A Study on Representation and Reasoning Techniques of Commonsense Episodic Knowledge: Challenges and Applications

MOHAMED GAWISH, ABDEL-BADEEH SALEM
Computer Science Department
Ain Shams University
Cairo, EGYPT
mygawish@cis.asu.edu.eg, absalem@cis.asu.edu.eg

Abstract: - In the recent decades, there has been much research on the representation of commonsense knowledge and on inference techniques to manipulate that knowledge. This paper discusses the nature of commonsense knowledge highlighting the main challenges exist in acquiring, representing and reasoning with commonsense episodic knowledge from the view of artificial intelligence. In addition, the paper analyses the different approaches and techniques dealing with commonsense episodic knowledge in many fields, such as problem-solving, human-computer interaction, temporal reasoning, script learning and story generation.

Key-Words: - episodic knowledge; commonsense reasoning; intelligent systems; knowledge representation; cognitive architectures; social situations; context

1 Introduction

Developing intelligent systems for any domain is the main focus of artificial intelligence (AI), which strongly requires gathering knowledge about that domain, selecting proper knowledge representation and reasoning techniques. Knowledge consists of facts, concepts, events, theories, procedures, and relationships. AI techniques are used to give the machine reasoning capability to think, learn and make inferences based on the gained knowledge for handling many intelligent tasks, such as problem-solving, story generation, analogy, negotiation ability, and decision making.

Interesting work is being done in giving computers the ability to reason about everyday situations they face as a human being does it, which require a high level of intelligence in humans [1, 2]. This ability is known as commonsense which is defined as the essential goal of intelligent behavior and thought [3]. In fact that many theories and researches have ended up that building computers that exhibit commonsense ability is more complicated than computers that can solve hard reasoning issues [4]. There is a continuous tendency in AI to provide computers with commonsense knowledge using reasoning mechanisms that can observe the world and acquire implicit knowledge about people, everyday life and the world [4, 5]. Such computers can exhibit commonsense behavior and interact effectively with humans in everyday life [4, 5]. There are many domains related to commonsense knowledge such as time, space, mental states, stereotypical situations, emotions, goals, kinds, and locations of objects and so forth [3, 6].

Chen [7] and Mueller [3] have argued that the world that we live in consists of a variety of objects, events (actions) and changes as a result of applying actions. In addition, most instances of commonsense knowledge involve action and change [3]. From that, episodic knowledge is one type of commonsense knowledge. In which, episodic knowledge is composed of previously experienced events, temporal relations among them and contextual knowledge [8]. For example, temporally dated events of “Entering a stadium” and “Eating at a restaurant.” Temporal relations play a crucial role in preserving the contextual knowledge associated with a particular sequence of events that describes a certain social situation [9]. Through experiences, people interact with different social situations and acquire cultural stereotypes regardless idiosyncratic variations of humanity [7].

The paper is organized as follows: Section 2 highlights the main challenges exist in dealing with commonsense knowledge. While Section 3 presents the requirements needed to explicitly represent that knowledge. A number of systems with different approaches and techniques in representing and reasoning with commonsense episodic knowledge are surveyed in Section 4. In addition, Section 5 surveys a number of cognitive architectures with different techniques in dealing with commonsense episodic knowledge. Finally, Section 6 elucidates our conclusions for this work.

2 Challenges

In contrast to expert knowledge given to computers by human experts, which is usually explicit [4]. The implicit nature of commonsense knowledge arises the need to explicitly represent it [3]. As has been pointed out in many researches such as [5,6, 10-12] that AI pioneers such as John McCarthy, Doug Lenat, and Marvin Minsky stated that the fundamental challenges in dealing with commonsense knowledge in AI are:

A. Representing Commonsense Knowledge.

First, it is required to find proper general ways to encode in machines the whole types of commonsense knowledge that humans have. What is the best technique to efficiently represent such implicit knowledge? Should it be encoded as a long list of if-then production rules? Should multiple representation techniques be used?

B. Capturing and Storing Sufficiently Commonsense Knowledge by Computers as Automatically as Humans.

Second, it is required to find sufficient knowledge acquisition techniques that can acquire this much knowledge. While humans learn numerous things by living on the planet, machines do not generally live on the planet as in they cannot see or control things on the planet almost well as humans can [6]. Despite the works done in the machine learning area, they were not able to promptly acquire information about the world as humans can. Since the beginning of AI area, the hurdle of implementing a believable computer program that can act intelligently in many different domains is to acquire a large amount of information of the world, this problem is so-called knowledge acquisition bottleneck [6].

C. Using and Reasoning with Commonsense Knowledge.

Third, it is required to find reasoning techniques that can simply use commonsense knowledge to learn or correctly act in new situations. It is worth mentioning that there are some automated reasoning techniques in specific narrow domains that can be compared to humans' behavior such as game playing domain [6]. But with regards to general reasoning technique with commonsense knowledge, it seems very hard to implement program that can, for example, understand a simple story as children can.

From the general view, the main challenges of commonsense knowledge are previously discussed. To more specific view, the main challenge of the episodic knowledge as a type of commonsense knowledge is finding a way to represent and reason with such knowledge while preserving the context. Context is the temporal execution of events that distinguish a certain situation from another situation and clarify ambiguous situations [9]. So, context organizes the episodic knowledge and cannot be segregated from it.

3 Representation Requirements

Regard to the first challenge of finding general representation techniques to represent the tremendous amount of commonsense knowledge. McCarthy and Hayes [12] define four types of adequacy which after that Davis [10] considered them as the main requirements that the knowledge representation techniques must satisfy to represent commonsense knowledge, which are:

A. Metaphysical (Representational) Adequacy.

This requirement means that the representation technique can represent all aspects of the commonsense knowledge without contradicting any fact of the reality [12].

B. Epistemic Adequacy (Expressivity).

A requirement for the representation technique that can express the available knowledge regardless it is complete or not (i.e., representing the partial knowledge). The time consuming and difficulty of getting and computing with complete information arise the need and importance of representing and reasoning from partial knowledge [10].

Expressivity is what distinguishes between representations used to represent commonsense knowledge and representations used to represent other types of knowledge [10]. In which, commonsense reasoning has a need to use different kinds of partial knowledge as it drives general rules that apply to a wide range of different situations. While other types of knowledge assumed to be complete or partial in extremely restricted domains [10].

C. Heuristic Adequacy (Effectiveness).

The necessity that the representation technique can be computationally used by a reasoning technique to implement effective inferences [10].

D. Acquisitional Efficiency.

A requirement for the representation technique to acquire new information easily and control knowledge acquisition [13].

