UNIVERSITÄT KOBLENZ · LANDAU

Fachbereich 4: Informatik

Towards Explainable Creativity: Tackling the Remote Association Test with Knowledge Graphs

Master Thesis

A thesis submitted for the
Master of Science in Web Science

Submitted by

Rakshitha Gnanasekaran
Mat.-Nr. 218203283
rgnanasekaran@uni-koblenz.de

First supervisor: Prof. Dr. Ulrich Furbach

Second Supervisor: Dr. Claudia Schon

UNIVERSITÄT KOBLENZ · LANDAU

Universitätsbibliothek in Koblenz
- Bestandserschließung -
Universitätsstr. 1
56070 Koblenz

Universitätsbibliothek in Landau
- Bestandserschließung -
Fortstr. 7
76829 Landau

Erklärung zur Abgabe einer elektronischen Publikation

Bitte kreuzen Sie die für Sie zutreffenden Angaben an und geben Sie das Formblatt unterschrieben bei der zuständigen Bibliothek ab. Die E-Mail-Adresse und die Telefonnummer werden für eventuelle Rückfragen benötigt, aber nicht publiziert.

Verfasser/Verfasserin:	Rakshitha Gnanasekaran
Titel der Publikation:	Towards Explainable Creativity: Tackling the Remote Association Test with Knowledge Graphs
E-Mail-Adresse	rgnanasekaran@uni-koblenz.de
Telefonnummer:	+49 17656750035
Tag der mündlichen Prüfung **(nur für Dissertationen):**	
Erstgutachter **(nur für Dissertationen):**	
Fachbereich/Institut:	Fachbereich 4: Informatik

Es handelt sich um eine ☐ Dissertation ☐ Habilitation ☒ sonstige Prüfungsarbeit ☐ sonstige Schrift

1. Hiermit übertrage ich der Universitätsbibliothek Koblenz-Landau (nachfolgend UB genannt) unwiderruflich das einfache Recht, meine Publikation sowie die von mir angefertigten Abstracts elektronisch zu speichern und in Datennetzen sowie bei Bedarf auf anderen Speichermedien öffentlich zugänglich zu machen. Darin enthalten ist auch das Recht der Weitergabe der Publikation an die zuständigen Pflichtexemplarbibliotheken sowie bei Bedarf an Sondersammelgebietsbibliotheken. Eine Vergütung erfolgt nicht. Meine sonstigen Urheberrechte werden nicht eingeschränkt. Ich kann die Arbeit jederzeit zusätzlich ganz oder teilweise veröffentlichen.

2. Die UB kann ihr übertragene Rechte an einen Dritten weitergeben, sofern dieser generell die Aufgabe der längerfristigen Archivierung der elektronischen Publikation von der UB übernimmt.

(Stand: 07/2019)

3. Die UB ist aufgrund ihrer längerfristigen Archivierungsaufgabe berechtigt, meine abgelieferten Dateien bei Bedarf in ein technisch moderneres Datenformat zu konvertieren. Daher akzeptiert die UB nur PDF-Dateien ohne Kopierschutz und ohne Verschlüsselung.

4. Die UB garantiert die Zitierfähigkeit des Werkes - insbesondere die Integrität des intellektuellen Inhalts bei jeder Art von Formatumwandlung - und stellt die Archivierung sicher, soweit es im Rahmen der technischen Entwicklung mit vertretbarem Aufwand möglich ist.

5. Eine kommerzielle Nutzung meiner Publikation durch die UB ist ausgeschlossen. Kostenerstattungen oder eine von Rechts wegen vorgeschriebene Erhebung von Gebühren stellen keine kommerzielle Nutzung dar.

6. Hiermit erkläre ich, dass mit der Bereitstellung des Dokuments und jedes Bestandteils (z.B. Abbildungen) keine Rechte Dritter (z.B. Miturheber, Co-Autoren, Verlage, Verwertungsgesellschaften, Drittmittelgeber) verletzt werden. Der Verfasser/die Verfasserin versichert, dass das Dokument keine personenbezogenen Daten Dritter enthält, die dem Datenschutz unterliegen. Der Verfasser/die Verfasserin versichert, in Zweifelsfällen oder bei Entstehen diesbezüglicher Rechtshindernisse die UB Koblenz-Landau unverzüglich in Kenntnis zu setzen. Der Verfasser/die Verfasserin stellt die Universität Koblenz-Landau von etwaigen Ansprüchen Dritter frei. Die UB Koblenz-Landau ist berechtigt, den Zugriff auf ein Dokument zu sperren und das Dokument zu löschen, soweit konkrete Anhaltspunkte für eine Verletzung von Rechten Dritter bestehen.

7. Ich erkenne an, dass jede Veränderung oder Rücknahme der Publikation nach dem Zeitpunkt der Veröffentlichung aus Gründen der Sicherung der Zitierfähigkeit ausgeschlossen ist.

8. Ich versichere, dass die von mir abgelieferte Arbeit vollständig in Form und Inhalt mit der vom Betreuer genehmigten Version übereinstimmt (**nur für sonstige Prüfungsarbeiten**).

9. Hiermit versichere ich, dass die Drucklegung der Dissertation in der endgültigen Fassung vom Promotionsausschuss genehmigt wurde und die elektronische Version mit den abgegebenen Druckexemplaren in Form und Inhalt im Rahmen der Regelungen der Promotionsordnung übereinstimmt. Der Doktorand überträgt der Hochschule das Recht, im Rahmen der gesetzlichen Aufgaben der Hochschulbibliotheken weitere Kopien von seiner Dissertation herzustellen und zu verbreiten bzw. in Datennetzen zur Verfügung zu stellen. (**nur für Dissertationen**).

10. Ich versichere, dass das vorstehende Erklärungsformblatt von mir nicht verändert wurde.

Koblenz, 10.09.2021	Rakshitha.
Ort, Datum	Unterschrift Verfasser/Verfasserin

Für sonstige Prüfungsarbeiten:
Ich stimme der Veröffentlichung o.g. Arbeit über den Hochschulschriftenserver sowie im Internet zu.

Koblenz, 10.9.2021	schon@uni-koblenz.de	C. Schon
Ort, Datum, Stempel des FB/Instituts	E-Mail-Adresse	Unterschrift Betreuer/Betreuerin

Universität Koblenz-Landau
WeST – Institute for Web Science & Technologies
Universitätsstraße 1
D-56070 Koblenz

(Stand: 07/2019)

Abstract

Remote Association Test (RAT) is a creativity test that assesses participants' creativity by measuring their associative ability. Various AI frameworks can perform computational creativity tests like the Remote Association Test to measure AI systems' cognitive and problem-solving abilities. However, the state-of-art, CreaCogs cannot propose explanations for these solutions. Thus, my master's thesis's proposed aim is to implement an AI system that can solve RAT computationally by acquiring knowledge from a common-sense knowledge base and word embeddings and constructing explanations for these RAT solutions. The proposed approaches are implemented and evaluated with the state-of-art and the normative data of Bowden and Jung (2013). The study concludes that knowledge from ConceptNet provides a plausible approach to solve RAT computationally and explain "why" an answer is related to the RAT query.

Acknowledgement

I want to thank my supervisor, Dr. Claudia Schon, for her support and the great opportunities she has given me over the last few years. I want to express my gratitude for her advice on providing insights into various methodologies to be carried out in my master thesis, feedback on how to present the research work clearly, and guidance throughout this thesis work.

I would like to acknowledge and express my profound thanks to my thesis supervisor Prof. Dr. Ulrich Furbach, for giving me the opportunity to do this master thesis.

Additionally, I would like to thank my friend, Mr. Rejnald Lleshi, for his genuine support throughout this master thesis.

Finally, I express my heartfelt thanks to my parents and my loved ones for their continuous encouragement, financial support, and assistance throughout my years of study.

