STAIRS 2016
D. Pearce and H.S. Pinto (Eds.)

doi:10.3233/978-1-61499-682-8-111

A Hybrid Approach to Commonsense Knowledge Acquisition

Christos RODOSTHENOUS [a,1], Loizos MICHAEL [a]
[a] *Open University of Cyprus*

Abstract. This work presents a knowledge acquisition platform and a certain game developed on that platform for endowing machines with common sense, by following a hybrid approach that combines crowdsourcing techniques, knowledge engineering, and automated reasoning. Short narratives are presented to players, who are asked to combine fragments of text into rules that would correctly answer a given question, to evaluate the appropriateness of gathered rules, and to resolve conflicts between them by assigning priorities. The text fragments that are used are a priori translated by a knowledge engineer into a machine-readable predicate form. Players are rewarded based not only on their inter-agreement (as in most games with a purpose) but also based on the objective ability of the rules to answer questions correctly, as determined by an underlying reasoning engine. Beyond discussing the knowledge acquisition platform and the game design, we analyze the common sense that has been gathered during the deployment of the game over a five-month period and we use the acquired knowledge to answer questions on unknown stories.

Keywords. commonsense knowledge acquisition, knowledge engineering, Games With A Purpose

1. Introduction

This work is concentrated on one of the most challenging problems of Artificial Intelligence, the acquisition of Commonsense Knowledge (CSK) aiming in the development of programs with common sense [1] and eventually leading to the development of Cognitive Systems [2]. Over the past two decades, there has been an increase in the efforts to gather CSK. These efforts were materialized into systems that included both manually processed knowledge like the Cyc [3] and crowdsourcing approaches like the Open Mind Common Sense project [4] and its successor ConceptNet [5]. New approaches included web mining for CSK and Games With A Purpose (GWAPs), with characteristic paradigms the Never Ending Language Learner [6] and Verbosity [7]. According to Zang et al. [8], CSK can be categorized into three basic types: (i) factual knowledge, (ii) ontological knowledge and (iii) knowledge rules. Knowledge rules is the most difficult to acquire type of knowledge, since acquiring it requires human knowledge engineers to try and transfer pre-existing knowledge to a specific rule in a symbolic language that can be read by a machine.

[1]Corresponding Author: Christos Rodosthenous, Open University of Cyprus, PO Box 12794, 2252, Latsia, Cyprus; E-mail: christos.rodosthenous@ouc.ac.cy.

We take the approach that CSK rules can be acquired by harnessing the power of the crowd and combining it with efforts and work from knowledge engineers and machines. More specifically, we investigate techniques to acquire CSK rules from short narratives using a GWAP and propose a specific methodology that allows the acquisition of CSK rules, the resolution of possible conflicts using CSK rule preferences and evaluation of the appropriateness of the acquired knowledge. In the following sections, we present related work on the field of CSK acquisition, the knowledge acquisition platform we developed, an implementation of a GWAP developed using that platform, the experimental setup used for gathering CSK along with the results of this effort and an example of using the acquired knowledge to answer questions on unknown stories. Finally, we present our conclusions and future work.

2. Related Work

Currently, there are several systems that deal with the problem of CSK acquisition. Most of them employ various techniques (see Figure 3) for gathering factual CSK and only few of them deal with the problem of CSK acquisition in the form of rules. The majority of these systems use knowledge engineers as a source of knowledge and use symbolic languages (e.g., CycL, First Order Logic notation etc.) for representing acquired rules.

Sharma and Forbus [9] investigated the usage of Plausible Inference Patterns (PIP) on fully grounded queries aiming in improving the performance of QA systems. By examining the examples that the authors presented (extracted from the ReasearchCyc knowledge base), one can observe that it requires a highly trained knowledge engineer to create such rules and this process can not scale enough to gather a substantial amount of CSK rules.

Witbrock et al. [10] proposed an automated system built on top of Cyc to extract CSK rules using machine learning techniques to ground facts. As the authors state, these rules are not guaranteed to be correct, so a review and validation process is needed. Even though this is still a manual process, it is much easier to review a CSK rule than creating it from scratch.