4 Commonsense-Based Applications

In the following, we describe some systems that take unconventional different approaches and techniques to representing, acquiring, and reasoning with large volumes of commonsense knowledge with focusing on the episodic knowledge. Each system endorses a different technique: CYC [14] is a large-scale formal logic, OMCS [15] is a knowledge base of sentences in natural language, ConceptNet [16] is a large-scale semantic network, LifeNet [2] is a probabilistic graphical model, EventNet [5] is a large-scale association network, ThoughtTreasure [29], OMEX [17] and StoryNet [6] are databases of linear scripts, and FrameNet [18] is a database of semantic frames.

A. CYC.

The first attempt at creating an immense knowledge base of diverse types of commonsense knowledge began in 1984. As of 1984, Lenat et al. [14, 19] have started their CYC project in building a large commonsense knowledge base. They believe that building an immense knowledge base of everyday facts about the world with a reasoning by analogy capability, may have the credit to overcome the knowledge acquisition bottleneck problem [14]. In which, understanding new information depends on what already know in advance. So, a team of knowledge engineers is manually handcrafting the commonsense knowledge as factual assertions in formal language [16].

Along a decade, most of the time has been spent in collecting approximately million commonsense facts and encoding them in first order logic [19]. While CYC is the largest commonsense knowledge base till now, it contains general episodic knowledge in the formal logic (facts and rules) rather than descriptions of events flow in different situations [17].

In Massachusetts Institute of Technology (MIT) Media Lab, the researchers believe that as the diversity of commonsense knowledge, it needs many types of knowledge representation that can represent every aspect of such knowledge [20]. The problem is how to build one system that can find a way to manipulate them [20]. So, they are exerting efforts on creating machines that can exhibit human-like commonsense intelligence in everyday life. Machines that can learn and reason about different aspects of human daily life. Machines that have knowledge of the kinds of objects and goals people have, the actions people can take and their effects, and so forth. More than eight commonsense tools with different representation and reasoning techniques are developed by MIT Media lab., such as OMCS-1, OMCS-2, OMCSNet, ConceptNet, LifeNet, EventNet, OMEX, and StoryNet.

B. OMCS-1.

As the end of 2000, the first version of Open Mind Common Sense project (OMCS-1) [20] (at http://www.openmind.org/commonsense) was started to acquire knowledge about the ordinary life over the web. By involving non-expert users, with no first preparing, in many activities that gather facts, rules, stories and descriptions in free-form simple English sentences [15, 20]. In contrast to CYC, OMCS knowledge is represented by natural language rather than formal logic. In addition, it constructed by nonexperts rather than knowledge engineers. A spelling check method, a word-sense disambiguation method and information extraction techniques are utilized to extract meaningful knowledge and to structure the extracted knowledge in knowledge representation technique to be usable for further commonsense-based applications [15]. This resulted in participating 9296 users and gathering 465,195 assertions of commonsense knowledge [15]. Which have been employed in many different applications such as ARIA [21] which is a photo retrieval system uses commonsense knowledge for managing personal photos of users and GOOSE [22] which is a goal-based search engine uses commonsense for returning the most effective results that satisfy the users' search goal [15, 20].

C. OMCS-2.

Due to the enthusiasm of participants in organizing, clarifying, assessing, and entering different sorts of knowledge [15]. The second version (OMCS-2) is motivated. In OMCS-2 [15], the knowledge acquired using templates in English rather than using free-form and still represented by natural language [15]. In which, participants are involved in 30 different activities and are restricted to enter knowledge into blank fields [15, 23]. For example, the effect of [drinking water] is [a feeling of satiety] and somewhere you find [a car] is [in front of the house] [15]. These templates extracted from OMCS-1 and helped in acquiring temporal and causal relations between stereotyped events [20]. OMCS-2 utilized analogical reasoning in generating a list of inference rules that used to feedback on entered knowledge [15]. At the beginning of 2005, the results were 750,000 assertions from over 16,000 participants around the world [5].

As a commonsense knowledge acquisition system, the quality of this corpus is evaluated by human judges. They graded 75% of data as largely true, 82% as mostly objective and 85% as making sense [23]. OMCS-2 is the second largest knowledge base of commonsense knowledge after CYC.

D. ConceptNet.

Based on commonsense knowledge of the OMCS corpus, the first attempt to create a free reasoning toolkit over text with automatically generated commonsense knowledge base is called OMCSNet [23] which was launched in 2002. As of 2004, a new version of OMCSNet with natural language processing was released under another name which is ConceptNet [6, 16].

ConceptNet is generated by extracting a set of binary relations which form a simple, handy semantic network [16, 23]. Nodes represent fragments of English sentence and edges represent the commonsense relation between these fragments [16, 23] as shown in Fig. 1 for example. ConceptNet consists of twenty binary relations such as ISA, PartOf, HasProperty, MadeOf, SubEventOf, HasEffect, HasAbility, LocationOf, DoesWant, HasFunction, and ConceptuallyRelatedTo [16, 23]. Once the system was launched, these relations have composed around 1.6 million edges and connected around 300,000 nodes [16, 23].

From the expressive power of English in representing knowledge in each node and the power of semantic network as a graph in inference tasks, ConceptNet has three textual reasoning tasks [16, 23]. The first task is context determination, finding out the context around one or several concepts by carrying out spread activation from the query concept to other concepts [16, 23]. The second task is a semantic similarity and analogy-making, which utilizes analogical inference by performing structure-mapping methods of Gentner [24] over a semantic network, for computing the degree of structural similarity between pairs of query concepts or returning a list of structurally similar concepts to the query concept [16, 23]. The third task is building inference chains by going over the network starting with one node then onto another node till finding paths between two query concepts [16, 23]. As a commonsense toolkit for textual reasoning, the quality of this toolkit was evaluated by human judges. They in average rated 68% of data as fairly comprehensive, and 24.8% as noisiness (i.e. include incorrect knowledge) [16, 23].

ConceptNet has been employed in many different interesting applications such as MAKEBELIEVE [46] which encourages users to interact with it in order to invent new stories, GloBuddy2 [25] is a mobile application using

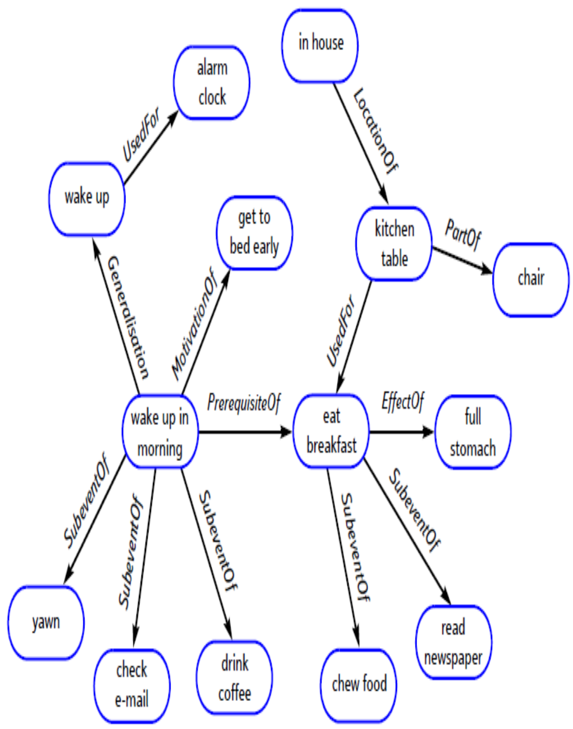

Fig. 1. A fragment of ConceptNet's semantic network. (Adapted from [16]).