Table of Contents

List of Figures

List of Tables

1 Introduction

According to the Cambridge dictionary, creativity[1] is defined as "the ability to produce an original and unusual idea, or to make something new or imaginative." When one hears the word creativity, our mind links this ability with highly creative and valued people like Beethoven, Mozart, or Picasso. However, in reality, creative ability is found in all humans in various activities like validating a mathematical proof, writing poems or novels, painting, acting, music, fashion, solving riddles, and much more. Traditionally, researchers claimed that creativity is strongly associated with the arts (Davies and Lynn Newton, 2018). Nevertheless, it is required in various other fields like archaeology or history to explain why an event has occurred or in engineering to develop innovative ideas or to solve complex problems.

Creative problem solving is one of the topics that interest both Cognitive Scientists and researchers in Artificial Intelligence. One aim of cognitive scientists is to build various A.I. systems that can solve creative problems and answer questions like 'How the human mind works while solving a creative task?' For Artificial Intelligence, creative problem-solving systems can help modeling agents solve complex problems with novel ideas

However, creativity is hard to be measured in humans as it cannot be defined entirely (since a perceiver consider a particular piece/artifact creative only based on its novelty and value: provides benefits to the content and different perceiver have different views) (Wiggins, 2006). But some empirical tests (Maier, 1931; Duncker, 1945) like the Remote Association Test (RAT), the Alternative Uses Test, the Empirical Insight Test, the Torrance Test of Creative Thinking or Riddles (Duncker, 1945; Kim, 2006) are available to evaluate the level of human creativity. Some of these creativity tests uses insight problems like "*Which would be worth more, a pound of 10-dollar pure gold coins or half a pound of 20-dollar pure gold coins; or would they be worth the same? Explain your answer"*(Gayle T. Dow and Mayer, 2004)[2] and these problems are sometimes difficult to solve. The participants take about thirty minutes to propose a solution, making it impossible to address more than a few problems in a single session, which leads to a lack of variety and generalization of conclusion (Oltețeanu, 2020). Nonetheless, Remote Association Test measures insight problems in a shorter duration where

[1] https://dictionary.cambridge.org/dictionary/english/creativity
[2] Answer: The insight is to use gold as a unit of comparison. So, a pound of gold is worth more than half a pound of gold.

many test queries can be answered in less than 30 seconds. For example, a test query like *Manners, Round, Tennis*[3] is provided to participants, and they must come up with the fourth word that relates these words. RAT hence provides a large variety and quantity of test items to the participants making it more widely used creativity test after the Torrance Tests of Creative Thinking (Kim, 2006).

According to Colton and Wiggins (2012), computational creativity is defined as "The philosophy, science, and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative." Computational creativity can be found in various fields like mathematics, music, art, poetry, design, architecture, and games. Like empirical studies measures human creativity, computational creativity can be measured using various models like the Wiggins' model, the FACE IDEA descriptive model, the CreaCog Framework. CreaCogs framework, the state-of-art, tries to measure computational creativity by building an integrated system that can solve the Remote Association Test. Yet, the CreaCogs framework fails to explain why a particular answer is chosen during the creative problem-solving.

1.1 Research Question and Approach

This section explains about the research question answered in my master thesis and the approaches used to solve them. The proposed aim of my master's thesis is to:

1. Build a system which can computationally solve Remote Association Test
2. Construct explanation for the Remote Association Test solutions.

What is Remote Association Test?

Remote Association Test (RAT) is a creativity test, which measures the associative ability of remote items, and according to Schooler & Melcher (1995) exhibiting better in this test corresponds to successfully solving insight problems. In RAT, a participant is provided with three remote words like *SWISS, CAKE* and *COTTAGE*[4] and participants should come up with a fourth word associated with all three initial test words. There are two kinds of RAT: Functional and Structural RAT. Functional RAT has a functional meaning on why an answer

[3] Answer: TABLE
[4] Answer: CHEESE

is related to the initial test words. For example, Figure 1, represents a functional RAT query; were as a structural RAT, have a language association by forming compound words. For example, *SWISS*, *CAKE,* and *COTTAGE,* and the answer *CHEESE* is associated with the initial test query by forming compound word.

This section explains the first part of the research question on how knowledge is acquired and prepared to develop a creative problem-solving system that solves RAT. CreaCogs framework, proposed by Olteţeanu and Falomir (2015), designs RAT Knowledge Base using bigram from a publicly available Corpus of Contemporary American English (COCA)[5]. However, building a knowledge base using bigrams would lack essential details on "how" and "why" a noun is related to another noun, making it difficult to create explanations from the knowledge base. Thus, the approach followed in this master's thesis is to use a common-sense knowledge base like ConceptNet or WordNet for knowledge acquisitions. Then the Initial RAT test words/query is sent to ConceptNet, and all the nodes that the test query is related to are retrieved. A detailed explanation is provided in Chapter 5. To validate the practicality of the proposed system, 144 normative compound RAT queries of Bowden and Jung (2003) and 48 normative functional RAT queries of Olteteanu, Schöttner and Schuberth (2019) will be fed into the designed systems, and then this system is evaluated based on the answer candidates.

The second part of the research question is the main objective of this master thesis, which would be to construct explanations for the answer candidate and the RAT queries. Explanations for RAT can be achieved by understanding the semantic relation between the test queries and the solution of RAT from ConceptNet. Supposedly, the initial query as in Figure 1 provided to the ConceptNet was *DAISY, TULIP,* and *VASE* [6]; the proposed system will look for an association for the initial three query words. Figure 1 illustrates that the nouns *Daisy*, *Tulip* and *Vase* have an association with *Flower* and various semantic explanations (compound relation), as *'Location*_of' and *'Type_of'*. So, the system should be able to construct an explanation like *'Daisy is a type of flower', 'Tulip is a type of flower',* and *'Things located at vase is a flower'.* This implication can be tested with various RAT queries, and the explanations proposed computationally can be presented to humans to judge and evaluate the feasibility of the presented approach.

[5] http://corpus.byu.edu/coca/
[6] Answer: Vase

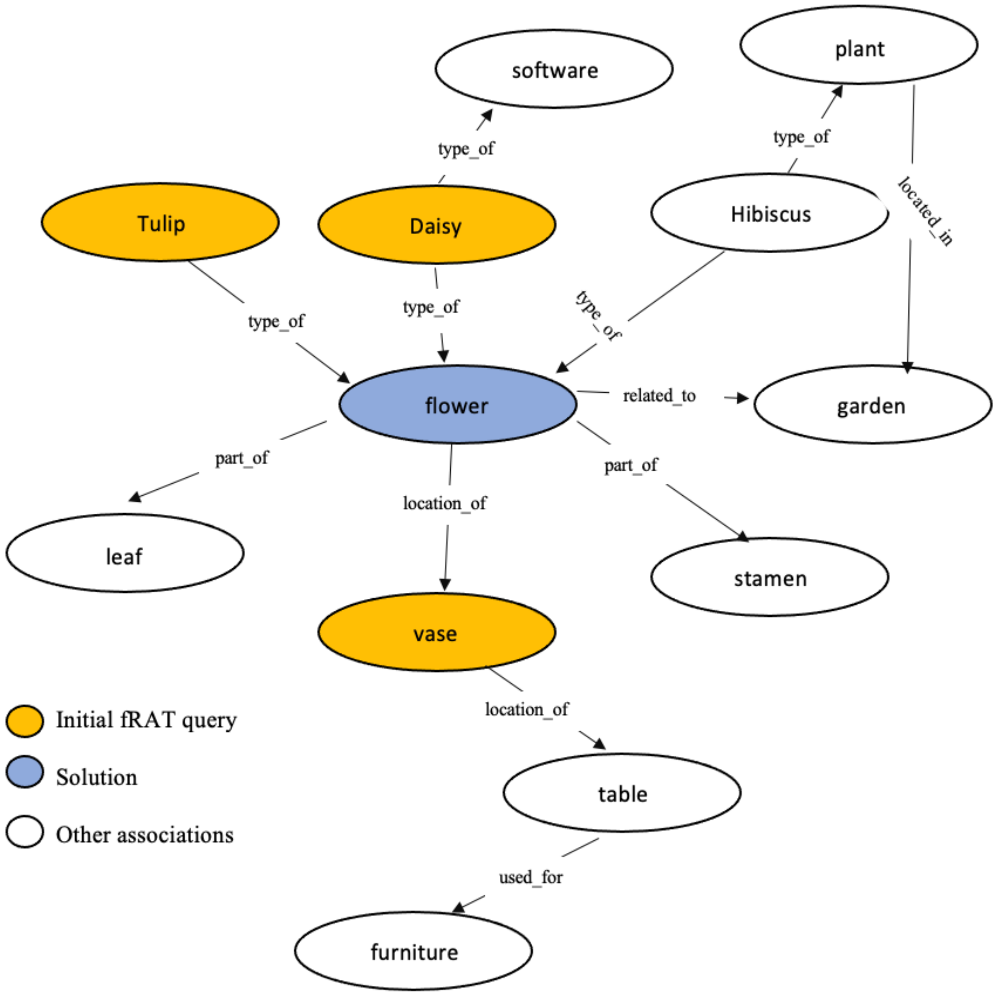

Figure 1 Example of functional RAT query using ConceptNet

1.2 Contributions and Findings

The contribution of these works are as follows:

1. This study presents a computational approach to solve RAT queries using knowledge from ConceptNet. The proposed approach can also apply to other knowledge bases like WordNet. The study also suggests solving RAT queries based on determining the nearest neighbour (calculating a similarity score) using Word Embeddings.
2. The study also presents a way to provide a reason "why" a particular answer candidate is chosen for a RAT query. This approach can also be applicable in other ontologies like DBpedia.

The empirical findings of this research are as follows:

1. Functional RAT solved by identifying the nodes with path length 2 had a better accuracy score compared to solving using knowledge graphs.
2. Structural RAT, being a compound word, solved using knowledge graphs did not perform as better as Functional RAT, as knowledge graphs like ConceptNet represent common-sense knowledge.
3. Explanations provided on "why" an answer candidate is related to a RAT query was plausible using ConceptNet. However other ontologies like Dbpedia can be also considered.

1.3 Overview

The remaining parts of this master thesis are structured as follows: Chapter 2 delves on theoretical backgrounds about human and computational creativity, followed by Chapter 3, which focuses on related works in computational creativity. In Chapter 4, the normative dataset for RAT is discussed, and Chapter 5 elaborates on Research methods and experimental results. In Chapter 6, Limitations for this approach and Future works are discussed. This master thesis is then concluded with Chapter 7 with a brief overview.

2 Theoretical Background

2.1 Measuring Human Creativity

Though creativity is hard to be measured in humans, Guilford in the year 1967, proposes that creativity can be evaluated and studied using psychometric approach with pen and paper (Guilford, 1967). Later some empirical tests like the Alternative Uses Test, the Torrance Test of Creative Thinking, Riddles, the Empirical Insight Test, or Remote Association Test RAT were available to evaluate the level of creativity by asking the participants to solve insight problems (Maier, 1931; Duncker, 1945).

2.1.1 Alternative Uses Test

Naturally, a creative problem does not have one perfect solution; instead, diverse possible solutions. Divergent thinking is an essential part of creative problem-solving. In the Alternative Uses Test (Guilford, 1967), participants are provided with an everyday object (ex: paper clip). Now, the participants should develop a list of non-obvious uses for this object in a specific duration of time(generally 1 to 3 minutes). For example, for *paper clips*, one can come up with the following responses (Dippo, 2013):

- Holding items
- Decorations
- Scraping things
- Unclog
- Weapon
- Bracelet
- Necklace

Depending on the novel ideas generated by the participants, human creativity is measured using four categories (Guilford, 1967; Oltețeanu, 2020):

- Fluency: Fluency is the number of uses the participants provided. In the above example, the Fluency score would be 7
- Flexibility: The flexibility score measures the number of different categories that the participants provided. In this example: *bracelet* and *necklace* have the same category as jewelry and are counted as one category.

- Originality: Originality is the measure of novel ideas in comparison to other human participants. Suppose *scrapping items* is provided by 15% of the population and *weapon* is provided by 5%, then *weapon* is considered to be highly imaginative.
- Elaboration: Elaboration measures how detailed the solution provided by a human participant is (however, there is a contradiction between fluency and elaboration).

2.1.2 Torrance Tests of Creative Thinking (TTCT)

Ellis Paul Torrance developed Torrance Test of Creative Thinking where a participant is provided with a drawing as shown in Figure 2 and asked to use or combine or complete it. There are two main parts in TTCT: TTCT-Verbal and TTCT-Figural (E. P Torrance, 1998; E.P. Torrance, 1998; Torrance and Ball, 1998; Kim, 2006)

TTCT Figural has three different types on how participants can be tested:

- ***Use:*** Picture construction (where participants are given one particular shape like a circle, and they are asked to construct a picture using that shape like in Figure 2).
- ***Combine:*** Repeated figures (where participants are presented with multiple shapes, and they are asked to use those shapes to construct a complete image like in Figure 2).
- ***Complete:*** Picture completion (where participants are provided with an incomplete figure and their task is to complete it, for example, in Figure 2).

TTCT Verbal has five different types of tasks:

- ask and guess (where participants ask questions based on a given drawings).
- guessing causes and consequences (where the participants are required to guess about the cause and consequence of the event related to a given drawing).
- product improvement (where participants have to come up with an improvement for the product, for example[7]: “try to improve a stuffed toy so that it will be more fun to play with”).
    - unusual uses (like Alternative Uses Test, participants should come up with as many uses as possible)

[7] http://home.iitk.ac.in/~sahus/se367/project/TTCT.pdf

- just suppose (participants are provided with an unusual idea, and they should predict plausible outcomes).

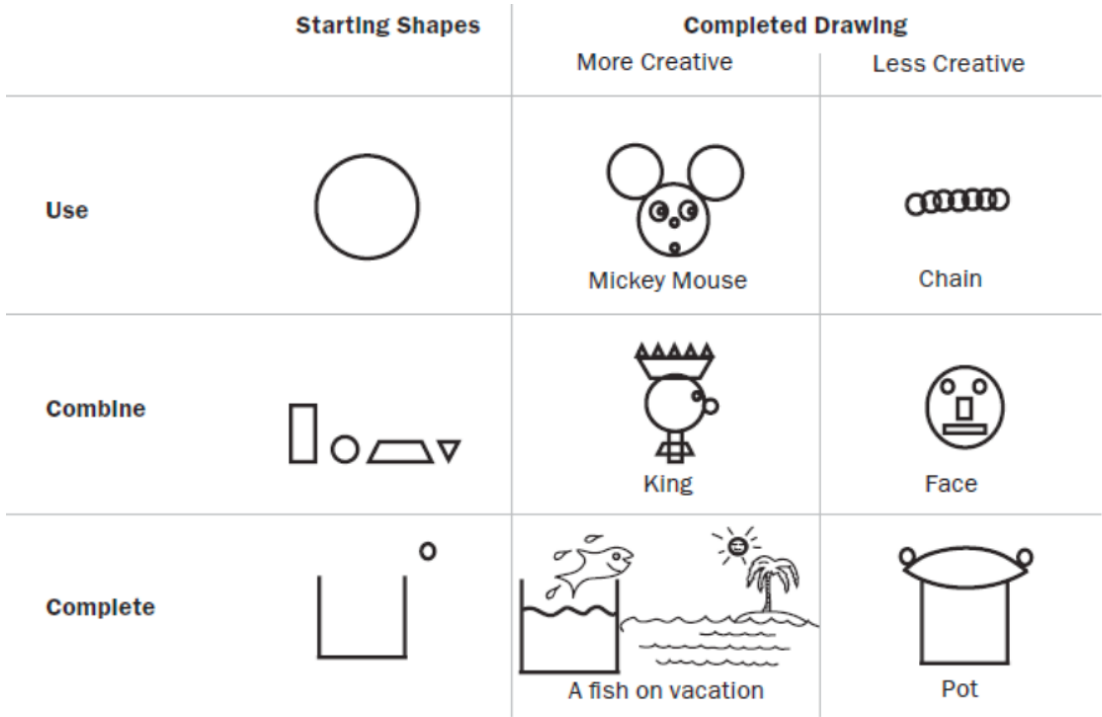

Figure 2 Example of TTCT (Jiménez, Fernández-cosials and Mínguez, 2017)

2.1.3 Riddles

Interpretation and performance on riddles are used to measure human creativity (Whitt and Prentice, 1977; Qiu *et al.*, 2008). Unlike other insight problems, there are no comprehensive set of riddles problems. However, there are two types of riddles depending on the type of resource needed to solve them: enigmas and conundrums.