In our previous work [11] we investigated the possibility that a fully-fledged crowdsourcing solution could be used for converting natural language (NL) to symbolic language (SL), acquiring CSK in the form of rules, generalizing these CSK rules and finally, checking their applicability and their validity. We conducted experiments using Aesop fables as the training dataset. We concluded that it is possible to gather CSK rules from the crowd, but it is difficult to use untrained subjects for converting natural language to symbolic language. The problem with the acquired CSK was that it was story specific and most CSK rules gathered could not be used in domains other than that of the stories that originated them.

Other attempts using crowdsourcing techniques and more specifically GWAPs include the “common consensus” game [12], a web based GWAP that aims in collecting and validating CSK about everyday goals. The game is based on an American TV game show called Family Feud, where players had to answer questions based on templates to extract goals. According to the authors, the game was launched for a test run with few players and the amount of unique answers retrieved were approximately 550. The extracted CSK goals are in natural language.

Other more recent attempts include the Rapport and the Virtual Pet games [13] which focus on social interactions between players. The Rapport game is based on user collaboration through a social media platform by using actions like questions, votes, etc. The virtual Pet game is deployed in a popular bulletin board and players perform actions like feeding a virtual pet and teaching it common sense, aiming in getting more common sense points. Contributions are stored in natural language and according to the authors, in a six-month period the Rapport game managed to gather 14000 statements and Virtual Pet 511734 statements.

Currently, there are only a few systems that deal with the acquisition of CSK rules using crowdsourcing, a technique that is already used for gathering CSK in the form of facts and ontologies. Reports from psychology [14] state that inference generation is a task-oriented process that follows the principle of cognitive economy enforced by a limited-resource cognitive system. Humans understand a story by integrating story related knowledge with CSK. This is due to the fact that humans have limited cognitive resources and that leads to the activation of only a small restricted subset of the available CSK [15]. Moreover, humans do not have a single CSK rule for each situation. They are more likely to have a series of rules that might be conflicting and at a given time only some of these CSK rules are activated and the rest are ignored. The notion of CSK rule preferences is introduced to describe this process. As far as we know, there are not any systems that deal with the acquisition of CSK rule preferences.

A number of systems employ techniques for reasoning with the acquired CSK. We focus on an argumentation based reasoner, the STAR system [16]. This system allows reasoning about actions and change while using CSK. Some of its major features include the ability of handling preferences between CSK rules, assigning time points for each story event and providing a question answering mechanism. The format used to code the CSK is similar to that of First Order Logic (FOL) notation.

3. Knowledge Acquisition Platform

In this section, we present the knowledge acquisition platform we designed and developed. This platform uses a hybrid approach to fascilitate the knowledge acquisition process and includes tools for conducting social and psychological experiments, deploying knowledge acquisition systems like GWAPs and other crowdsourcing applications. Users can create their own template and design their application using numerous objects and tools. Researchers have access to a graphical user interface that allows access to a series of natural language processing tools and semantic parsers like the Stanford Parser [17], OLLIE [18] and Boxer [19]. Moreover, the platform provides tools for integrating automated reasoning engines like the STAR system [16]. There is also an option for importing corpora into the system for further processing. Currently, we have implemented connectors with ConceptNet for importing CSK facts into test domains, Wordnet and Triangle-COPA [20]. The infrastructure can be enhanced to allow other datasets to be used according to the researchers' needs. This platform has already been used and tested by several researchers while creating GWAPs and preparing CSK acquisition experiments.

In this platform, we use a high-level version [21] of the Event Calculus [22] for representing the acquired CSK rules, aiming to exploit formal reasoning systems (e.g., [14]). Both implication and causal type of rules are supported [11].

For managing, controlling and monitoring the knowledge acquisition process, we have implemented an administration console for presenting information in real time and in visual form. The administration console has integrated features for managing acquired data and preparing them for further processing. These features include CSK rules filtering, junk rules detection and experiment preparation. Researchers are able to control each experiment workflow, parameterize and monitor it, view the results and analyze them. The administration console is built on top of the Joomla[2] framework. This design, allows the use of existing infrastructure for security, presentation and integration with experiment data. Researchers can also set options for the experiment, like configuring the corpus used and choosing the reasoning engine by selecting an available webservice (e.g., the STAR system webservice).

