It implements tableau algorithm for expressive DL. It provides services for many TBoxes and ABoxes. It provides all reasoning services and also provides explanation facility. The explanation facility has been implemented for SUMO ontology (Pease and Niles, 2002). It uses built-in data structures of tableau reasoners for providing explanation facility. It supports nRQL (new Racer Query
Language) for providing conjugative query reasoning. It is implemented in Java (JRacer), LISP (LRacer) and also in C. The comparison of various reasoners is given in Table 2.1.

Table 2.1. Comparison of Reasoners

<table>
<thead>
<tr>
<th>Features</th>
<th>Fact++</th>
<th>Pellet</th>
<th>Hermit</th>
<th>RacerPro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressivity</td>
<td>$SHOIQ(D)$</td>
<td>$SRIOIQ(D)$</td>
<td>$SHOIQ+$</td>
<td>$SRIQ(D-)$</td>
</tr>
<tr>
<td>Reasoning Algorithm</td>
<td>Tableau</td>
<td>Tableau</td>
<td>Hypertableau</td>
<td>Tableau</td>
</tr>
<tr>
<td>SWRL support</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Classification</td>
<td>Slow</td>
<td>Average</td>
<td>Fast</td>
<td>Average</td>
</tr>
<tr>
<td>Consistency Checking</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DIG</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

This section discussed about various reasoning techniques, algorithms and DL reasoners and its unique features. From the literature, it is noted that the reasoners perform the reasoning services only based on the concepts which are represented in the knowledgebase. However, the semantic relationships provide more semantics to the domain knowledge. As it is evident from the above discussion about reasoners, this research work concentrated on a reasoning methodology based on semantic relationships, which may be used to develop a complete reasoner system. Next section discusses various semantic relationships which are exploited from literature to modern knowledge based systems.

### 2.5 SEMANTIC RELATIONSHIPS

Semantic relationships are the key component in the representation of the domain knowledge. There are wide ranges of semantic relationships that have been exploited from literature to modern computational theory. The semantic relationships are used to structure the concepts and to provide well defined meaning to realize more about concepts. It is also used to discover implicit knowledge and helps to develop structured information from unstructured information. The relations such as class inclusion, similar, case, contrast and part-whole relations are widely deliberated by
(Chaffin and Herrmann, 1984). Totally 31 semantic relationships are identified and they are clustered hierarchically.

In (Sheth et al., 2004), the semantic relationships are classified based on the information content. It has been broadly classified into content independent relationships and content dependent relationships. The content independent relationships are used to relate the performance, scalability, location of the document, properties, etc. The content dependent relationships are based on the information and/or with its representation. They have classified content dependent relationships into direct content dependent relationships and content descriptive relationships. The direct content dependent relationships are fuzzy in nature. The content descriptive relationships are used to specify real world entities. It can be developed with the use of taxonomies, ontologies and thesaurus. These relationships are used to associate concepts within a domain or across multiple domains which can be called as intra-domain relationships and inter-domain relationships respectively. The mappings of the semantic relationships between concepts are accomplished through various proportions such as arity, cardinality, structural composition, proximity, data properties, etc. The identified semantic relationships have been used for document enhancement.

The relationship between the concepts in Government ontology is constructed using E-Government thesaurus (Gao and Zhao, 2008). The relationships such as equivalent, hierarchical and related relationships are exhibited between the terms. The semantic relationships such as kind-of, part-of, instance-of and attribute-of relationships are defined between the concepts in ontology. In this paper, the semantic relationships are classified into 15 categories such as genus-species, part-of, population-individual, superior- subordination, synonym, antonym, similarity, causal, agent, application relation, contradiction, timing, spatial, instance-of and attribute relations. The relationships are helpful for the construction of Government ontology and also help to provide more detailed descriptions about the concepts.

Semantic relationships that exist in the database provide more semantics for an application in database management systems. It is widely used to accommodate real world knowledge (Storey, 1993). The relationships in DBMS are commonly referred as data abstractions. The most common abstractions in DBMS are inclusion,
association and aggregation. The relationships also represent minimum and maximum cardinality among entities in the database. Class, meronym and spatial relations are identified in inclusion semantic relations. Meronym relations are distinguished based on its functionality, separability, similarity and timeliness. Some of the meronym relations are component-object, feature-event, member-collection, portion-mass, phase-activity, place-area and stuff-object. Other widely used relations are possession, attachment and attribution, synonym, antonym and case relations involving agent and actions.

The semantic relationships also play a major role in various domains such as biology, chemistry, etc. In OBO (Open Biological Ontologies), the relations are classified into foundational, spatial, temporal and participation relations (Smith et al., 2005). The defined relationships in OBO are located_in, adjacent_to, contained_in, derives_from, transformation_of, preceded_by, has_participant and has_agent relations. The GO (Gene Ontology) provides controlled vocabulary which can be applied to several domains of cellular and molecular biology (Harris et al., 2004). It incorporates the relations such as is_a, part_of and regulates relations.