ConceptNet for providing English speaking tourists with a dynamic language phrasebook, and What Would They Think? [26] which utilizes ConceptNet and natural language processing to model person's attitudes by automatically analyzing personal texts (e.g. weblogs, and e-mails). ConceptNet, however, can only gain and learn new knowledge from OMCS corpus and no learning method is developed to enrich its knowledge base by reasoning from its prior experiences formulated in the semantic network.

E. LifeNet.

Like ConceptNet, another commonsense knowledge base is generated based on the OMCS corpus so-called LifeNet [2]. LifeNet is concerned with temporal reasoning using probabilistic belief updating techniques [27]. LifeNet is a knowledge base of propositions extracted from the OMCS corpus and the OMCSNet [23]. LifeNet represents these propositions as a probabilistic graphical model with temporal and atemporal relations (before and after relations) connecting them together with joint transition probabilities [2] as shown in Fig. 2 for example.

The generation of LifeNet knowledge base passed through four steps. The first step is generating propositions, that is, a list of propositions is created by collecting some of the English sentences from the OMCS corpus and arguments of the OMCSNet

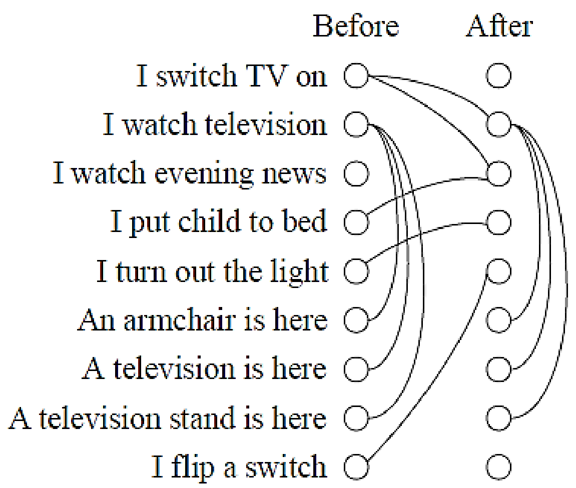

Fig. 2. A sample of LifeNet. (Adapted from [2])

knowledge base relations [2]. It is worth mentioning that in this step a set of automated and manual filters are run on the collected list for repairing or deleting many of types of generation errors. The second step utilizes eight binary relations of OMCSNet in order to generate propositions rules [2].

Temporal and atemporal proposition rules are created based on EffectOf, SubEventOf, LocatedAt, OftenNear, FirstSubevetOf, LastSubeventOf, UsedFor, and Requires OMCSNet relations [2]. Due to errors in this step, a probability value that determines to what extent the rule is accurate is assigned to each resulting proposition rule in the third step [2]. These three stages generated around 130,000 proposition rules. Which in turn converted into a large probabilistic graphical model of 78,332 nodes and 415,248 edges in the last step. For more details of the generation process, see [2].

With comparison with ConceptNet, ConceptNet is more expressive than LifeNet [6]. This is because LifeNet focuses only on temporal relations between different situations [6]. However, LifeNet as a commonsense inference system for temporal reasoning, its quality was also evaluated by human judges. Due to resolving the knowledge base errors and the simplicity of knowledge representation, the judgment of humans rated 89% of propositions and 78% of inferences make sense [2]. These results were in favor of LifeNet and proving that knowledge of LifeNet is more accurate (less noisy) than ConceptNet which its knowledge was in average rated 24.8% as noisiness.

For extending LifeNet knowledge base, a web site called Open Mind LifeNet is created to collect more propositions and repairing the existing ones from the general public as done in the original OMCS corpus [2]. Three more activities are added to acquire knowledge about new propositions, temporal and atemporal edges. Although, the work of Open Mind LifeNet on extending the knowledge base of LifeNet, but as mentioned that this work depends only on the volunteers and there is no an automated learning algorithm that can drive new knowledge from existing experience.

F. EventNet.

Focusing on commonsense temporal reasoning, another toolkit for predicting temporal relations between ordinarily occurring events was released on 2005. This toolkit is so-named EventNet [5]. EventNet is more concerned with predicting the past and future events, rather than building a huge knowledge base of all possible temporal nodes and relations [5]. EventNet depends on LifeNet knowledge base in extracting the temporal information and creating an association network for representing its knowledge [5]. Using the temporal information of LifeNet an association network, of 10,000 nodes and 30,000 temporal links, is created by temporal nodes of the English form and weighted links between nodes [5].

EventNet utilizes spread activation algorithm to do two main inference operations. In contrast to LifeNet of predicting from only one event at a time [5]. The first operation of EventNet is plan recognition which can infer from the user's current event or a set of events, a set of possible associated subsequent (next) events, antecedent (previous) events, and temporally related events. The second operation is finding paths between two given events as a sequence of temporally related events between the two [5]. Additionally, EventNet can create new links between two semantically similar events by utilizing synonyms from WordNet [32] and ConceptNet analogies to efficiently search the knowledge base for plausible paths between events [5]. The quality of EventNet inference toolkit was evaluated by human judges with the comparison with human inference. They in average rated 62% of EventNet temporal reasoning make sense, while human inference was rated 69% [5].

Although the rich work is done in CYC, OMCS, ConceptNet, LifeNet, and EventNet for capturing and representing commonsense knowledge, there is a shortcoming when dealing with episodic knowledge in different social situations. This flaw is due to (a) having general knowledge about situations rather than rich descriptions of events flow, (b) neglect the context of the current social situation, (c) fairly sparse of knowledge in separate knowledge bases [29], (d) lack of relations that connect events in a believable manner, and (e) difficulty in reasoning with used knowledge representations.

For that, a new perspective of using scripts, inspired by Schank and Abelson [28], has been considered in assembling and representing commonsense knowledge for building commonsense knowledge bases such as ThoughtTreasure [29], OMEX [17], and StoryNet [6]. In ThoughtTreasure, knowledge engineers entered about 100 scripts of different activities. While OMEX and StoryNet, prompting the general public to enter structured stories which then represented by scripts, instead of factual English assertions as CYC, and OMCS.