Furthermore, there are options for filtering acquired CSK based on evaluation results, type, contributions, etc. These can be grouped in a custom setup option so that they can be reused by others. The platform keeps track of all actions and keeps data in a database where both backup, security and indexing features are enabled. The knowledge acquisition platform allows researchers to build a number of crowdsourcing applications for engaging human participants in contributing knowledge.

4. A GWAP: Robot Trainer

We used the knowledge acquisition platform described previously to develop a GWAP called Robot Trainer[3]. This game aims in harnessing human player activities for contributing CSK. Player takes the role of a teacher that aims in training a robot that will travel in deep space for a long journey, so as to avoid the destructive consequences of the death of our solar system. The trained robot will be able to transfer the human knowledge needed for the continuity of our species and culture in other planets, along with embryos that will evolve into humans after arriving in their new habitat.

The goal of the player is to teach the robot how to answer simple questions on short narratives by explaining the way we think for answering such questions. Players have to construct CSK rules using natural language phrases, help the robot resolve possible conflicts with these CSK rules and evaluate the appropriateness of their fellow players contributions. Players can join the game by creating an account using their email address or their social media accounts. When authenticated, players are redirected to the "Introduction screen" of the game. There, they get to view a short two minute introduction video, take the online tutorial, select a level to play or share their game status with others in social media.

Data and Game Mechanisms Selecting an appropriate dataset for a knowledge acquisition game is not a trivial task. We seek for a dataset that has a predefined dictionary of terms and stories with situations that change through the course of time when certain events occur. Such a dataset is the Triangle-COPA which includes a set of one hundred short stories with animations and questions. These stories focus on the interactions between two triangles, a circle and a box with a door. This dataset can be extended with more stories and animations using the Heider-Simmel Interactive Theater[4]. Each story

[2]Joomla is an open source Content Management System available at https://www.joomla.org
[3]The game is available online at http://cognition.ouc.ac.cy/robot.
[4]Heider-Simmel Interactive Theater is available online at http://hsit.ict.usc.edu

is also accompanied by its representation in ISO-standard Common Logic Interchange Format, prepared by the authors of the dataset.

For using this dataset, we needed to convert each story, phrase and question in symbolic form. This is a very time consuming and prone to errors job, since it requires stories to be entered by hand by a knowledge engineer and that currently restricts the mass addition of new stories to the system and hence the scaling up of CSK acquisition. We also made some adjustments to the initial dataset, like changing predicates that were actually the negation of others (e.g., *unhappy* and *dislike* changed to *not happy* and *not like*) to reduce the number of predicates and help the automated reasoning engine. We selected a subset of that dataset that included twenty one narratives with a common theme. We randomly selected sixteen narratives for feeding the game database and five narratives that will later be used for evaluating the effectiveness of the acquired CSK rules. When a player constructs a CSK rule in natural language, it is automatically converted to symbolic language using the conversions entered initially by the knowledge engineer.

For generalizing CSK rules, we use the platform's internal mechanism to substitute all instances of "shapes" in contributed rules with variables. These variables are of type person (e.g., person(big_triangle)) since in the Triangle-COPA dataset each shape actually behaves as a person.

For the game to start, a player chooses one of the three available levels: Elementary (see Figure 1a), Advanced (see Figure 1b) and Examination (see Figure 1c). Any level can be chosen at any time and players are not required to complete a level before proceeding to the next one. We present each level in the next paragraphs, using real examples from the game and screenshots of level specific information.

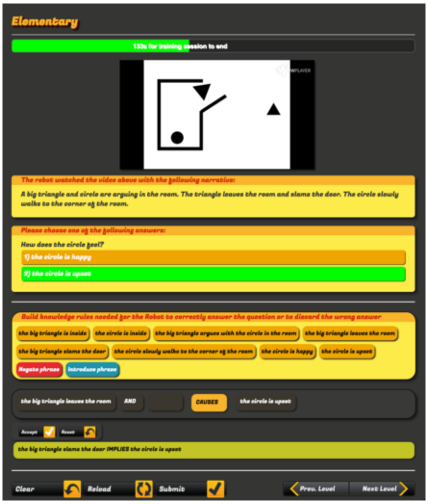
4.1. First Level (Elementary)

At the first level, a short story is selected randomly from the pool of available stories. Players read the short story accompanied by a short animation and then answer a question about that story. The next step is to build and submit CSK rules using phrases prepared by the knowledge engineers. Players can build CSK rules by dragging phrases on the body or the head of the rule (see Figure 1a). Before submitting the CSK rule, a player must choose whether it is a causal or an implication rule. When players believe that the available phrases are not sufficient for building appropriate CSK rules that answer the question, they can search for new phrases (based on a predefined dictionary of 122 phrases) by typing the first three letters of the phrase, select the desired phrase template and then select the subjects involved (e.g., big triangle (BT), little triangle (LT), circle (C) etc.).