Enigmas are a category of a riddle that is expressed in allegorical or metaphorical language. Example: "Which creature has one voice and yet becomes four-footed and two-footed and three-footed?[8]" Conundrums are language-based riddles where the words have a different meaning. Example: "What gets wetter as it dries?[9] "

[8] Answer: Human. As a baby, one crawls in 4 feet; as an adult, one walks in 2 feet; as an elder, one walks with a walking stick.
[9] Answer: Towel

2.1.4 Empirical Insight Test

When solving an Insight problem, a participant should change their perspective in a novel way to achieve the desired solution. By this novel approach, creativity in humans can be measured and depending on the type of knowledge the participants evoke (Olteţeanu, 2020); insight problems can be divide into a different category. Some examples can be Mathematical Insight problems, Verbal insight problems, Spatial Insight problems, and much more. The following are examples of insight problems (G. T Dow and Mayer, 2004). Mathematical Insight problem: *In the Smith family, there are seven sisters, and each sister has one brother. If you count Mr. Smith, how many males are there in the Smith family?*[10]. Verbal Insight problem: *Three women - Joan, Dana, and Sandy - have among them three children - Sam, Traci, and David. Sam likes to play with Dana's son. Sandy occasionally babysits for Joan's children. Who is Traci's mother?*[11]. Spatial Insight Test: As shown in figure 3a, *how can you arrange six identical pencils in such as way as to form 4 identical triangles whose sides area are all equal without modifying the pencils in any way?* Answer: Figure 3b.

Figure 3a Question

Figure 3b Solution

2.1.5 Remote Association Test (RAT)

Supposedly one comes across the words *SWISS*, *CAKE,* and *COTTAGE*[12]. What do they have in common? It is an absurd question unless one has engaged in the experiment conducted by Mednick and Mednick on Remote Association Test. Mednick and Mednick proposed the Remoted Association Test (RAT) in 1971 to measures creativity based on the participants' ability to propose remote associations (Mednick, 1962). According to Schooler J.W (1995), "performing good in the RAT has shown to correspond with the ability to solve insight

[10] Answer: Two (the father and the brother)
[11] Answer: Joan
[12] Answer: Cheese

problems successfully"(Schooler J.W., 1995). To examine this associative ability, participants are provided with three query words like *DEW, COMB, BEE,* and these participants should draw a fourth term that associates with the initial three query words. In this case, the correct answer would be *HONEY* because *HONEY* can be associated with the query as *'Honey Dew,' 'Honeycomb,' 'Honeybee'* (Mednick, S.A., Mednick, 1971). RAT has been translated to various languages other than English (Nevo and Levin, 1978; Hamilton, 1982; Akbari Chermahini, Hickendorff and Hommel, 2012). Since RAT highly relies on the existing association within the language, it cannot be directly translated. A query like *FISH, MINE, RUSH* can only be translated and adapted in a language if there is an existing relationship. If the compound word '*Goldfish'* cannot be found in the language, then the query becomes impractical and might involve generating a new set of queries for that language. Worthen and Clark (1971) studied that the test queries in a RAT are a blend of structural and functional associates. In structural associates, words occur together, forming compound words: *'Goldfish'* or *'Steering Wheel'*. Moreover, these words have a syntactic structure and lack a functional relationship. In functional associates, there is a functional relationship between the words: *flower* is located in *vase*[13]*, or eggs* are laid by *birds*[14]; moreover, both these queries go beyond language association.

2.2 Measuring Computational Creativity

Computationally innovative AI systems can be found in various fields like mathematics (Colton, 2012a), music (Pearce and Wiggins, 2004), art (Colton, 2012b), poetry (Colton, Goodwin and Veale, 2012), design and architecture (Schneider, Fischer and König, 2011), video games (Cook and Colton, 2014), etc, and more AI systems are being developed every year. Below is a brief overview of few AI systems.

- **AM Model:** The field of AM model is 'elementary set and number theory' in mathematics. The database of AM model has about 115 fundamental mathematical concepts (like heuristics and equality), which was used by the AI system to reengineer concepts from set and number theory (Lenat, 1976).

[13] https://www.conceptnet.io/c/en/flower
[14] https://www.conceptnet.io/c/en/eggs

- **HR Model:** The HR model was named after two mathematicians, Hardy and Ramanujan, and is used to produce new mathematical theories using a pre-defined set of production rules (Colton, 2012a).
- **The painting fool:** The creative system[15] designed by Colton is used to create art in real-time by drawing stokes one by one to a canvas. This AI system can paint or stimulate in different styles like pencil, charcoal, acrylics, chalks based on emotions because the knowledge base of this AI system is annotated based on emotions (Colton, 2012b).
- **The poetry system:** Colton's poetry system uses a template (structure), word associations, and smileys. The poem is built by this AI system using templates and constraints like sentiment, rhyme, stress, word frequency. The database that this AI system uses is usually a corpus or pre-established lexicon. Below is an example of a poem(Colton, Goodwin and Veale, 2012): *"Relentless attack a glacier-relentless attack the wild unprovoked attack of a snake the wild, relentless attack of a snake a relentless attack, like a glacier the high-level function of eye sockets a relentless attack, like a machine the low-level role of eye sockets a relentless attack, like the tick of a machine the high-level role of eye sockets a relentless attack, like a bloodhound"*

As we have seen in Section 2.1. (Measuring human creativity) on how human creativity is measured empirically, computational creativity can also be measured using various systems. There are many assays on assessing computational creativity. Few such systems are

- The Wiggin's model of universe of possibilities and Transformational Creativity (Wiggins, 2006).
- Ritchie's typicality criteria and the inspiring sets (Pereira *et al.*, 2015).
- The FACE and IDEA descriptive model (Colton, Charnley and Alison Pease, 2012).
- Evaluating Machine Creativity (Pease *et al.*, 2001)
- CreaCogs Framework (Oltețeanu, 2020)

2.3 Knowledge Graphs

Common-sense knowledge is defined as a usually and generally used knowledge about everyday life. According to a commoner, common sense is perceived as 'good sense and sound

[15] http://www.thepaintingfool.com

judgment'; however, according to AI scientists, common-sense knowledge indicates millions of basic facts and understanding. Common-sense knowledge is built from spatial, physical, social, and temporal aspects of everyday life (Liu and Singh, 2004). Candies are sweet; Black coffee with no sugar is bitter; To drink water from a bottle, one must first open the cap are some examples of basic understanding of the facts from human experience, which are used to form common-sense knowledge.

When a sentence like 'I ate chips while watching Netflix' is provided to a common-sense deprived computer system, it will not make logical insights since it might link chips to computer chips rather than crisps or when a sentence like 'I got fired today' a computational solver can make the assumption that it is a negative emotion by spotting the negative word 'fired'. So when such a sentence is provided to a computer with common-sense knowledge, it should be able to make conclusions like 'maybe happy because did not like the job,' or 'because one is incompetent,' or ' sad because when fired, no salary.' There are various knowledge bases available to make this interpretation, and a brief overview of a few are as follows.

Graphs (Example Figure 4) are used to represent the knowledge base, where the nodes depict conceptual entities like *car*, *drive, shift, vehicle etc.,* and the edges describe the nature of relations between the nodes like *UsedFor, PartOf, IsA etc.* ConceptNet, WordNet, and Cyc are the most distinctly used large-scale semantic knowledge bases in works of literature.