SO (Sequence Ontology) uses mereological relations, temporal and spatial interval relations which was described by (Mungall et al., 2011). The part_of, has_part, integral_part_of and has_integral_part relationships are included in mereological relations. The temporal relations defined in SO are transcribed_from, transcribed_to, processed_from, processed_to, ribosomal_translation_of and ribosomal_translation_to. The relations like contains, overlaps, adjacent_to, started_by, start, finishes, maximally_overlaps, disconnected_from and is_consecutive_sequence_of are incorporated in spatial interval relations. The widely used semantic relationships in several domains are shown in Table 2.2.

Table 2.2. Classification of relationships based on Applications

<table>
<thead>
<tr>
<th>RELATIONSHIPS</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Literature</td>
<td></td>
</tr>
<tr>
<td>Contrast relation</td>
<td>Finding similarity and diversity of the</td>
</tr>
<tr>
<td>Contradictory antonyms, Contrary antonyms, Directional, Reverse antonyms</td>
<td></td>
</tr>
</tbody>
</table>
### Similar relation
- Dimensional similarity relation, Attribute similarity relation

### Class inclusion relation

### Case relation
- agent –action relations, agent-instrument relation, agent-object relation

### Part- whole relation
- Heterogeneous relations
  - Functional relations, spatial relations
- Homogeneous relations
  - Individual, group relations

### Content independent relationship

### Content dependent relationship
- Direct content dependent relationship
- Content descriptive relationship
  - Direct semantic relationship, Complex transitive relationship, Interdomain multi-ontology relationship, Semantic proximity relationship

### Spatial relations
- located_in, overlap, contained_in, adjacent_to

### Temporal relations
- transformation_of, derives_from, preceded_by

### Participation relations
- has_participant, has_agent
Mereological relations
- part_of, has_part, integral_part_of, has_integral_part

Spatial interval relations
- contains, overlaps, adjacent_to, started_by, start, finishes, maximally_overlaps, disconnected_from, is_consecutive_sequence_of

Temporal relations
- transcribed_from, transcribed_to, processed_from, processed_to, ribosomal_translation_of, ribosomal_translation_to

Develop sequence ontology to provide interoperability

In addition to the deliberated domains, the relationships also play a crucial role in knowledge engineering to enhance the knowledge. The domain knowledge cannot provide more semantics without semantic relationships. Apart from the discussed relations, many other relations like active, associative, causal, polysemy, paradigmatic, possession, kinship, theme, predicate, measure, certainty, etc have been proposed by many authors based on their perception and applications. Figure 2.2 shows the primary classification of semantic relationships. The study of these semantic relationships indicates that any relations identified have to fit into any one of this category. Among these, the widely used relations are Hyponym, Hypernym, Meronym Cause-effect, Spatial and Temporal relations. Former four relations are commonly used in all domains. Spatial relations are widely used in Geo-spatial based applications. Temporal relations are usually used where the ordering of the events are essential.

**Fig 2.3. Classification of Semantic Relationships**
The semantic relationships are used to enhance the open world semantics. The domain knowledge is developed with the motive of resolving specific problem. The concepts, their relationships, properties and instances are incorporated in the knowledge to achieve the specific goal. The realization of complete domain knowledge could not be achieved only with concepts. The inference about the semantic relationships in the represented knowledge will help the user to understand more about particular domain. The properties of the semantic relationship are also crucial in inferring the consequences. Generally, inference is carried out with the unique semantic relationship with transitive properties. Thus, there is a need of reasoning algorithms which infer novel semantic relationships.

This section described the importance of semantic relationships and several semantic relationships in various domains. The next section describes reasoning based on semantic relationships.

2.6 REASONING WITH SEMANTIC RELATIONSHIPS

The reasoning algorithms are implemented across various concepts and same kind of relationships in domain knowledge. It checks satisfiability and consistency of the given concepts and infers subsumption relations. Very few researchers make an effort to retrieve novel relationships. Still, the work is towards “is-a” and “part-of” relations only. Inference about spatial, temporal and other relations is ubiquitous (Carreiras and Santamaría, 1997) (Bittner et al., 2004) (Goodwin and Johnson-Laird, 2005).

Generally, the spatial descriptions are represented in the form of linguistic representations and mental models (Byrne and Johnson-Laird, 1989). The mental model is appropriate only for spatial domain (Rips, 1994). (Carreiras and Santamaría, 1997) conducted two experiments which were designed to predict the relations using model theory of reasoning and the formal rules of inference theories. In these experiments, they have used spatial and non-spatial relations. It comprises the relations such as higher than, in front of, to the right of, to the left of, more than and less than. Many researches concentrate on spatial and temporal reasoning. Most of the spatial and temporal relations satisfy transitive properties and it enables the system to infer the relationships more easily.