G. ThoughtTreasure.

In 1998, Mueller declared its work in building a commonsense knowledge base, called ThoughtTreasure and begun in 1994, for reasoning purposes, natural language processing, and computational linguistics tasks [29]. ThoughtTreasure contains about 35,000 English words and phrases and 51,000 assertions [29]. For encoding the diverse sorts of commonsense knowledge, it utilizes four knowledge representation techniques: physical spaces are represented by 2-dimensional grids, rules of thumb are represented by finite automata, linguistic facts by logical assertions, and activities (situations) by scripts [29].

Scripts are organized into a top-level hierarchy [29]. In which more specific scripts such as riding a car inherit the information of more general scripts such as riding a vehicle. About 100 scripts of different situations have been entered by knowledge engineers within the knowledge base [29]. Each script contains a chronological sequence of events that constitute the script rather than scenes, humans and physical objects that are taking part of the script, entry conditions, results of carrying out the script, personal goals, emotions, locations, and duration of the script [29]. Each information of the script is of the form *[predicate-name argument1 argument2 ...]* [29. For example, blackout script is shown in Fig. 3[29], in which it inherits disaster script as stated in a kind of (ako) predicate, its duration is 3,600 seconds. Three emotions associated with it (anger, unhappy-surprise, and worry). Two events make up the script, the first event (event01) indicates that the power cuts and three emotions associated with this event, while the second one (event02) is the subevent that indicates that the reaction of power cuts is fetching another light source. The script has one player (human) and one physical object (electricity network) and occurs in an apartment, house or office.

```
Object blackout

[English] power failure, blackout; [French] black out,
panne de courant, panne d électricité
[ako ^ disaster]
[duration-of ^ NUMBER:second:3600]
[emotion-of ^ [anger human]]
[emotion-of ^ [unhappy-surprise human]]
[emotion-of ^ [worry human]]
[event01-of ^ [anger human]]
[event01-of ^ [electronic-device-broken
               electricity-network]]
[event01-of ^ [unhappy-surprise human]]
[event01-of ^ [worry human]]
[event02-of ^ [fetch-from human na light-source]]
[performed-in ^ apartment]
[performed-in ^ house]
[performed-in ^ office]
[period-of ^ NUMBER:second:3.1536e+07]
[role01-of ^ human]
[role02-of ^ electricity-network]
```

Fig. 3. A ThoughtTreasure script. (Adapted from [29])

Organizing scripts in a hierarchical manner and representing events in logical assertions make it hard to find automated methods for inferring new knowledge and the context, learning and adding new events without ignoring or breaking the logical sequence of the events and applying the script in similar activities.

H. OMEX.

Unlike using the graph with binary relations or factual assertions or logical rules to represent commonsense knowledge as done in ConceptNet, LifeNet, OMCS and CYC. On the approach of OMCS, OMEX (Open Mind Experiences) [17] is another crowdsourcing approach by MIT Media Lab researchers to capture the commonsense knowledge from the general people through the Internet. OMEX is the first knowledge acquisition tool intended to collect story-like knowledge from non-expert volunteers [17]. Its idea came from the argument that humans express and exchange their knowledge in the form of stories [17].

Wherefore, its goal is to build a large-scale commonsense knowledge of stories by allowing volunteers to write down stories and explanation of these stories in structured English sentences [17]. Thirty-two story templates were built by hand and prompt users to fill in the blank places with English sentences to complete the story [17]. The user can add a new story, explain a story by answering some questions about it, evaluate a story, report that a story has grammar or syntax errors and repair errors of a story [17].

I. StoryNet.

Under building a large-scale story knowledge base, StoryNet [6] is the second attempt to capture commonsense knowledge as structured stories from the general public through the Internet by MIT Media Lab. Additionally, for extending its knowledge, StoryNet utilizes the ConceptNet and LifeNet as commonsense resources for gathering more stories about different social situations such as riding a bicycle, entering a restaurant, and visiting a doctor [6].

StoryNet is created based on Aristotle's theory, in which Aristotle [47] argued that humans use stories to remember and organize their past in the form of a sequence of episodes [48]. StoryNet represents its knowledge using linear scripts [6]. Each script is a list of temporally linked events that model the flow of events in a particular situation [6] as shown in Fig. 4 for example. Unlike ThoughtTreasure, one situation may have multiple (one or more) scripts that represent different scenarios of events in that situation [6].

Singh et al. [6] have declared their desire to integrate case-based reasoning with StoryNet for generating new stories, merging two scripts together, and other script manipulations. But till now no notable progress is published. In addition, no evaluation of collected story knowledge is published.

Even though, the beneficial effect of OMEX and StoryNet in collecting a large-scale commonsense knowledge in the form of stories. There is no an automated process for inferring the context of the current entered story based on the prior experiences instead of prompt volunteers to answer some questions that can determine the context [6]. In addition, using linear scripts in representing situations may increase the complexity of finding the most similar script to the current situation or may make it difficult to find the plausible reaction to the current faced event to give computers a human-like behavior in social situation.

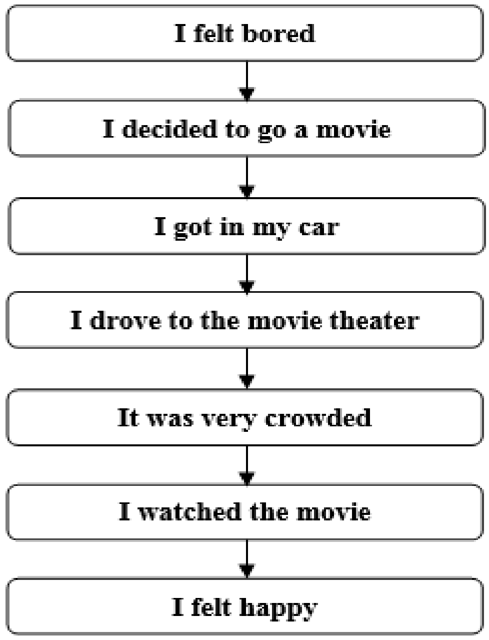

Fig. 4. A StoryNet script. (Adapted from [6])

J. FrameNet.

In 1997, a three-year project called FrameNet [18] was begun for capturing human commonsense into schematic conceptual scenarios in different domains including perception, social context, cognition, communication, space, time and emotion [18, 30]. FrameNet introduced a new approach to collecting and representing the commonsense knowledge. Based on large electronic English corpora such as British National Corpus (BNC) and Concise Oxford Dictionary (COD), FrameNet extracts information about several English lexical items (English words) using both manual and automatic procedures [30]. The extracted information is then represented by semantic frames structure by humans (lexicographers and linguists) [18]. Each frame describes one word and consists of two attributes, frame-elements attribute which contains names of main elements that can participate in describing such word and example sentences in scenes attribute that describes the situation of the frame [18].