Narrative: The little triangle was limping.
Question: Why was the little triangle limping?
Answers: [A] The little triangle is angry. [B] (*correct*) the little triangle is injured.
Rule (SL): limp(LT) `IMPLIES` injured(LT)

4.2. Second Level (Advanced)

Moving to the next level, players are instructed to help the robot in resolving possible conflicts when using specific pairs of overlapping CSK rules. These pairs are selected



(a) Elementary level.

(b) Advanced level.

(c) Examination level.

Figure 1. Robot Trainer level screenshots.

randomly by searching the pool of already acquired CSK rules. Next, we describe the selection algorithm in detail. Consider the following CSK rules: (a) BODY A `IMPLIES` HEAD A and (b) BODY B `IMPLIES` HEAD B. Then the following overlapping CSK rule pairs could lead to possible conflicts: (a) HEAD B = -HEAD A, (b) HEAD A exist in BODY B and (c) BODY B = BODY A.

A new story is created dynamically by using phrases from the overlapping CSK rules, and the player must decide if these rules are conflicting or not. If they are, the player must choose which of the two CSK rules should be discarded (less preferred), otherwise the player must state that this pair is not conflicting.

Narrative: Person A is angry and Person A plays with Person B.
Possible conflicting rule 1 (NL): Person A is angry `IMPLIES` NOT_TRUE_THAT[Person A is happy]
Possible conflicting rule 2 (NL): Person A plays with Person B `CAUSES` Person A is happy

Player's response: Rule 1 is preferable to rule 2

4.3. Third Level (Examination)

The third level of the game is the Examination. Players are instructed to evaluate the appropriateness of CSK rules added by their fellow players for helping the Robot know which rules can be generally used and which are too specific. In Figure 1c, the game level is presented showing the evaluation options. When a player selects any of these: “Completely nonsense”, “Generally false”, “Unhelpful”, “I dont know”, “Somewhat true” and “Generally true”, they are also asked to make one change to the CSK rule for making it more useful. Changes that are allowed are: add a phrase, remove a phrase, negate a phrase and change the rule type. There is also the option to proceed without doing any changes, for cases where any single change will make the rule less useful.

Rule (NL): Person A hits Person B `IMPLIES` Person A is angry at Person B
Player’s evaluation: “Somewhat true”
Change proposed: “add more phrases”

Whenever a player contributes a change on a specific CSK rule, this change is presented to a fellow player in the first level as a “tip” while building the same rule.

Help Facility We incorporated a number of help tools to the game for players to feel more comfortable in playing it. More specifically, players can read the intro of each level and then play a demo with guidance from the game itself. After doing so, they can choose to skip this step in next levels and enable it again if needed from their profile settings. Moreover, players are presented with the goals of each level throughout the game. They also have the option to view the online tutorial for a quick description of the game area, modules and controls. At any point, players have the option to contact us and provide feedback or report a bug of the game.

Player Incentives and Motives Robot Trainer is a GWAP and as with any other game of this type, players are motivated to play it for fun and of course to compete with other players. The game offers a flexible scoring framework for assigning points for various actions, like: contributing rules, contributing new rules, contributing rules that answer a specific true/false question, matching contributions of other players, contributing rules within a timeframe etc. Whenever a player contributes a CSK rule, the game automatically produces a STAR system program using the story information, the player contributed CSK rules and the story question in symbolic form. This program is sent via a webservice for execution to the reasoner server. When processing is completed, the results are returned to the game, the player receives a notification and the relevant points are added to the total score if the story question gets answered. Players can view a detailed score sheet for better understanding their score and prepare their game tactics. Besides the above score scheme, players are also informed and gain points when other players contribute CSK rules that match theirs. Players get real time information on where they stand compared to their fellow players and their progress in each mission, using the high score module. The most points are given for players that confirm other players’ contribution.