(Bittner, 2002) defined qualitative relations between temporal regions. In this, Jointly Exhaustive and Pairwise Disjoint (JEPD) relations are defined. He used approximate
theory to reason the temporal location of an event or process with partition of the time line. Approximate reasoning is used to improve the quality and robustness of the inference. In this, four types of topological relations are discussed and it has provided formal basis about approximate location with time using approximate temporal reasoning.

(Bittner et al., 2004) provided inferences about foundational relations between individuals, universals and collections. They have considered individuals as parts, universals as instances and collections as members. Based on it, they have identified the relations which are mentioned in Table 2.3. In this, they have used FOL and provided an axiomatic theory which infers the relations between universal, collection and individuals.

Table 2.3. Relations between Individuals, Universal and Collections

(Bittner et al., 2004)

<table>
<thead>
<tr>
<th>Term1</th>
<th>Term2</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>Individual</td>
<td>part-of</td>
</tr>
<tr>
<td>Individual</td>
<td>Universal</td>
<td>Instance-of</td>
</tr>
<tr>
<td>Individual</td>
<td>Collection</td>
<td>Member-of</td>
</tr>
<tr>
<td>Universal</td>
<td>Universal</td>
<td>hypernymy</td>
</tr>
<tr>
<td>Collection</td>
<td>Individual</td>
<td>Partition-of</td>
</tr>
<tr>
<td>Collection</td>
<td>Universal</td>
<td>Extension-of</td>
</tr>
<tr>
<td>Collection</td>
<td>Collection</td>
<td>Inclusion</td>
</tr>
</tbody>
</table>

(Goodwin and Johnson-Laird, 2005) presented model theory based reasoning over spatial and temporal relations. It is infeasible if, the semantics of the relations is considered based on properties. They have identified spatial relations such as in the same as, beyond, not beyond, next in line to, directly on top of, nearest to, next to, on the right of, at. It holds the logical properties such as transitive, intransitive, nontransitive, symmetry, asymmetry, non-symmetric. The properties of the relations play a key role in knowledge inference. Lack of these logical properties does not imply that a relation yields no inferences. But, it differs based on the context. Authors have mentioned about the temporal relation Happens – Before which satisfies the transitive property leading to achieve correct and simple reasoning. Table 2.4 shows some of the relations (Goodwin and Johnson-Laird, 2005) and its corresponding
inferences. The reasoning is accomplished through mental model theory. In this, the individuals have constructed a model for relational reasoning. In this, the knowledge is represented with predicate calculus.

Table 2.4. Relations, its properties and its Inferences

(Goodwin and Johnson-Laird, 2005)

<table>
<thead>
<tr>
<th>Relations (R)</th>
<th>Properties</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the same place as</td>
<td>Transitive</td>
<td>a R b, b R c → a R c</td>
</tr>
<tr>
<td>Next in line to</td>
<td>Intransitive</td>
<td>a R b, b R c → ¬(a R c)</td>
</tr>
<tr>
<td>Next to</td>
<td>Nontransitive</td>
<td>a R b, b R c → No Conclusion</td>
</tr>
<tr>
<td>Next to</td>
<td>Symmetric</td>
<td>a R b → b R a</td>
</tr>
<tr>
<td>Taller than</td>
<td>Asymmetric</td>
<td>a R b → ¬(b R a)</td>
</tr>
<tr>
<td>In the same place as</td>
<td>Reflexive</td>
<td>a R a</td>
</tr>
<tr>
<td>Next to</td>
<td>Irreflexive</td>
<td>a R ¬a</td>
</tr>
<tr>
<td>Happens before</td>
<td>Transitive</td>
<td>a R b, b R c → a R c</td>
</tr>
<tr>
<td>Taller than</td>
<td>Transitive</td>
<td>a R b, b R c → a R c</td>
</tr>
</tbody>
</table>

Gene ontology (Harris et al., 2008) contains many concepts and relationships between those concepts. The ontology is represented in OWL. It has functional relations such as part_of, is_a, has_part, occurs_in and regulates. In this ontology, reasoning can be performed over is-a, part-of and regulates relations using logical inferences which is shown in Table 2.5.