As shown in Fig. 5, for example, that TRANSPORTATION frame specifies MOVER, MEANS and PATH as the three elements that take part of any transportation, with a description of the frame emphasizes the relation between these elements by scene: MOVER move along PATH by MEANS. The inheritance relationship is defined between frames, where less specific frame inherits all the properties of a more general and complex one. For example, DRIVING and RIDING frames are specific kinds of TRANSPORTATION frame as shown in Fig. 5.

FrameNet focuses on the scenes that describe the lexical item with lack information about subevents of each scene; this may have occurred due to its much concern in building a computational lexicography system more than capturing all various sorts of commonsense knowledge.

K. Chen's work.

Another work in representing commonsense episodic knowledge using frames is introduced by Chen [7]. Where the actors, props, and events

```
frame(TRANSPORTATION)
frame_elements(MOVER(S), MEANS, PATH)
scene(MOVER(S) move along PATH by MEANS)
---
frame(DRIVING)
inherit(TRANSPORTATION)
frame_elements(DRIVER (=MOVER), VEHICLE
(=MEANS), RIDER(S) (=MOVER(S)), CARGO
(=MOVER(S)))
scenes(DRIVER starts VEHICLE, DRIVER con-
trols VEHICLE, DRIVER stops VEHICLE)
---
frame(RIDING_1)
inherit(TRANSPORTATION)
frame_elements(RIDER(S) (=MOVER(S)), VE-
HICLE (=MEANS))
scenes(RIDER enters VEHICLE,
VEHICLE carries RIDER along PATH,
RIDER leaves VEHICLE )
```

Fig. 5. Some frames from FrameNet. (Adapted from [18])

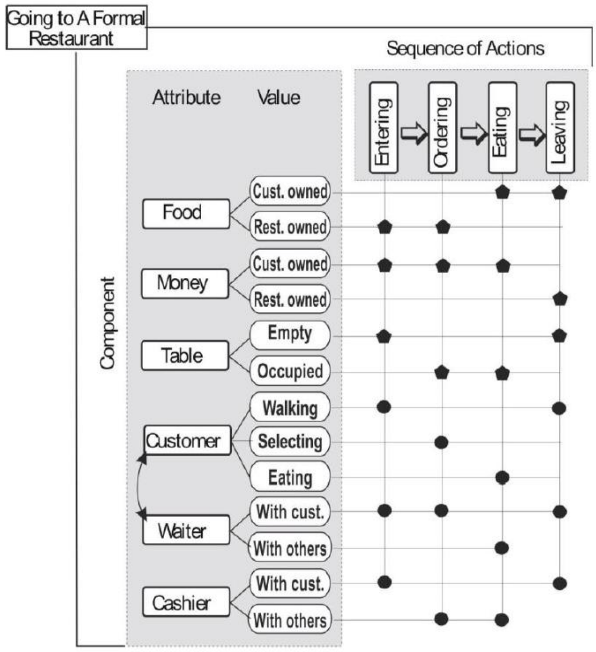

Fig. 6. A fragment of going to a restaurant frame. (Adapted from [7])

formulating a certain social situation are represented by two interconnected frames as shown in Fig. 6.

The first frame is so-called component frame which contains actors and props as multi-valued attributes with their possible values [7]. In addition, it indicates constraint relations between different sets of values [7]. For example, as shown in Fig. 6, the constraint relations between customer's values and waiter's values, in which if the value of customer is eating, then the value of waiter is with others. The second frame is so-called sequence frame and contains the temporal sequence of actions (scenes) as attributes and the causal connections between these scenes. The causal connections for a certain scene defined by a set of specific values corresponding to the attributes in the component frame [7]. For example, as shown in Fig. 6, "eating" in the action sequence takes four specific values from the four attributes in the component frame: "*customer eating*", "*waiter interacting with others*", "*food owned by the customer*" and "*money owned by the customer*".

Despite the significant effort made in this work for representing social situations using frames, while preserving the context and the causal relations. But the frame structure is very complicated to be extended and to be used for inferring the context associated with the social situation. In addition, each frame contains a linear sequence of events representing a certain situation which leads to a lack of diversity of social dynamics associated with that situation.

L. Li et al. work.

In 2012, Li et al. [31] presented a very interesting work in automatically learning a multi-branched cognitive script-like episodic knowledge from crowdsourcing as shown in Fig. 7. For the purpose of story generation, this work aimed to build an extensive commonsense knowledge base of multi-branched stories from linear crowdsourced narratives using a three-step process [31]. The first step in this work is collecting linear crowdsourced story examples of a particular situation by allowing participants to submit a short real-world linear sequence of stereotyped events (linear script) of the given common situation in structured English sentences [31].

The second step is to cluster the core set of events from the entered linear scripts based on semantic similarity. In which, the components (verb, actor, and non-actor noun) of each event are extracted using Stanford parser [33], then the semantic similarity of the events is computed using semantic gloss information from WordNet [32], finally k-Medoids algorithm is utilized to result in clusters, each cluster represents an event that can occur in the given situation [31].

The third step is to construct the multi-branched script by learning the typical temporal ordering of these clusters (events). In which, a technique tries to learn before relations between all pairs of events [31]. For every pair of events (e1, e2), there are two hypotheses of before relation between them either *before(e1, e2)* is true or *before(e2, e1)* is true.

Fig. 7. A multi-branched script of restaurant situation. (Adapted from [31])

Based on counting the observation of evidence for each hypothesis, the probability of each hypothesis is computed using *k/n*, where *k* is the number of observations that supports the hypothesis and *n* is the total number of observations. Finally, the relation that its probability exceeds a threshold (0.5) is selected.

At the end of this process, a multi-branched script, representing different social dynamics and preserving the context associated with a certain situation, is constructed as shown in Fig. 7. While this work is considered as a notable work in script learning. But its learning process is based on intra-domain, where the multi-branched script is constructed from a set of linear scripts in the same domain.

Table 1 represents a summary of all previously discussed systems that take unique, different approaches and techniques to acquiring, representing and reasoning with large volumes of commonsense knowledge with focusing on the episodic knowledge. Although the great efforts are done by the above techniques. But none of them prove its competence to gain all the vast, diverse information that people have.

5 Cognitive Architectures-Based Applications

Due to diversity dilemma of commonsense knowledge, a research direction of building computational models of human cognition, so-called cognitive architectures, is manifested since the early 1970s. A cognitive architecture is a model of a set of human cognitive capabilities such as memory, learning, interaction, perception, problem-solving and emotion [34]. A cognitive architecture is also considered as an information-processing system for computationally modeling many aspects of human performance. It is desired that when the model of cognitive capabilities combined with knowledge, it can play a commonsensical role in exhibiting effective human-like behavior in complex environments [34].