Cyc[16] is commonly used for establishing common-sense knowledge into a logical framework. It has about 1.6 million facts associating with more than 11800 concepts. CycL, a representation language for Cyc, is used for mapping text which needs to be reasoned into a proprietary logical representation of Cyc (Liu and Singh, 2004). This mapping of text to Cyc representation leads to the difficulty of textual reasoning in Cyc. Another disadvantage of Cyc is that it is not fully available to the public.

WordNet[17] is also widely used in the computational linguistic community as semantic knowledge. This common-sense knowledge base is a collection of 200 000 distinct words of primary nouns, verbs, and adjectives, and the database of words is connected through semantic-

[16] www.cyc.com

[17] https://wordnet.princeton.edu/download/current-version

relations like *synonyms* or *is-a* hierarchical relations (Kilgarriff and Fellbaum, 2000; Liu and Singh, 2004).

On the other hand, ConceptNet[18] is a large-scale common-sense knowledge graph with about 1.6 million edges connecting more than 300 000 nodes (Liu and Singh, 2004), available for free and easy to adopt. ConceptNet and WordNet vary significantly on the node and link type. WordNet includes just words with a clear focus on linguistic knowledge and is disambiguated. In contrast, ConceptNet (Figure 4) focuses more on common-sense knowledge by having compound relations like *UsedFor, DesireOf, ISA, LocationOf, PartOf,* which tries to explain an object, place, or place property, or an action (Singh, Barry and Liu, 2004)**.**

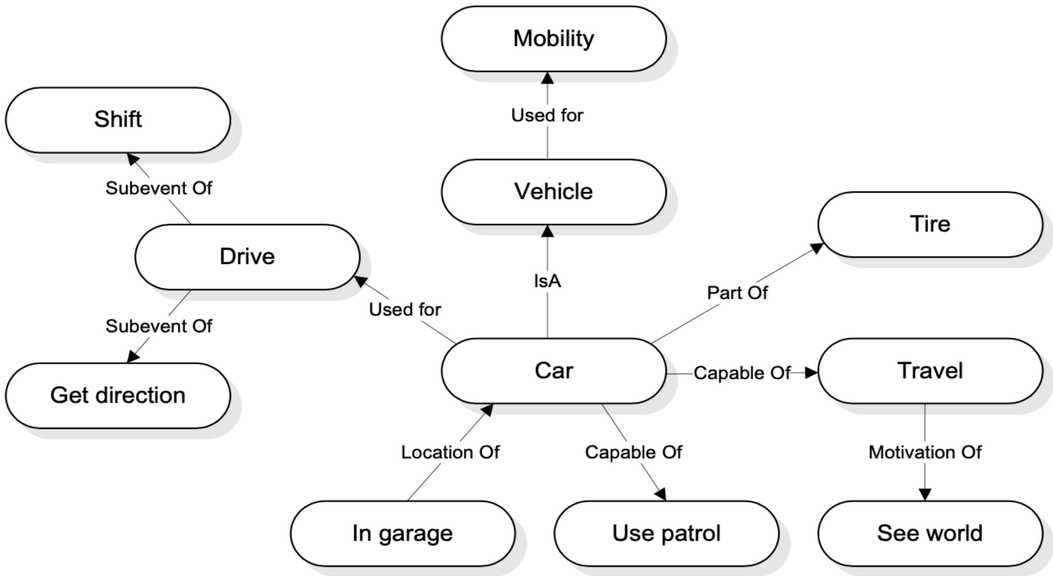

Figure 4 Snippet of ConceptNet (Ebersold, 2008)

[18] https://conceptnet.io

3 Related Works

3.1 CreaCogs

Participants of RAT claims that the 4th word associated with the initial three query words comes to mind spontaneously(Olteţeanu, 2020). Olteţeanu (2020) proposes a hypothesis that a variety of words used in RAT might have a previous association in the agents' memory. When these initial 3 query words are given to the agent, an association to these query words is activated. Sometimes this association can activate all three query words creating a convergence. These convergences are stimulated from long-term memory. Creative upward search is a methodology that implements this hypothesis in AI systems like the CreaCogs framework, which solves creativity problems(Olteţeanu, 2020). There are four cognitive AI systems that are relevant to RAT: comRAT-C is a cognitive system that determines the associated term for the initial queries in RAT, comRAT-G generates RAT queries, fRAT propose functional RAT, and visual RAT broadens RAT to the visual domain.

3.1.1 ComRAT-C

Olteţeanu and Falomir (2015) proposed a computational RAT solver popularly known as ComRAT-C, which investigates a cognitive system's associative ability. To test this A.I. cognitive system, initially three queries $w_{a,}$ $w_{b,}$ and w_c are fed into the system, and the system then determines an answer word w_{ans}. (Olteţeanu and Falomir, 2015)

ComRAT-C has three types of knowledge structure: Concepts, expressions, and links. Concepts are one-word lexical terms. Expressions are compound words or two Concepts occurring together in a language. Links are bidirectional connections between concepts in an expression.

During knowledge procurement, comRAT-C is continuously provided with bi-grams from a corpus[19]. Initially, an expression class is constructed, and comRAT-C checks whether both the concepts in the expression are present in the knowledge base. If present, a link between the

---

[19] Corpus of Contemporary American English (COCA): http://corpus.byu.edu/coca/

concepts is created. Else, the unknown concept is added to the knowledge base, and then a link is created. For example, if a ComRAT-C is provided with a bi-gram *Cottage Cheese*, the system checks whether both *Cottage* and *Cheese* concepts are present in the knowledge base. If one or both the concepts are not present in the knowledge base, then the unknown concept is added, and a bi-directional link is created between them. If both the concepts are present in the knowledge base, the system creates a bi-directional link between them. By successfully providing bi-grams, links are created between concepts belonging to the same expression. After creating the knowledge base, the initial three query words (w_a, w_b, and w_c) are given. ComRAT-C searches the query words in its knowledge base. If found, the query word activates the concepts to which they are linked. Subsequently, all the expressions are activated. As seen in Figure 5, this activation of expression is irrespective of the concepts' position: first term or last term.

Figure 6 shows how concepts are activated in ComRAT-C. The three initial query words *COTTAGE, SWISS, CAKE* are depicted in green. These concepts now activate various other concepts that are linked in blue. Two initial concepts activate the concept in yellow *Chocolate*. While the answer, *Cheese*, the concept in red, is activated by all three concepts. Suppose ComRAT-C does not find a three-word convergence, then it proposes a concept with two- word convergence.

(Cottage ∗) OR (∗ Cottage)	(Swiss ∗) OR (∗ Swiss)	(Cake ∗) OR (∗ Cake)
cottage **cheese**	Swiss Alps	cake batter
cottage garden	Swiss army	cake decorating
cottage industries	Swiss ball	cake flour
cottage ...	Swiss chard	cake layer
... cottage	Swiss **cheese**	carrot cake
... cottage	Swiss chocolate	**cheese** cake

Figure 5 Example of activation(Olteţeanu, 2020)

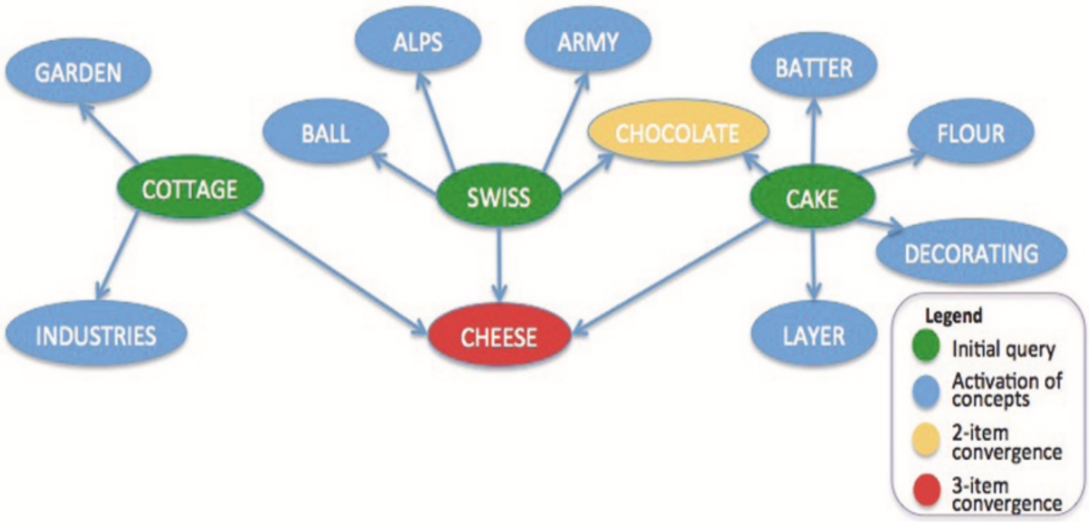

Figure 6 Visual Depiction of concept activation (Zunjani and Olteţeanu, 2019)