Table 2.5. GO Relations and its Inferences

(Harris et al., 2008)

<table>
<thead>
<tr>
<th>Relation (R1)</th>
<th>Relation (R2)</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>is-a</td>
<td>is-a</td>
<td>a R b, b R c → a R c</td>
</tr>
<tr>
<td>is-a</td>
<td>part-of</td>
<td>a R1 b, b R2 c → a R2 c</td>
</tr>
<tr>
<td>is-a</td>
<td>regulates</td>
<td>a R1 b, b R2 c → a R2 c</td>
</tr>
<tr>
<td>part-of</td>
<td>is-a</td>
<td>a R1 b, b R2 c → a R1 c</td>
</tr>
<tr>
<td>part-of</td>
<td>regulates</td>
<td>a R1 b, b R2 c → a R2 c</td>
</tr>
<tr>
<td>part-of</td>
<td>part-of</td>
<td>a R b, b R c → a R c</td>
</tr>
<tr>
<td>regulates</td>
<td>is-a</td>
<td>a R1 b, b R2 c → a R1 c</td>
</tr>
<tr>
<td>regulates</td>
<td>part-of</td>
<td>a R1 b, b R2 c → a R1 c</td>
</tr>
<tr>
<td>regulates</td>
<td>regulates</td>
<td>a R b, b R c → a R c</td>
</tr>
</tbody>
</table>
(Kaza, 2008) considered the relationships between the action and their consequences such as spatial, temporal and functional relationships. They have used heuristic reasoning approach to perform spatial and temporal reasoning. In this, they have considered the relationship with partial ordering i.e. transitive, reflexive and antisymmetric.

(Goodwin and Johnson-Laird, 2013) discussed about the Boolean relations and, or and not. It is widely used to create the new concepts. They have employed mental model theory for the reasoning and tried to investigate the possibilities to which premises refer and derive conclusions accordingly (Johnson-Laird, 2010).

It has been noted that reasoning can be carried out in a well-defined manner in case, if the relation holds transitive property and sometimes with reflexive and symmetry properties. It is complex in case of other relational properties. Hence, the system is required to discover the semantic relationships between concepts even it persists other properties such as intransitive, irreflexive and antisymmetric properties. Recognizing the semantic relationships is essential for interpreting the structure and evolution of certain concepts.

This section depicted the research works aimed at reasoning with semantic relationships. Next section describes the techniques used to extract the semantic relationships from various corpuses which will be useful for reasoning.

2.7 RELATIONSHIP EXTRACTION

The goal of extracting semantic relationship is to learn relations from structured and unstructured information. In structured information, the information is represented using various representation languages. Extracting semantic relations between concepts in unstructured information is a crucial task in NLP. The turnover of unstructured information into structured information is achieved by annotating semantic information. The extracted relations can be used to answer the questions, learn the interaction between entities and to build the database which helps the user to explore the knowledge. The extracted relations are extensively useful in gene-disease interaction, protein-protein interaction. The relations are extracted from unstructured information using supervised approach or unsupervised approach. Supervised approaches include feature based approaches and kernel based approaches.
Unsupervised approaches include various techniques such as Snowball (Agichtein and Gravano, 2000), DIPRE (Brin, 1999), KnowItAll (Etzioni et al., 2004) and TextRunner (Yates et al., 2007). Relation extraction approaches are shown in Figure 2.3.

Fig 2.4. Relationship Extraction Approaches

2.7.1 SUPERVISED APPROACHES

The prominent increase in the amount of information available in web makes a necessity of extracting information from text or structured information. Extracting semantic relationships from the extracted information helps to enhance the user’s perception towards the particular domain. The relations are extracted based on features and kernel strings in supervised approaches.

2.7.1.1. Feature Based Approaches

(Miller et al., 2000) applied information retrieval system to select the articles from corpus. The retrieved articles are annotated with augmented tree structures. The annotation is accomplished manually. Trees are generated with the models which were proposed by (Collins, 1997). (Kambhatla, 2004) proposed maximum entropy
model for extracting relations from the pair of entities within each sentence. In the feature based relation extraction, the syntactic features are extracted from the given sentences which include entities, type of entities, words between entities, number of words between entities and corresponding path in the dependency tree. The relations are extracted from the set of features for small feature space only.

2.7.1.2. Kernel Based Approaches

String kernels are used for relation extraction which was described by (Lodhi et al., 2002). (Cumby and Roth, 2003) proposed an approach to construct kernels for learning relationships from structured data. It could be implemented with kernel based algorithms like Perceptron or SVM (Support Vector Machine). It is efficient even when the expanded feature space is exponential in size. (Zelenko et al., 2003) proposed a method to extract the relations from shallow parse tree with the use of kernel functions. In this, they have extracted person-affiliation and organization-location relations from the text. They have used tree kernels in conjunction with SVM and Voted Perceptron learning algorithm to extract the relations. The kernel is used to calculate the similarity between two shallow parse tree structures for the given two entities. Shallow parse tree is more reliable and robust to extract the relations.