5. Experiments

The CSK acquisition experiment was active for a period of five months (153 days). During that period, players registered and played Robot Trainer GWAP. The following section presents the game, players and CSK rules analytics. Before the deployment of the game, we decided to have a short calibration period for 25 days, so that possible problems, bugs and minor improvements could be applied before deploying the final version of the game and running the experiment.

5.1. Calibration Period

During the calibration period, 24 people played the game and contributed 410 CSK rules, 182 of which were unique. The majority of contributed rules were implication (56%), whereas the causal rules were 44%. At that point, we suspected that the fact that the default option set for building a new CSK rule was the implication type, led to these results. To verify our suspicion, we decided to change this setting to the final version of the game. At the end of the calibration period we interviewed the players to get a better understanding of how they understood the game and the different levels. From the interviews, we concluded the following: (i) The first level (Elementary) was the most interesting for players, (ii) The second level (advanced) had a lot of information that was not necessary for completing the task and (iii) The third level (Examination) lack the option to select that nothing can be done to make the rule more useful. We changed the second level to make it easier for the players and we redesigned the third level for allowing evaluation of the appropriateness of the CSK rules. Also, we selected a scale similar to that of the developers of ConceptNet5 while evaluating the acquired CSK rules, so that we can compare our findings with theirs [8].

5.2. Experiment Period

In this section, we present the acquired data from the experiment period. These data include player analytics, CSK rules analytics and examples of CSK rules acquired. We also present an example of using these CSK rules to answer multiple choice questions on unknown stories.

For the period of 153 days, 799 persons played the game from various regions of the world. More specifically, we had players from Asia (72.25%), Europe (11.10%), America (10.99%), Africa (4.29%) and Oceania (0.73%). This fact, along with the fact that the experiment was not conducted in a closed, supervised environment (e.g., a lab or classroom) allowed players to contribute CSK rules without researchers intervening or influencing players in this process. The majority of players preferred the first level (Elementary). Currently, more than 50% of the registered players contributed to the game on any level. On average, a player needed 2.08 minutes for contributing a CSK rule, 0.50 minutes for contributing a CSK rule preference and 0.42 minutes for evaluating the appropriateness of a CSK rule. In terms of average contributions, a player contributed 10 rules, resolved 7 conflicts and evaluated 13 rules.

During the experiment period, players contributed 1847 CSK rules, 893 of which were unique. A CSK rule is unique if there are not any other CSK rules with the same head and body in the acquired CSK database. CSK rules with the same head and body but with different order of predicates are not considered unique. Another important finding, is that players chose to contribute simple CSK rules (i.e., rules with only one predicate at the body). Over 74.60% of the acquired unique CSK rules had a maximum of two predicates at the body of the rule. The type of acquired CSK rules is another metric we took into consideration. The majority of CSK rules contributed (67.30%) were causal. Comparing this result to that of the calibration period, we observe that most players followed the default option set while contributing CSK rules.

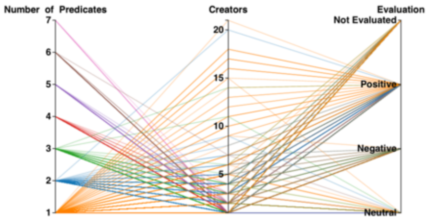


Figure 2. An overview of the acquired CSK rules per number of predicates, creators and evaluations.

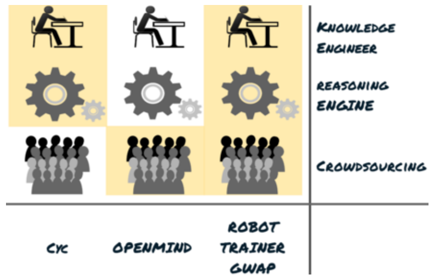


Figure 3. Comparison of three systems based on the knowledge acquisition technique they use.

Robot Trainer also collects CSK rule preferences between possible conflicting pairs. From the 893 unique CSK rules we have gathered, we detected 31199 overlappings that could lead to possible conflicting pairs (see Section 4.2). The majority of overlapping CSK rules (60.37%) were of type B (a predicate at the head of one rule is present in the body of the other). Players were presented with some of these pairs during the second level of the game. Players contributed on resolving 1053 (371 unique) of them. From those 371 pairs, 52 (14.02%) were reported as not possible to lead to conflicts.