3.1.2 ComRAT-G

RAT is the second most used test to measure creativity (Arden et al., 2010) and it is also widely used for evaluating creativity in the literature (Ansburg and Hill, 2003; Barton Cunningham et al., 2009). Olteţeanu and Falomir (2015) designed a ComRAT-C that solves RAT queries computationally. In 2019, Zunjani and Olteţeanu studied that the knowledge base used in ComRAT-C can not only solve RAT queries, but they can also propose various initial RAT queries. Zunjani and Olteţeanu created an A.I. system called ComRAT-G, which can create RAT queries by reversing the convergence (Zunjani and Olteţeanu, 2019). At first, the system retrieves nouns that can be a possible answer (w_{ans}) with the words they are linked to ($w_{q}s$). ComRAT-G chooses w_{ans} that has at least three $w_{q}s$ (Zunjani and Olteţeanu, 2019). For example, ComRAT-G iterates over the knowledge base and retrieves '*STAR*'(w_{ans}). The system also retrieves it's linked query words ($w_{q}s$): *MOVIE, ROCK, POP, NEUTRON, FORMATION, BASKETBALL, POWER, FOOTBALL, WITNESS, FILM, SYSTEM, CLUSTERS, CLUSTER, PLAYER, TRACK, TENNIS, SHOOTING, GUEST, ANISE, CHILD* (Olteţeanu, 2020). In this way, ComRAT-G can propose various altered initial queries like '*MOVIE, ROCK, FILM*' or '*PLAYER, CLUSTER, NEUTRON,*' which has the same associative answer *STAR*. Unlike ComRAT-C, which can be validated by the study done by Bowden and Jung, the altered initial queries generated by ComRAT-G cannot be validated (Olteţeanu, 2020). ComRAT-G is the first cognitive A.I. system to generate RAT queries and enhance creativity and psychometric test (Oltețeanu and Schultheis, 2019).

3.1.3 ComRAT-G$_F$

Test queries in RAT is a combination of structural and functional associates. However, according to Worthen and Clark (1971), a mix of both the types of associates in an initial three query word like Mednick's is not efficient enough to find the level of creativity and proposed a new fRAT based on functional associates (Worthen and Clark, 1971). However, the fRAT examples proposed by Wooden and Clark went missing during transportation, and later Olteţeanu, Schultheis and, Dyer (2018) created an A.I. system that regenerated fRAT computationally (Olteţeanu, Schultheis and Dyer, 2018).

ComRAT-G$_F$ is a cognitive A.I. system to generate fRAT (functional RAT) instead of structural RAT. fRAT is generated by using the same ComRAT-C system. Instead of extracting Nouns, words from a dataset: (Nelson, McEvoy and Schreiber, 2004) were extracted from ComRAT-C's knowledge base. This dataset listed words from human participants when given a cue word. For example: when participants are presented with a cue word *APPLE*, they came up with the words *TREE*, *PIE*, *TART*. Another example: when cued with *ABUNDANCE*, words like *FAMINE*, *FOOD*, *FULL* were presented by participants. The knowledge base for comRAT-G$_F$ has a similar structure to ComRAT-G; similarly, words having at least three associations were considered a plausible answer and can be used as fRAT queries (Zunjani and Olteţeanu, 2019). Studies conducted by Olteţeanu, Schultheis and Dyer (2018) showed that ComRAT-G$_F$ creates a highly reliable fRAT, which also correlates with the accuracy and response time of ComRAT-G's structural query RAT.

3.1.4 vRAT

RAT, a language-based test, also helps humans in problem-solving using creativity. However, some tricky problems might need linguistic as well as visual creativity. Visual Remote Association Test, called vRAT, is a cognitive A.I. system proposed by Olteţeanu, Gautam, and Falomir (2015) that evaluates visual and linguistic creativity.

For vRAT, the initial elements e_1, e_2, e_3, and the answer to be found e_{ans} are visual elements. Figure 7 shows an example of vRAT, where a participant is presented with three images *BATHTUB*, *GLASS* and *BEACH* (e_1,e_2,e_3). Participants formulate an associative term *WATER*.

For a participant to solve a linguistically based RAT query, one must know at least two out of three associative words. While in vRAT, a participant makes an association based on their experience with the objects shown in the picture (Oltețeanu, 2020).
As shown in Figure 3, these visual elements were designed carefully such that the answer is not depicted in the picture (The bathtub and the glass are empty, i.e., no water, a part of the beach where no sea or ocean is depicted).

Figure 7 Example of vRAT: BATHTUB, GLASS, BEACH (Olteţeanu, Gautam and Falomir, 2015)

vRAT uses the following approach (Olteţeanu, Gautam and Falomir, 2015):

- Images replace words and, scenes replace expressions.
- Instead of a structural or functional linguistic relationship, a visual(sensory) relationship is created.
- Participants carry out visual relationships rather than linguistic relationships.

vRAT can be used to produce queries in the visual domain, thus broadening creativity.

4 Dataset

Mednick built two sets of tests, each consisting of 30 items (Mednick and Mednick, 1967; Mednick, 1968). In Mednick's RAT test, each test query consists of three words (*wa, wb,* and *wc*), and the answer candidate can be linked to the test query in several aspects. For example, the test query *SAME, TENNIS, HEAD* has the solution *MATCH; MATCH* is associated with *SAME* because they are synonyms, *MATCH* is associated with *TENNIS* because of semantic association, and *MATCH* is associated with *HEAD* as they form a compound word: '*matchhead'*. Later Bowden and Jung further claimed that Mednick's RAT queries were a combination of two types: language-based association (structural) and associations that arise beyond language (functional).

4.1 Structural Remote Association Test

Bowden and Jung wanted a more significant number of consistent test queries than Mednick's initial set: meaning, the solutions are always associated with the initial test queries in a single way, for example, *DREAM, BREAK, LIGHT* where the solution, *DAY*, is associated with the test query by forming compound word: *DAYDREAM, DAYBREAK, DAY LIGHT*. These queries that are formed based on language association are called the compound Remote association test by Bowden and Jung, which resulted in 144 compound Remote Association Test queries (Bowden and Jung-Beeman, 2003).

4.2 Functional Remote Association Test

Worthen and Clark, used 20 of Mednick's RAT queries as functional RAT. However, these set of RAT queries from the annex of Worthen and Clark's paper were lost during transportation when the National Auxiliary Publications Service was dissolved to the Library of Congress. Hence currently there are no normative dataset for functional RAT queries (Olteteanu, Schöttner and Schuberth, 2019).

Olteteanu, Schöttner and Schuberth, 2019 proposed a computational model which generates functional RAT queries, using the dataset from Nelson et al. This dataset listed words from human participants when given a cue word. For example: when participants are presented with a cue word *APPLE*, they came up with the words *TREE*, *PIE*, *TART*. Another example: when cued with *ABUNDANCE*, words like *FAMINE*, *FOOD*, *FULL* were presented by participants. Later Olteteanu, Schöttner and Schuberth evaluated these functional RAT queries with human participants and proposed 48 normative functional RAT queries.

A list of complete functional and structural RAT queries are attached in the Appendix.