(Culotta and Sorensen, 2004) extended the algorithm which was proposed by (Zelenko et al., 2003). They have estimated the kernel function with dependency parse tree and provided better accuracy. (Zelenko et al., 2003) and (Culotta and Sorensen, 2004) explored implicit feature space for computing kernel functions.

(Bunescu and Mooney, 2005) proposed an approach to extract the relationship between the entities by computing the shortest path between the constructed dependency parse tree. (Mooney and Bunescu, 2005) proposed generalized subsequence kernel for extracting semantic relationship between two entities. They have used the patterns such as fore-between, between, between-after for extracting the relations between entities in the given sentences. (Zhao and Grishman, 2005) applied kernel functions by combining information from different syntactic resources and it provided better accuracy for relation detection. (Zhou et al., 2007) proposed context sensitive tree span and context sensitive convolution tree kernel for extracting the relations. (Chun et al., 2011) proposed composite kernel approach for calculating the similarities using predicate argument structure patterns and phrase structure in
convolution tree kernel. The kernel based methods are applied by (Zhou and Zhang, 2007) (Zhang et al., 2008) (Zhou et al., 2010) for extracting semantic relations.

Kernel based method for relation extraction methods explore large feature space in polynomial time without representing the features explicitly and also outperform as compared to feature based relation extraction.

The various relations are extracted from biology based corpus which was carried out by (Huang et al., 2006) (Rinaldi et al., 2007) (Abulaish and Dey, 2007). The semantic relations are extracted between Nominals from various knowledge sources which was described by (Zhang, 2008) (Girju et al., 2010). The semantic relations are also extracted by constructing the rules (Huang, 2009). (Zhou et al., 2008) used hierarchical learning strategy for extracting semantic relations.

The supervised approaches are computationally challenging and difficult to extend with the new relations. It can’t scale well with the increasing amount of input data.

The abundant amount of unlabeled data is available. Hence, the semi-supervised approaches have been implemented. Next section describes various semi-supervised approaches.

2.7.2 SEMISUPERVISED APPROACHES

A semi-supervised approach uses bootstrapping method. In semi-supervised approach, the initial data is labeled and then train the remaining unlabeled data using labeled one. It uses the algorithms which were proposed by (Yarowsky, 1995) and (Blum and Mitchell, 1998). It helps to use weak learners output for next iteration as training data.

The DIPRE (Brin, 1999) and Snowball (Agichtein and Gravano, 2000) systems adopted semi-supervised approach which requires small set of labeled data for training unlabeled data. KnowItAll (Etzioni et al., 2004) and TextRunner (Yates et al., 2007) applied semi-supervised approach for extracting relations from large collection of data.

2.7.2.1 DIPRE

(Brin, 1999) proposes a system called DIPRE (Dual Iterative Pattern Relation Expansion) for structured relation extraction. This system is proposed to extract author and title relation from web i.e from the collection of HTML documents.
DIPRE learn patterns from the tuples defined the documents which consist of order, author, book, prefix, suffix, middle patterns. The extracted tuples are grouped together according to the order and middle tuple. It applied wild card expression for extracting more relations from web. DIPRE uses Yarowsky algorithm (Yarowsky, 1995) for relation extraction.

2.7.2.2 Snowball

Snowball (Agichtein and Gravano, 2000) architecture is similar to DIPRE. It extracts organization and location relations from text documents. In this, each tuple is represented as a vector and compute similarity function to group the tuples instead of using exact match. It extracts the tuples in the form of <prefix, organization, middle, location, suffix>. It identifies all organization and location entities from text documents using named-entity tagger. The system forms the tuple from the extracted organization and location pair. The relation tuple is matched with all extracted patterns and retain only the relation which has the highest similarity value. It is more flexible than DIPRE.

2.7.2.3 KnowItAll

KnowItAll (Etzioni et al., 2004) automatically extracts large collection of concepts and relationships from web. KnowItAll is a language and domain independent and support scalability. The ontology and small number of templates are given to extract the rules for each concept and relations in ontology. It includes the modules like extractor, search engine interface, assessor and database to extract the relations. The communication is accomplished using thread through asynchronous message passing. It extracts the relations using relation specific extraction rules. The set of relations should be specified in advance.

2.7.2.4 TextRunner

TextRunner (Yates et al., 2007) automatically extracts set of concepts and relations from the large collection of text documents without using dependency parser. It automatically labeled the training data as positive or negative during learner module. Then, the labeled data is used by the binary classifier for extraction. The relations are generated by the extractor and the reliable relations are retained. The probability for each retained relations are assigned by using probability model of redundancy. The
redundancy based method is used to extract the relations from web (de Boer et al., 2007). Generally, Semi-supervised approaches are ineffectual and unacceptable because it has limited number of labeled data. So newly added data is liable to lose the original meaning of the data which can be called as semantic drift (Curran et al., 2007).