Players also evaluated a number of CSK rules while playing the third level. For 1847 contributed CSK rules, players provided 1501 evaluations. For better filtering and presentation of the results, we grouped “Somewhat true” and “Generally true” as “Positive” evaluations, “Completely nonsense”, “Generally false”, “Unhelpful” as “Negative” evaluations and “I dont know” as “Neutral” evaluations. When a CSK rule has equal number of “Positive” and “Negative” evaluations, it is considered as “Neutral”. In terms of unique CSK rules, 415 (46.47%) of 893 CSK rules or 350 (39.19%) if “Neutral” answers were ignored, were evaluated by at least 1 evaluator. Players evaluated 221 (63.14%) CSK rules as “Positive” out of definite responses (i.e., the responses discarding “Neutral” evaluations).

In a similar evaluation process, Witbrock et al. [10] reported that reviewers marked 7.5% of the acquired CSK rules as “correct” and 35% as “correct with minor adjustments”. Moreover, comparing these results, with results from the evaluation of ConceptNet 5 [8], our methodology lays at the middle of the range (60%-70%) of facts gathered from WordNet, Wiktionary (English-only), and Verbosity in ConceptNet database, and reviewed by evaluators. Comparison between the two systems cannot lead to safe conclusions, since ConceptNet deals with gathering CSK of different type.

In terms of CSK rule evaluation speed, Robot Trainer allows the evaluation of 143 CSK rules per hour. In similar measurements [10], a reviewer evaluated 20 CSK rules per hour.

Furthermore, players added a “Positive” evaluation to simple CSK rules (i.e., rules with one or two predicates in the body) instead of more complex ones. More specifically 62.07% of the “Positive” evaluated CSK rules had one predicate and 22.99% had two predicates. Figure 2 presents an overview of the acquired CSK rules and shows that most contributors prefer building simple CSK rules (orange and blue lines).

5.2.1. Examples

In this section, we present examples of CSK rules acquired during our experiments.

R1: injured(A) `CAUSES` limp(A)
R2: hug(A,B) `IMPLIES` like(A,B)
R3: hit(A,B) `IMPLIES` angry(A)
R4: pull(A,B) `CAUSES` -happy(B)
R5: hug(A,B) `CAUSES` happy(A)
R6: argueWith(B,A) `CAUSES` -happy(A)

R1 was contributed by 21 players and evaluated by 12. 58.33% evaluated this CSK rule as "Somewhat true" and "Generally true". R2 was contributed by 17 players and evaluated by 35. 85.71% evaluated this CSK rule as "Somewhat true" and "Generally true". R3 was contributed by 16 players and evaluated by 24. 79.16% evaluated this CSK rule as "Somewhat true" and "Generally true".

R4 is an example of a not so useful CSK rule gathered during the acquisition process. It was contributed by 6 players and evaluated by 19. 26.31% evaluated this CSK rule as "Somewhat true" and "Generally true". This CSK rule most probably would not be included in any knowledge database due to its low evaluation score.

In terms of CSK rule preferences acquisition, consider the CSK rules R5 and R6. Five players contributed on resolving possible conflicts between R5 and R6. More specifically, players were presented with a short story where B argues with A and A hugs B. 60% reported that R5 is preferable to R6.

5.2.2. Question Answering Using the Acquired CSK

The acquired CSK rules can be used to answer questions on unknown narratives. For demonstrating this, we prepared the following experimental setup: First, we used the 5 randomly selected narratives from the Triangle-COPA dataset that were not seen by the game players. Each of these stories was accompanied by a multiple choice question with 2 possible answers. Then, we created a knowledge pool using CSK rules acquired previously using the Robot Trainer GWAP. More specifically, we selected CSK rules that were evaluated by at least 2 evaluators, the majority of the evaluators added a "Positive" evaluation and they had a maximum of 4 predicates in the body of the CSK rule. We also used CSK rule preferences that were chosen by the majority of the contributors.