5 Research Methods and Experimental Results

Worthen and Clark (1971) studied that the test queries in RAT are a blend of structural and functional associates. Functional RATs are words when they have a functional meaning or relationship between them. For example: Consider the query: *DAISY*, *TULIP*, *VASE* (*wa, wb,* and *wc*), and the answer *FLOWER*, this query is a functional RAT as *DAISY* and *TULIPS* are type of *FLOWERS* and *FLOWERS* are contained in *VASE*.

In structural RAT, words occur together, forming compound words: *'GOLDFISH'* or *'STEERING WHEEL.'* Moreover, these words have a syntactic structure forming a structural relationship and lack a functional meaning. These structural RATs are also called compound associates, as these RAT forms compound words.

This section describes the various research methods that are adopted to solve the research questions. Along with the approaches that are employed, their experimental results are also discussed in this section.

5.1 Research Question 1: "Build a system which solves functional RAT."

A cognitive system needs knowledge, and acquiring data required for an AI system is called knowledge acquisition. In the below approach to computationally solve functional RAT knowledge acquisition is done using ConceptNet.

5.1.1 Approach 1 – Using the three-word intersection

To answer the research question where the objective is to design a model which solves functional RAT following steps are done:

Constructing a set of connected nodes: The functional RAT query (*wa, wb,* and *wc*) is looked up using ConceptNet, and then the proposed system retrieves all the nodes that the functional RAT query (*wa, wb,* and *wc*) is connected to.

connected(w_a) = set of nodes connected to node w_a in ConceptNet. Similarly,

connected(w_b) = set of nodes connected to node w_b in ConceptNet.

connected(w_c) = set of nodes connected to node w_c in ConceptNet.

From Figure 8 for query *QUESTION*, *REPLY*, *SOLUTION*,

connected(w_a) = enquire, interrogate, cross-examine, sentence, statement, answer.

connected(w_b) = statement, answer, sentence, response, talk

connected(w_c) = solvent, salination, method, statement, answer

Determining the answer candidates: The system finds the intersections between these retrieved node sets and determines the answer word(s) (*Wans*). There are two ways in which an answer node can be activated: Three-word intersection and Two-word intersection. A three-word intersection is when the answer node is an intersection between all the three query words (*wa, wb* and *wc*) in a functional RAT query. A two-word intersection is when the answer node is an intersection between any two RAT query words (that is either between (*wa, wb) or (wb, wc) or (wa, wc)*).

As in Figure 8, if the functional RAT query (*wa, wb,* and *wc*) is *QUESTION*, *REPLY*, *SOLUTION*

connected(w_a) = enquire, interrogate, cross-examine, sentence, **statement**, **answer**.

connected(w_b) = statement, **answer**, **sentence**, response, talk

connected(w_c) = solvent, salination, method, **statement**, **answer**

where **answer** and **statement** are activated by three-word intersection between all the three query words.

A clearer understanding of this process with a visual depiction of how the proposed approach works is shown in Figure 8.

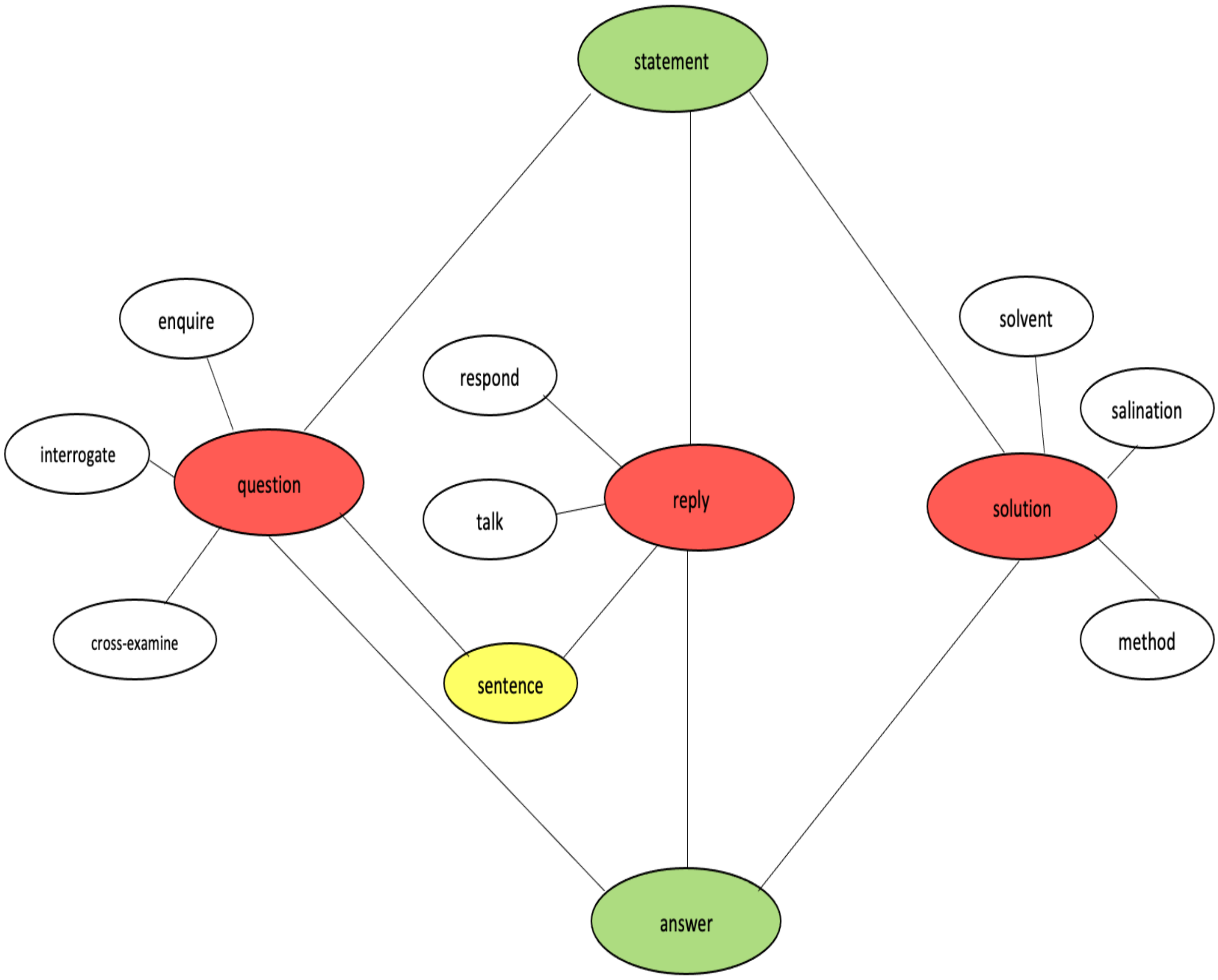

Figure 8 Visual Depiction of Functional RAT

In Figure 8, the functional RAT query words: *QUESTION*, *REPLY*, *SOLUTION* (*wa, wb,* and *wc*) are shown in red. The node *SENTENCE,* shown in yellow, is triggered by a two-word intersection of the functional RAT query words *QUESTION* and *REPLY* (*wa,wb*). The nodes *ANSWER* and *STATEMENT,* shown in green, is triggered by three-word intersection of the functional RAT query (*wa, wb,* and *wc*) which is also the answer candidates. Since the first approach followed to answer the research question is using a 'three-word intersection' only answer candidates which are activated by the three-word intersection is chosen as the answer candidates (*Wans*).