Our proposed SeRelaC system uses semi supervised approach for relation extraction from unstructured information. The semantic relations are also captured from structured documents like OWL, RDF which could be used for mining relations, question answering, etc. The extracted relations should be classified to increase the efficiency of information retrieval, predictive power and to strengthen the communication among entities.

2.8 RELATIONSHIP CLASSIFICATION

There are various possible diversified semantic relationships between concepts. The classification of semantic relationships is a crucial and important task. Generally, the semantic relationships are classified in a supervised manner. To improve the classification accuracy, the semantic resources such as WordNet, FrameNet, VerbNet, etc. have been widely used. The standard relation classification relies on training data i.e. in the form of tagged information.

(Craven, 1999) has applied Naïve Bayes classification algorithm for the classification of the semantic relationships. The documents are represented as Bag-of-Words and the corresponding relations are extracted from MEDLINE database. The semantic relationships are classified in normalized noun phrases using SVM in (Girju et al., 2004). (Zhang, 2004) used weakly supervised approach which consists of supervised learner and bootstrapping algorithm for classifying the semantic relationships. They have used SVM for classification which was evaluated over ACE corpus. (Rosario and Hearst, 2004) have used MeSH (Medical Subject Headings) for relation classification of MEDLINE database. The semantic relations are labeled and then semantic relationships are classified, based on the lexical semantic relationships by machine learning method (Marsi and Krahmer, 2005).

(Davidov and Rappoport, 2008) proposed a pattern cluster method for Nominal Relationship (NR) classification. In this, pattern clusters are discovered from the large
dataset independent of the training data using unsupervised approach. The discovered clusters are learning a model from the training dataset and it could be used by the classification. The classification is performed by cluster labeling, cluster HITS value and unsupervised clustering. They have evaluated their approach with SemEval-07. It could be used to extract the instances using semantic relationships. (Yan et al., 2008) have used kernel method for relation classification and evaluated with SVM.

(Do and Roth, 2010) proposed TAREC (TAxonomic RElation Classifier) which is used to classify taxonomic relations between a given pair of terms using machine learning based classifier. In this, the relations such as ancestor and sibling relations are considered. TAREC algorithm is evaluated with YAGO ontology and provides enhanced accuracy. (Hendrickx et al., 2009) introduced multi way classification of semantic relationships between the pair of nominals. The relations are classified into cause-effect, instrument-agency, product-producer, content-container, entity-origin, entity-destination, component-whole, member-collection, communication-topic category.

(Szarvas and Gurevych, 2010) computed semantic relatedness between entities and applied for semantic relation classification. They have applied maximum entropy classification algorithm over SemEval2010 dataset for relation classification. The semi-supervised machine learning approach is applied to exploit both labeled and unlabeled data for semantic relation classification (Chen et al., 2010). It adopted ASO (Alternating Structure Optimization) algorithm for relation classification. The relations are classified using SemEval2007 and SemEval2010 dataset.

The classified semantic relationships have been used to improve the communication among concepts and to improve the predictive power. Next section describes the effectiveness of semantic relationships in information retrieval.

### 2.9 INFORMATION RETRIEVAL WITH SEMANTIC RELATIONSHIPS

The keyword based search is the most popular method to retrieve the documents from the document set. In the traditional search, the documents are retrieved which contains maximum number of given keyword(s). The keyword based search is most popular for its simplicity and ease of accessibility but it has failed to retrieve the documents based on its semantics. As the information on the web increases dramatically, it
becomes more and more complex to the user to find more relevant information. In semantic search, the given query is expanded with synonym of the given keyword to improve the accuracy of the information retrieval. Sometimes, it may also lead to retrieval of irrelevant documents. The integration of semantic relationships in semantic search will help to reduce the number of irrelevant documents. Hence the given query should be expanded with semantic relationships for effective information retrieval.

The strength of the relations are measured by link based analysis in (Rocha et al., 2004). They have used measures such as cluster, specificity and hybrid to find the relevance of the relations. They have implemented spread activation algorithm to extract the relations and to assign weight to each semantic relationship. The algorithm is domain dependent. The querying relationships between concepts was presented by (Anyanwu et al., 2005).

(Ning et al., 2009) described an approach to extract the shortest path between concepts. They have proposed approximation algorithm to retrieve answer for the given query. (Chatvichienchai and Tanaka, 2009) proposed an approach to search the set of documents for the given query based on document type, keywords, semantic relationship between given keywords and the documents. They have used to locate the office documents which are based on XML. The query is manually annotated with relations like equals, less than, greater than and searches the documents according to the generated query.