We aim in correctly answering as many questions as possible using only the CSK acquired from players and the STAR system. The STAR system returns three possible results for each question: "accepted", "rejected" and "possible". A question is answered if any of the following conditions are met: (a) the STAR system responds with a different definite result ("accepted" or "rejected") for both answers or (b) the STAR system responds with a definite result ("accepted" or "rejected") for one of the 2 possible answers and the result for the other answer is "possible". Responses are the result of the STAR system reasoning process, that finds arguments to support or defeat a possible answer. From the 5 narratives processed with the automated reasoning engine, we retrieved answers to all questions. From the 5 questions, we retrieved correct answers for the 4 of them. Details of the processing procedure are depicted in Table 1.

For better understanding the question answering process, we present the "Argue and trudge" narrative example with a subset of the CSK rules used in the reasoning process.

Narrative (NL): The little triangle wants to go out and party with its friends but it's mom wants it to do its homework. The little triangle goes to sulk in the corner.
Narrative (SL): argueWith(BT,LT) at 1, inside(LT) at 1, moveTo(LT,corner) at 2.
Question (NL): Why does the little triangle trudge to the corner of the room?
Answers (SL): (A) -happy(LT) at 4 or (B) happy(LT) at 4

Table 1. Processing results for the 5 narratives.[R], [A] and [?] indicate that the answer is rejected, accepted and possible respectively. The answer in bold text is the correct one.

Narrative title	Question/Answer	Process time
Date night	**[A] friend(LT,C)** | [?] stranger(LT,C) and stranger(C,LT)	0.45 min
Cold outside	[?] -happy(LT) | **[A] cold(LT)**	0.23 min
Run and hug	[R] -happy(BT) | **[A] excited(BT)**	1.10 min
Argue & Trudge	**[A] -happy(LT)** | [R] happy(LT)	0.27 min
Punch wall	**[?] -goal(angry(BT),BT)** | [A] excited(BT) and happy(BT)	3.83 min

The following rules are used for building the argument to support that the little triangle is happy at time point 4.

R1: fight(X,Y) implies -happy(Y)
R2: argueWith(Y,X) implies fight(Y,X)
R3: argueWith(Y,X) implies -happy(X)
R4: argueWith(Y,X), moveTo(X,corner) implies -happy(X)

6. Discussion and Future Work

In this paper, we presented a hybrid methodology and a knowledge acquisition platform that bridges three different approaches of gathering CSK; knowledge engineers, automated reasoning and crowdsourcing. We presented a GWAP developed using the platform components and the results of the acquisition process. Results of this methodology are comparable to other systems, with the difference that this system is not only used to gather CSK in the form of rules, but it gathers CSK rule preferences and evaluates a CSK rule appropriateness. Acquired CSK can be used for story understanding tasks [23], question answering systems and more complex applications like cognitive agents.

There are still several problems needed to be solved for deploying this methodology in large scale, like the problem of automating story conversion from natural language to symbolic language. The process of transforming the Triangle-COPA dataset to symbolic language suitable for the STAR system, required a number of changes to the original dataset and hence added overhead to the overall effort needed.

In future versions of the game, we plan to make some changes to the mechanism that selects the CSK rules that will be evaluated or presented for possible conflict resolution. Currently, we use a random selection algorithm, but this mechanism allows specific CSK rules to be evaluated by many, whereas others are not evaluated. This happens when a CSK rule is introduced early in the game or by many, and players are presented with this rule more often. The solution to this, is to change the selection algorithm to present CSK rules that have the lowest number of evaluations first.

We envision using the game in other concepts, like teaching. The game can be used as a learning activity for non-English language speakers. The fact that the game uses simple English phrases is ideal for practising while studying English language courses. Also, the fact that the game allows players to answer questions can be modified to allow self assessment of students.

Furthermore, we aim in extending this methodology along with the knowledge acquisition platform presented, so as to provide tools for researchers to build similar applications and test various methodologies both in gathering CSK and for using it in real applications.

References

[1] J. McCarthy. Programs with Common Sense. In *Proceedings of the Symposium on the Mechanization of Thought Processes*, pages 75–91, 1959.

[2] L. Michael, A. Kakas, R. Miller, and G. Turán. Cognitive Programming. In *Proceedings of the International Workshop on Artificial Intelligence and Cognition*, volume 1510, pages 3–18, 2015.