It has been argued that the main cognitive processes in any cognitive architecture are the both memory and learning processes. This is because the crucial role of memory in storing all the different sorts of knowledge that acquired or modified through the learning process. Episodic memory is exhibited by few cognitive architectures, which embody how episodic knowledge is encoded, learned, retrieved and used to improve behavior. In the following, we survey some cognitive architectures that take almost the same learning mechanisms to alter elements of human memory with focusing on the episodic memory.

Soar [35] is a general cognitive architecture that has a general mechanism for learning from experience which applies to any task it performs and guides its behavior. In Soar, all long- term knowledge being represented as production rules, so learning involves creating new facts and productions. ACT-R [36], EPIC [37] and EPIC-Soar [38] are other architectures that also use episodic memory within which knowledge is represented as production rules.

A. SOAR.

As of the beginning of 1980 and an extension to 2008, there has been a series of developments and evolution in creating a general cognitive system, so-called Soar [35], which ended up with the ninth version (Soar 9.0 [39]). Soar stands for State, Operator and Result system, in which all tasks are defined as trails to achieve goals through carrying out a sequence of actions (operators) starting from the initial state (problem) and leading towards the goal state [35].

Soar consists of a single long-term memory and a single short-term memory (working memory) [39].

Table 1. Comparison between different commonsense-based applications.

System	Approach	Knowledge Acquisition	Knowledge Representation	Reasoning Technique	Goal(s)
CYC ***(Lenat et al., 1985)***	Large-scale	knowledge engineers	First-order logic	-	Knowledge acquisition
OMCS ***(Singh et al., 2002)***	Crowdsourcing	Nonexperts (the general public)	Natural Language Sentences (English)	-	Knowledge acquisition
OMCSNet ***(Liu et al., 2002)*** **ConceptNet** ***(Liu et al., 2004)***	Crowdsourcing	OMCS corpus	Semantic Network	Analogy using structure-mapping method	Knowledge reformulation and Textual reasoning
LifeNet ***(Singh et al.,2003)***	Crowdsourcing	OMCS corpus and OMCSNet	Probabilistic graphical model (Bayesian Network)	Belief Propagation	Knowledge reformulation and Temporal reasoning
EventNet ***(Espinosa et al.,2005)***	Crowdsourcing	LifeNet	Association network	Spread Activation	Temporal reasoning
ThoughtTreasure ***(Mueller, 1998)***	Large-scale	Knowledge engineers	Linear scripts as logical assertions	-	Knowledge acquisition
OMEX ***(Singh et al.,2003)***	Crowdsourcing	Nonexperts (the general public)	Linear scripts	-	Knowledge acquisition
StoryNet ***(Singh et al.,2004)***	Crowdsourcing	LifeNet and ConceptNet	Linear scripts	-	Knowledge acquisition
FrameNet ***(Baker et al.,1998)***	Large-scale	Electronic English corpora	Semantic frames	-	Knowledge acquisition
Chen, 2004	-	-	Frames	-	Knowledge representation
Li et al., 2012	Crowdsourcing	Nonexperts (the general public)	Multi-branched cognitive scripts	-	Script Learning

"-" means that a system does not have this property.

All knowledge in long-term memory is represented as production rules, so learning consists of matching and firing rules for creating new facts and rules [39]. While short-term knowledge is represented as a symbolic graph structure of the current state perceived from perception, and the knowledge retrieved from the long-term memory [39]. Soar started with a procedural long-term memory, while episodic and semantic memories were added in 2007.

In Soar, episodic memory provides the ability to remember the history of previous states (episodes) and the temporal relations between these experienced states [39, 40]. Soar compares the current state in working memory to all stored states in episodic memory, and then selects and retrieves the best matching state based on a partial matching algorithm [41]. Depending on the retrieved state, a sequence of states in temporal order is also retrieved [41]. Soar utilizes multiple learning mechanisms for different types of knowledge such as chunking mechanism that compiles the sub-goals of the problem-solving process into production rules. In addition, episodic learning for acquiring episodic knowledge, in which, the contents of working memory are recorded as snapshots in episodic memory through episodic learning process [39, 41].

B. ACT-R.

In 1993, Anderson compiled all his theories, since 1976, of understanding human cognition in one platform so-named ACT-R (Adaptive Control of Thought-Rational) [42]. ACT-R is a cognitive architecture for emulating human cognition, which can be used to implement various models of cognitive tasks such as perception, memory, thinking, problem-solving, and reasoning [42]. It also achieved great progress in implementing intelligent tutoring systems [43]. There has been a series of developments of ACT-R to improve its behavior, which ended up, in 2008, with the fifth version (ACT-R 5.0 [36]).

ACT-R consists of a set of modules that are integrated to produce a coherent cognition. Each module is associated with a buffer for processing a different kind of information. The core modules of ACT-R are a declarative memory and procedural memory [36].

In contrast to Soar, ACT-R makes a distinction between declarative and procedural knowledge and encode them using different representation techniques [42]. In which, it encodes declarative knowledge as chunks, whereas procedural knowledge encoded as production rules. Each declarative chunk has a numeric parameter that reflects its past usage and measures the activation strength of the chunk to be retrieved from memory to solve the current problem [40, 42]. Chunks that are repeatedly accessed receive a high activation [44]. The chunk with the highest activation is chosen, if several chunks are applicable to the current problem at the same time [42].

Similarly, each procedural rule has a probability of success (utility value) that measures the interest of the rule of achieving the desired goal [36, 44]. In ACT-R, declarative memory is implemented as a short-term memory [35] for storing factual knowledge about the past experiences or objects of the current observed environment [36, 43]. While procedural memory is a long-term memory used for storing rules about how to solve problems and plays an important role as a production system for handling different cognitive tasks by coordinating the processing of all other modules [42].

During solving a problem, ACT-R test the conditions of each production rule against chunks in the short-term memory [40, 42]. Once the production rules that exactly match against the chunks are determined, ACT-R computes the utility value of each matched rule [40, 42]. Then it selects the production rule with the highest utility value and executes its actions [40, 42]. Achieved goals are stored as new facts in declarative memory; examples can be generalized to new rules. Through ACT-R experience, learning involves creating new facts of achieved goals in declarative memory and new production rules of solved problems in procedural memory, as well as updating activation and utility values associated with chunks and production rules, respectively [40, 42]

While the modules of ACT-R can work in a parallel manner at the same time [36]. ACT-R has two bottlenecks, compared with Soar. First, the limited size of any buffer to only a single declarative chunk that can be retrieved at a time. Second, only one production rule is also selected and fired at a time [36]. Additionally, in contrast to Soar, the declarative memory of ACT-R does not have a separate episodic memory nor a method for distinguishing a memory of prior events nor the ability to preserve the temporal relations between different units of knowledge (chunks) nor a retrieval method for retrieving temporally related knowledge [41].