Conceptual semantic space (Huang and Zhang, 2010) has been proposed to expand the query using semantic relations. In this, semantic tree i.e. Tree for Associational Semantics Model (TASM) is constructed and query is expanded using the tree. It calculates the similarity between the given queries and the set of documents and extracts the documents according to it. (Lee et al., 2010) proposed semantic association based search and visualization methods which help the user to understand why and how the results are retrieved. In this, the user query is seeded and then semantic spanning is performed. It is used to find the semantic path and rank the documents with similarity and distance.

Relation based search engine called OntoLook was implemented by (Li et al., 2007). OntoLook process the keyword with the semantic relationships which is offered by
semantic web. Search engines make use of RDF documents for relation based search. The concept-relation graph is constructed using the concepts and the relationships between the concepts. The subgraph is constructed with some quantitative relations between concepts. OntoLook system fetches the relations between a given keyword pair and form property keyword candidate set. The corresponding URL is retrieved through ontology database. It reduces many keyword isolated pages by incorporating the semantic relationships. The user should provide all keywords to extract the relevant pages. In addition, it does not consider the semantics and importance of the relationships. Ranking method is not incorporated in OntoLook search engine.

(Lee et al., 2014) proposed a method to extract the set of semantic paths between target resources and given query keyword(s) to form semantically enhanced query with ontology schema. Many semantic paths are extracted and it can be pruned using length threshold and weight threshold. Pruning semantic paths effectively reduce the search space. In order to access top-k results, keyword index (extended inverted index) has been used through off-line processing.

This section described the way in which the semantic relationship has been exploited substantially, in information retrieval. Our proposed system RISeR incorporates similarity measures and reasoning algorithm for retrieving more relevant documents. Next section describes the methods for ranking with semantic relationships.

2.10 RANKING WITH SEMANTIC RELATIONSHIPS

Ranking is used to rank the documents according to the relevance of the given query. Once the relevant documents are retrieved, it should be ranked based on the ranking function which includes term frequency and tf-idf. Google search engine uses link analysis algorithm such as PageRank (Brin and Page, 1998) (Page et al., 1999), HITS (Kleinberg, 1999), Reverse Page Rank (Fogaras, 2003), Object Rank (Balmin et al., 2004), PopRank (Nie et al., 2005) algorithms to rank the retrieved webpages. The user searches the documents with multiple keywords. The retrieved documents are ranked and displayed based on the maximum number of occurrences of the given keyword(s). Most of the times, it may display irrelevant documents in the first page of a search. It can be reduced by the identification of semantic relationship between the given keyword(s). The ranking with semantic relationships displays the most relevant
documents on the top of a search. Little effort has been made to rank the documents using semantic relationships.

The relationships are extracted through semantic metadata and annotation from RDF graphs and ranking techniques are used to identify the most relevant documents (Aleman-Meza et al., 2003). In this, they have assigned weights for context specification, subsumption, path length and trust. The ranking is based on the context specified by the user and it is evaluated according to the user’s interest. Generally, there exist several paths between the concepts. (Aleman-Meza et al., 2005) proposed a method to rank the documents based on semantic and statistical metrics. They have considered context, subsumption and trust for semantic metrics. Statistical metrics includes rarity, popularity and association length. The documents are ranked using these measures. In this approach, they have proved that it displays the retrieved results according to the user’s interest.

Semantic relationships are based on the notation of RDF property sequence (Anyanwu et al., 2005). The semantic relationships may be ranked with shortest path, longest path, least frequently occurring paths, etc. (Anyanwu et al., 2005) proposed SemRank ranking approach to measure the amount of information conveyed to the user. In general, predictability measure has been used to rank the documents in ontology. The predictability is measured with uniqueness of the retrieved result and discrepant structure of the result. (Wu et al., 2008) proposed CARRank (Concept And Relation Ranking) algorithm to identify important concepts and relations in ontology. They have considered the importance of the relations for displaying the important concepts. Relevance measure has been proposed by (Aleman-Meza et al., 2010) to rank the documents. It can be determined by the relevance of a concept with respect to other concepts in the same document. It can be measured with the semantic path and semantics of the relationships that associate concepts in ontology.

The ontology based ranking methods such as Ontokhoj (Patel et al., 2003), OntoSelect (Buitelaar et al., 2004) and AKTiveRank (Alani et al., 2006) are also used in semantic web. (Lee et al., 2014) proposed a ranking method for semantic search with number of semantic paths, keywords coverage and discriminating the power of keywords. The semantic path weight is assigned with properties of the semantic paths. The relevance between two concepts can be decided by the length of the semantic paths. If the
semantic path length increases, the relevance between two concepts decreases. They have used Threshold algorithm (Fagin et al., 2003) which is based on extended inverted index to answer top-n results efficiently. They have constructed ontologies from DBLP data and IMDB data for evaluating the algorithm.

This section dealt with the significance of semantic relationships in ranking. In our proposed RaSeR system, semantic relation strength, relatedness degree and semantic relation length have been used to rank the documents. Next section describes the role of semantic relationships in the question generation.