[3] D. Lenat. CYC: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.

[4] P. Singh. The Public Acquisition of Commonsense Knowledge Push Singh The Diversity of Commonsense Knowledge. Technical report, USA, 2002.

[5] H. Liu and P. Singh. ConceptNet - A Practical Commonsense Reasoning Tool-kit. *BT Technology Journal*, 22(4):211–226, 2004.

[6] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Jr. Hruschka, and T. M. Mitchell. Toward an architecture for never-ending language learning. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 2010.

[7] L. von Ahn, M. Kedia, and M. Blum. Verbosity: A Game for Collecting Common-Sense Facts. In *Proceedings of the Conference on Human Factors in Computing Systems*, page 75, Québec, 2006.

[8] L. J. Zang, C. Cao, Y. N. Cao, Y. M. Wu, and C. G. Cao. A Survey of Commonsense Knowledge Acquisition. *Journal of Computer Science and Technology*, 28(4):689–719, 2013.

[9] A. Sharma and K. D. Forbus. Graph-Based Reasoning and Reinforcement Learning for Improving Q/A Performance in Large Knowledge-Based Systems. In *Proceedings of the AAAI Fall Symposium*, 2010.

[10] M. J. Witbrock, C. Matuszek, A. Brusseau, R. C. Kahlert, C. B. Fraser, and D. B. Lenat. Knowledge Begets Knowledge: Steps towards Assisted Knowledge Acquisition in Cyc. Technical report, 2005.

[11] C. T. Rodosthenous and L. Michael. Gathering Background Knowledge for Story Understanding through Crowdsourcing. In *Proceedings of the 5th Workshop on Computational Models of Narrative*, volume 41, pages 154–163, Canada, 2014.

[12] H. Lieberman, D. A. Smith, and A. Teeters. Common Consensus: A Web-Based Game for Collecting Commonsense Goals. In *Proceedings of the ACM Workshop on Common Sense for Intelligent Interfaces*, USA, 2007.

[13] Y. Kuo, J. C. Lee, K. Chiang, and R. Wang. Community-based Game Design: Experiments on Social Games for Commonsense Data Collection. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 15–22, France, 2009.

[14] I. Diakidoy, A. Kakas, L. Michael, and R. Miller. Story Comprehension through Argumentation. In *Proceedings of the 5th International Conference on Computational Models of Argument*, pages 31–42, UK, 2014.

[15] R. J. Gerrig. The Scope of Memory-Based Processing. *Discourse Processes*, 39(2-3):225–242, 2005.

[16] I. Diakidoy, A. Kakas, L. Michael, and R. Miller. STAR: A System of Argumentation for Story Comprehension and Beyond. In *Proceedings of the 12th International Symposium on Logical Formalizations of Commonsense Reasoning*, pages 64–70, 2015.

[17] M. De Marneffe, B. MacCartney, and C. D. Manning. Generating Typed Dependency Parses from Phrase Structure Parses. In *Proceedings of the Language Resources and Evaluation Conference*, pages 449–454, Italy, 2006.

[18] Mausam, M. Schmitz, R. Bart, S. Soderland, and O. Etzioni. Open language learning for information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534, USA, 2012.

[19] J. R. Curran, S. Clark, and J. Bos. Linguistically Motivated Large-scale NLP With C&C and Boxer. In *Proceedings of the 45th Annual Meeting of the ACL*, pages 33–36, Czech Republic, 2007.

[20] N. Maslan, M. Roemmele, and A. S. Gordon. One Hundred Challenge Problems for Logical Formalizations of Commonsense Psychology. In *Proceedings of the 12th International Symposium on Logical Formalizations of Commonsense Reasoning*, USA, 2015.

[21] L. Michael. Computability of Narrative. In *Proceedings of the 2nd Symposium on Computational Models of Narrative*, USA, 2010.

[22] M. Shanahan. The Event Calculus Explained. In Michael J. Wooldridge and Manuela Veloso, editors, *Artificial Intelligence Today*, Lecture Notes in Computer Science, pages 409–430. 1999.

[23] E. T. Mueller. Story understanding through multi-representation model construction. In *Proceedings of the HLT-NAACL workshop on Text meaning*, pages 46–53, USA, 2003.