C. EPIC-SOAR.

In 1994, a cognitive architecture that commonly shares most of the features of ACT-R was launched under the name EPIC (Executive-Process/Interactive-Control) [37]. EPIC is a cognitive information-processing system that aims at building models of human-computer interaction for practical objectives [37, 45]. Through a set of interconnected perceptual processors (visual, auditory and tactile) working in parallel, EPIC perceive the surrounding environment and consequently produce actions via motor processors [37, 45].

EPIC has three memory stores: a declarative memory, a procedural memory and working memory [45]. Declarative memory is a long-term memory of verbal knowledge that, describe particular tasks, represented in terms of propositions [45]. The procedural memory contains procedural knowledge, for actually performing the tasks, represented as production rules [45]. Working memory contains all of the temporary information received from perceptual processors or retrieved from declarative memory, which needed for the production rules of the model [45]. Similar to Soar, EPIC is a multi-match, multi-fire production system [45]. Instead of only one production rule be fired at a time as in ACT-R, when conditions of more than one production rule match the current information of working memory, all of these satisfying rules be fired, and all of their actions be executed [45]. So, EPIC has distinguished from ACT-R in parallelism the production system in which a set of rules can be performed concurrently.

Although the success of EPIC in building models of human-computer interaction, it does not gain any benefit of its new derived facts and rules in enriching its memory stores (experiences). While, EPIC has perceptual and motor processors, it also needs to have a problem solving and learning capabilities for better performance and behavior. As noticed that, in contrast, Soar has problem-solving and learning capabilities and neither perceptual nor motor processors. Thus, an attempt to integrate the best of both systems led to a new hybrid cognitive architecture so-called EPIC-SOAR [38].

EPIC-SOAR is an integration of perceptual and motor processors of EPIC with Soar which works as the main cognitive engine [38]. In which, EPIC send messages of sensory inputs from perceptual processors to Soar, and, in turn, Soar receives the inputs, processes them and then returns motor commands to EPIC.

Table 2. Comparison between different cognitive architectures.

System	Retrieval	Knowledge Representation	Learning Technique	Goal(s)	Domain
SOAR ***(Laird et al.,2008)***	Partial matching algorithm	Production rules	Chunking and Episodic Learning	Problem Solving	Domain-specific
ACT-R ***(Anderson et al.,2004)***	Partial matching algorithm	Chunks and Production rules	Activation Learning	Problem Solving	Domain-specific
EPIC ***(Kieras et al.,1994)***	-	Propositions and Production rules	-	Human-Computer Interaction	Domain-specific
EPIC- SOAR ***(Rosbe et al.,2001)***	Partial matching algorithm	Production rules	Chunking and Activation Learning	Human-Computer Interaction and Problem Solving	Domain-specific

"-" means that a system does not have this property.

Finally, EPIC accepts the commands and executes them. It is worth mentioning that EPIC-SOAR takes and applies the idea of assigning an activation level to each declarative knowledge from ACT-R, in order to improve the behavior of the system [38]. EPIC-SOAR has been created to meet the demand of building air traffic control simulation system by the USA Air Force Research Laboratory [38].

Table 2 represents a summary of previously discussed cognitive architectures. Production rules were used as the knowledge representation technique for encoding episodic knowledge. This kind of knowledge representation allows for exact matching between the premises of the rules of the current situation and those in the episodic memory, which can be considered as a drawback because of its ignorance of contextual information, which is an essential element in social situations [9]. Events are not fixed, and they differ according to persons involved in the social situation and the context of that situation [31]. So, it seems impossible to track all rules that capture all those different norms. On the other hand, it is notoriously hard to model social and cultural situations by hand. For example, a simple model of "*eating at a restaurant*" behavior uses more than 80 production rules to capture the social dynamics related to that situation [31].

6 Conclusion

This paper analyzes the main challenges, issues and applications of the commonsense episodic knowledge from the artificial intelligence perspective. The main results based on our analysis, multi-branched cognitive script is the most efficient and proper representation technique for representing commonsense episodic knowledge. As it is a simple structure that can (a) preserve the distinct context of each social situation, (b) clearly represent a variety of different social dynamics associated with the social situation, (c) express complex social situations, and (d) be easily extended.

Concerning reasoning technique with episodic knowledge, the expressive power of multi-branched script in representing the diversity of social dynamics and the timing of events execution associated with a certain social situation, strongly allows it to be used to support reasoning from different domains (cross-domain). Which in turn permits to gain a vast amount of episodic knowledge.

References:

[1] McCarthy J. L., Programs with common sense, *In Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, London, 1959, pp. 77–91.

[2] Singh P., & Williams W., LifeNet: a propositional model of ordinary human activity, *In Proceedings of the Workshop on Distributed and Collaborative Knowledge Capture (DC-KCAP) at KCAP*, 2003.

[3] Mueller E. T., *Commonsense reasoning: an event calculus based approach*. Morgan Kaufmann, Elsevier, 2014.

[4] Davis E., & Morgenstern L., Introduction: progress in formal commonsense reasoning, A*I Journal: Special Issue on Logical Formalizations and Commonsense Reasoning, 153*(1-2), 2004, pp. 1-12.

[5] Espinosa J., & Lieberman H., EventNet: inferring temporal relations between commonsense events, *In Proceedings of the Mexican International Conference on Artificial Intelligence*, Springer Berlin Heidelberg, 2005, pp. 61-69.

[6] Singh P., Barry B., & Liu H., Teaching machines about everyday life, *BT Technology Journal*, *22*(4), 2004, pp. 227-240.

[7] Chen X., Scripts and Conceptual Change, In: P. Li, X. Chen, and H.X. Zhang (Eds.), *In Science, Cognitive, and Consciousness*, Nanchang, 2004, pp. 96-117.

[8] Tulving E., What is episodic memory?, *Current Directions in Psychological Science, 2*(3), 1993, pp. 67-70.

[9] Hodhod R., Magerko B., & Gawish M., Pharaoh: Context-based structural retrieval of cognitive scripts. *International Journal of Information Retrieval Research (IJIRR), 2*(3), 2012, pp. 58-71.

[10] Davis E., *Common sense reasoning*. In: Shapiro, E. (ed.) Encyclopedia of Artificial Intelligence. Wiley, Chichester, 1992, pp. l288–l294.