### 2.11 QUESTION GENERATION WITH SEMANTIC RELATIONSHIPS

Question generation system helps the facilitator to generate sensible questions from structured or unstructured information. Question Generation from unstructured information is the most challenging one that involves Natural Language Understanding and Natural Language Generation. The text is mapped into symbols with natural language understanding and natural language generation is used to map symbols into text. Question generation system map symbols for declarative sentences into interrogative sentences. The symbols are used to represent the semantics of the natural language which can be processed by the system. Existing research towards question generation systems is widely based on template and syntax. Recently, the researchers focused towards semantics based question generation system. The generated questions should help the user to test various cognitive skills such as knowledge, comprehension, application, analysis, synthesis and evaluation. It will be more useful if the questions are developed on the basis of Bloom’s taxonomy (Bloom and Krathwohl, 1956).

#### 2.11.1 Template based QG System

Self-questioning strategy was proposed by (Mostow and Chen, 2009) to generate the questions from narrative sentences. It helps to enhance understanding at the children level. The templates include *what, why, how*. (Chen, 2009) extended the work to generate the questions such as What-would-happen-when, What-would-happen-if, Why-x and When-would-x-happen questions. Template based question generation system is widely used for specific application domain. In this, templates should be generated manually to generate the questions.
2.11.2 Syntax based QG System

(Wyse and Piwek, 2009) proposed a system Ceist which is used to generate the questions. In this, suitable sentences are extracted using pattern matching and pre-determined questions are generated using rules. They have used OpenLearn data resource for question generation. (Heilman and Smith, 2009) proposed a framework to generate comprehension questions automatically. The declarative sentences are extracted from text and converted using rules and the questions are statistically ranked (Heilman and Smith, 2010). The syntax based approaches employ syntactic tree (Klein and Manning, 2002) method. Multiple choice questions are generated automatically using NLP methodology (Mitkov and Ha, 2003). Question generation system is based on syntactic and semantic information which were extracted using Definite Clause Grammar (DCG) (Kunichika et al., 2003). (Gates, 2008) used phrase structure parser and tree manipulation language for generating the questions. Automatic question generation system is used in educational technologies (Graesser et al., 2005) and dialog systems (Walker et al., 2001).

2.11.3 Semantics based QG System

Semantic based method is performed on the semantic representations of the given sentence and helps to map declarative sentences into interrogative sentences. (Schwartz et al., 2004) introduced content question generator using logical form which can be used to identify the semantic relationships among various segments of a given sentence. The logical form can be used to generate WH questions.

(Papasalouros et al., 2008) proposed an approach to automatically generate multiple choice questions. They have followed the strategies such as class, properties and terminology. The semantic relationships between concepts are considered to generate the questions. (Xu et al., 2009) generated rules and used semantic hierarchy of a sentence to generate questions automatically by changing the clause type and converting into interrogative sentence.

(Teitsma et al., 2011) developed Situation Awareness Question Generator (SAQG) for generating questions from ontology. SAQG is used to determine the situation and to generate yes-no type questions. (Al-Yahya, 2011) developed OntoQue engine for generating multiple choice questions from domain ontologies. Analogy questions are
generated from ontology (Alsubait et al., 2012). They have proposed an approach to extract the pair of concepts using relatedness function and generated multiple choice questions based on the similarity which is computed using analogy function.

(Yao and Zhang, 2010) interpreted natural language sentence as Minimal Recursion Semantics (MRS) structure by parsing and through generation, MRS structure can be interpreted as natural language sentence. The authors identified the issues such as sentence simplification, question transformation and ranking. The solutions are proposed to resolve these issues. It includes connected dependency graph for sentence simplification, transformation of declarative sentences into interrogative sentences for question transformation and MaxEnt (Maximum Entropy) model for ranking the questions. (Yao et al., 2012) extended their work and developed MrsQG for transformation of sentences, sentence decomposition and automatic question generation with ranking. An ontology based approach is used for generating analogy questions (Alsubait et al., 2012). In QG system, semantic relationships play a vital role to generate WH questions.

2.12 SUMMARY

In this chapter, the reasoning techniques are classified and brief description about reasoning techniques are provided. The basic notations of Description Logic are provided which will be used for reasoning. The descriptions of reasoning algorithms are provided. The various reasoners are discussed and comparison of reasoners is made. The importance of semantic relationships and its classification are provided. It also discusses about the ways in which the reasoning are performed. The various approaches used for extracting the relationships are discoursed. It briefly explained about relationship classification. A brief literature survey is provided on information retrieval and ranking with semantic relationships. Finally, it exploited the usage of semantic relationships in the question generation system. Next chapter elaborates on the problem statement and research methodology of this work.