[11] Lieberman H., Usable AI requires commonsense knowledge, *In Workshops and Courses: Usable Artificial Intelligence held in conjunction CHI'08: Proceedings of the Conference on Human Factors in Computing Systems*, 2008

[12] McCarthy J. L., & Hayes P., *Some Philosophical Problems from the Standpoint of Artificial Intelligence*, In B. Meltzer and D. Michie (eds., Machine Intelligence 4, Edinburgh University Press, Edinburgh, 1969, pp. 431-450.

[13] Rich E., Knight K., & Nair S. B., *Artificial Intelligence*. Third Edition, McGraw-Hill, pp. 82, 2009.

[14] Lenat D. B., Prakash M., & Shepherd M., CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks, *AI magazine, 6*(4), 1985, pp. 65-85.

[15] Singh P., Lin T., Mueller E. T., Lim G., Perkins T., & Zhu W. L., Open Mind Common Sense: Knowledge acquisition from the general public, *In Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, Lecture Notes in Computer Science (Volume 2519), Heidelberg: Springer-Verla, 2002, pp. 1223-1237.

[16] Liu H., & Singh P., ConceptNet—a practical commonsense reasoning tool-kit, *BT technology Journal*, *22*(4), 2004, pp. 211-226.

[17] Singh P., & Barry B., Collecting commonsense experiences, *In Proceedings of the 2nd International conference on Knowledge capture*, ACM, 2003, pp. 154-161.

[18] Baker C. F., Fillmore C. J., & Lowe J. B., The berkeley framenet project, *In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, Association for Computational Linguistics, 1998, pp. 86-90.

[19] Lenat D. B., CYC: A large-scale investment in knowledge infrastructure, *Communications of the ACM, 38*(11), 1995, pp. 33-38.

[20] Singh P., The public acquisition of commonsense knowledge, *In Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, Palo Alto, CA: AAAI, 2002.

[21] Lieberman H., & Liu H., Adaptive Linking between Text and Photos Using Common Sense Reasoning, *In Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems*, (AH2002) Malaga, Spain, 2002, pp. 2-11.

[22] Liu H., Lieberman H., & Selker T., GOOSE: A Goal-Oriented Search Engine with Commonsense, *In Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems (AH2002),* Springer Berlin Heidelberg, 2002, pp. 253-263.

[23] Liu H., & Singh P., *OMCSNet: A Practical Commonsense Reasoning Toolkit* (Tech. Rep. SOM02-01), MIT Media Lab Society Of Mind Group, 2002.

[24] Gentner D., Structure-mapping: A theoretical framework for analogy, *Cognitive Science, 7*(2), 1983, pp. 155-170,.

[25] Lieberman H., Faaborg A., Espinosa J., & Stocky T., Commonsense on the go, *BT Technology Journal, 22*(4), 2004, pp. 241-252.

[26] Liu H., & Maes P., What would they think?: a computational model of attitudes, *In Proceedings of the 9th international conference on Intelligent user interfaces*, ACM. 2004, pp. 38-45.

[27] Pearl J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufman, 1988.

[28] Schank R. C., & Abelson R., *Scripts, goals, plans, and understanding*, Psychology press, Hillsdale, New Jersey:Erlbaum, 1977.

[29] Mueller E. T., *A database and lexicon of scripts for ThoughtTreasure*, CogPrints ID

cog00000555 http://cogprints.soton.ac.uk/, 1999.
[30] Fillmore C. J., Johnson C. R., & Petruck M. RL, Background to framenet, *International Journal of lexicography, 16*(3), 2003, pp. 235-250.
[31] Li B., Lee-Urban S., Appling D. S., & Riedl M. O., Automatically Learning to Tell Stories about Social Situations from the Crowd, *In the LREC Workshop on Computational Models of Narrative*, 2012.
[32] Fellbaum C., *WordNet: An electronic lexical database*, MIT Press, Cambridge, England, 1998.
[33] De Marneffe M. C., MacCartney B., & Manning C. D., Generating Typed Dependency Parses from Phrase Structure Parses, *In Proceedings of LREC,* 2006, pp. 449-454.
[34] Rosenbloom P. S. Rethinking cognitive architecture via graphical models, *Cognitive Systems Research, 12*(2), 2011, pp. 198-209.
[35] Laird J. E., Newell A., & Rosenbloom P. S, Soar: An architecture for general intelligence, *Artificial intelligence, 33*(1), 1987, pp. 1-64.
[36] Anderson J. R., Bothell D., Byrne M. D., Douglass S., Lebiere C., & Qin Y., An integrated theory of the mind, *Psychological review, 111*(4), 2004, pp. 1036-1060.
[37] Kieras D. E., & Meyer D. E., *The EPIC architecture for modeling human information-processing and performance: A brief introduction* (EPIC Tech. Rep. No. 1, TR-94/ONR-EPIC-1), Ann Arbor, University of Michigan, Department of Electrical Engineering and Computer Science, 1994.
[38] Rosbe J., Chong R. S., & Kieras D. E., *Modeling with Perceptual and Memory Constraints: An EPIC-Soar Model of a Simplified Enroute Air Traffic Control Task*. SOAR Technology Inc. Report, Ann Arbor, Michigan, 2001.
[39] Laird J. E., Extending the Soar cognitive architecture, *Frontiers in Artificial Intelligence and Applications, 171*, 2008, pp. 224-235.
[40] Langley P., Laird J. E., & Rogers S., Cognitive architectures: Research issues and challenges, *Cognitive Systems Research, 10*(2), 2009, pp. 141-160.
[41] Nuxoll A., & Laird J. E., Extending cognitive architecture with episodic memory, *Ann Arbor, AAAI, 1001*, 2007, pp. 1560-1565.
[42] Anderson J. R., & Lebiere C. J., *The atomic components of thought*, Psychology Press, Mahwah, New Jersey: Erlbaum, 1998
[43] Duch W., Oentaryo R. J., & Pasquier M., Cognitive Architectures: Where do we go from here?, *In Proceedings of the 1st conference on Artificial General Intelligence*, 2008, pp.122-136.
[44] Anderson M. L., Gomaa W., Grant J., & Perlis D., An Approach to Human-level Commonsense Reasoning, *In Paraconsistency: Logic and Applications* , Springer Netherlands, 2013, pp. 201-222.
[45] Kieras D. E., Meyer D. E., Mueller S., & Seymour T., *Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor performance*, In P. Shah & A. Miyake (Eds.) Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, Cambridge: Cambridge University Press, 1999.
[46] Liu H., & Singh P., MAKEBELIEVE: using commonsense to generate stories, *In Proceedings of the Eighteenth National Conference on Artificial Intelligence*, AAAI Press, 2002, pp 957-958.
[47] Aristotle, *Poetics*, Translated by Leon Golden, Englewood-Cliffs: Prentice-Hill, 1968
[48] Kearney R., *On Stories*, New York: Routledge, 2002.