Processing the answer candidates (*Wans*)**:** According to the method based by Olteţeanu and Falomir (2015), only nouns from the *Wans* are selected to be the final output. WordNet is used for this lexical analysis of *Wans*. For example, the functional RAT query, *BENCH, SOFA, STOOL* (*wa, wb,* and *wc*) provides CHAIR*, SEAT, SITTING, FURNITURE* to be the answer candidates (*Wans*)*; the* system is made to neglect the word *SITTING*[20] from the final *Wans* as it is an adjective. Table 1 shows few examples of functional RAT (*wa, wb,* and *wc*) and the answer that the model provides (*Wans*) from the proposed approach from Section 5.1.1.

	w_a, w_b, and w_c	w_{ans}
1	question, reply, solution	statement, answer
2	bud, dandelion, petals	plant, flower
3	discuss, gossip, telephone	talk
4	bench, sofa, stool	chair, seat, furniture
5	flu, nauseous, virus	-
6	sensitive, sob, weep	-
7	crown, royalty, throne	-
8	dictionary, verse, vocabulary	-
9	fault, incorrect, justice	-
10	marsh, saliva, slippery	-

Table 1 Examples of Function RAT

As shown in Table 1, some of the functional RAT queries (6 - 10) have no three-word intersection leading to no answer candidate. In contrast, some queries have more than one node from the ConceptNet as their answer candidates (W_{ans}) and to determine which of the node from the answer candidates (W_{ans}) best suits the functional RAT query (*wa, wb,* and *wc*), GloVe embedding is used. The Glove is one of the most widely used Word Embeddings which maps words into higher-dimensional vector space. Euclidean or similarity score can be calculated for the vectors in this vector space. By calculating the similarity scores between the answer candidates (from the W_{ans}) and the words in functional RAT query (*wa, wb,* and *wc*), a single node from W_{ans} is selected. Figure 9 explains this similarity calculations. For queries that have no intersection, another research methodology is discussed in Section 5.1.2.

[20] https://www.thesaurus.com/browse/sitting

Calculating Cosine similarity scores for answer candidates: To choose one node from the W_{ans} that the proposed approach provided, cosine similarity between w_a, w_b, and w_c, and the W_{ans} are calculated as in Table 2.

w_a, w_b, and w_c	W_{ans}	similarity score
question, reply, solution	answer	question-answer: 0.797 reply-answer: 0.627 solution-answer: 0.449
	statement	question-statement: 0.47 reply-statement: 0.372 solution-statement: 0.227

Table 2 Cosine similarity score for QUESTION, REPLY, SOLUTION

Comparing cosine similarity between the answer candidates: Then, the highest average among the W_{ans} is chosen to be the final answer node (*answer)*. For example, consider functional RAT query *QUESTION, REPLY, SOLUTION* (w_a, w_b, and w_c); the average score of the query with *ANSWER* is 0.624 and with *STATEMENT* is 0.356. Hence the model chooses the final answer (*answer)* (*answer* column as in Table 3) to be *ANSWER*. Table 3 shows few examples for other functional RAT queries and the column names *'answer'* is the final answer that the model chooses after calculating the similarity score.

Given the functional RAT query w_b , w_b , w_c

connected(w_a) = set of nodes connected to node w_a in ConceptNet
connected(w_b) = set of nodes connected to node w_b in ConceptNet
connected(w_c) = set of nodes connected to node w_c in ConceptNet

W_{ans} = connected(w_a) ∩ connected(w_b) ∩ connected(w_c)

cos_sim(x,y) = cosine similarity between x and y

If w_1 , w_2 ∈W_{ans},

cosine_similarity_w_1= (cos_sim(w_a,w_1) + cos_sim(w_b,w_1) + cos_sim(w_c,w_1)) /3

cosine_similarity_w_2= (cos_sim(w_a,w_2) + cos_sim(w_b,w_2) + cos_sim(w_c,w_2)) / 3

If cosine_similarity_w_1 > cosine_similarity_w_2 then w_1 is chosen to be ***answer***

Figure 9 Similarity calculation

Figure 9 shows how *answer* is chosen from the W_{ans}. 'cos_sim' from the Figure 9 is the cosine similarity function which calculates the similarity score using GloVe word embedding.

Table 3 Cosine Similarity for few functional RAT queries

w_a, w_b, **and** w_c	W_{ans}	**similarity score**	***answer***
question, reply, solution	answer statement	question-answer: 0.797 reply-answer: 0.627 solution-answer: 0.449 question-statement: 0.47 reply-statement: 0.372 solution-statement: 0.227	answer
bud, dandelion, petals	flower, plant	bud - flower: 0.504 dandelion - flower: 0.486 petals - flower: 0.696 bud - plant: 0.387 dandelion - plant:0.34 petals - plant: 0.331	flower
discuss, gossip, telephone	talk		talk
bench, sofa, stool	chair, seat, furniture	bench - chair: 0.546 sofa - chair: 0.671 stool - chair: 0.567 bench - seat: 0.49 sofa - seat: 0.468 stool - seat: 0.421 bench - furniture: 0.412 sofa - furniture: 0.67 stool - furniture: 0.503	chair

Experimental Results

To evaluate the approach proposed in Section 5.1.1., the *answer* (from Table 3) provided by the proposed model is compared with the solution from the normative data of Olteteanu, Schöttner and Schuberth (2019). In Table 4, the '*ground solution'* column is the solution from the normative data.

w_a, w_b, and w_c	w_{ans}	*answer*	**ground solution**
question, reply, solution	answer statement	answer	answer
bud, dandelion, petals	flower, plant	flower	flower
discuss, gossip, telephone	talk	talk	talk
bench, sofa, stool	chair, seat, furniture	chair	chair
flu, nauseous, virus	-	-	sick
sensitive, sob, weep	-	-	cry
crown, royalty, throne	-	-	king
dictionary, verse, vocabulary	-	-	wrong
fault, incorrect, justice	-	-	words
marsh, saliva, slippery	-	-	wet

Table 4 Experimental Results (Approach 1: functional RAT)

When reviewed this first research approach to solve functional RAT query using three-word intersection with the 48 queries from the normative data proposed by Olteteanu, Schöttner and Schuberth (2019), the designed model answered 12 queries correctly providing an accuracy of 25%. While 36 queries (like the last six queries from Table 3), didn't have any intersection, and this limitation is tried to be solved in section 5.1.2.

Evaluating the answer candidates W_{ans}

How is W_{ans} evaluated? The answer candidate (W_{ans}) is evaluated with *ground solution* from Olteteanu, Schöttner and Schuberth (2019) and if the correct ground solution is present in the list of answer candidates (W_{ans}) from the proposed approach then it is considered as a correct solution.

When evaluated the W_{ans} with the normative data, the proposed model provided with the same accuracy of 25%.

5.1.2 Approach 2 – Using the two-word intersection

There are some functional RAT queries where three-word intersection does not provide any answer candidate, meaning W_{ans} is empty. Examples of few such queries are represented in Table 5. Even as a human participant, when given a functional RAT query, and supposedly the participants know only two of the query word association, a mild guess of the third item can be made. Hence to make this computationally possible, a method based by Olteţeanu and Falomir (2015) is used to solve this limitation. When the proposed system fails to find a three-word intersection, then the system looks for two-word intersections. As depicted in Figure 8, the designed approach looks for two-word intersection which is shown in yellow.

Table 5 Some Functional RAT queries with no three-word intersection

w_a, w_b, and w_c	W_{ans}
flu, nauseous, virus	-
sensitive, sob, weep	-
crown, royalty, throne	-
dictionary, verse, vocabulary	-
fault, incorrect, justice	-
marsh, saliva, slippery	-

.

Constructing a set of connected nodes: The functional RAT query (*wa, wb,* and *wc*) is looked up using ConceptNet, and then the proposed system looks for a two-word convergence, that is, the proposed model now looks for nodes between all the possible combinations of the initial functional RAT query ((*wa, wb) or (wb, wc) or (wa, wc))*.

Processing the answer candidates (W_{ans}): Then nodes which are not nouns are removed (like in Section 5.1.1). Then, the proposed system finds the intersection between the processed set of connected nodes, providing answer candidates (W_{ans}). For example, Table 6 shows W_{ans} for functional RAT query *FLU, NAUSEOUS, VIRUS*, which does not have any three-word intersection.

w_a, w_b, and w_c	w_{ans}
flu, nauseous, virus	-
flu, nauseous	sick
flu, virus	disease, influenza, person
nauseous, virus	virose

Table 6 Example of function RAT query with two-word intersection

Calculating cosine similarity score between the answer candidates and choosing an *answer*: As mentioned in section 5.1.1., to choose between the various *Wans,* the average of cosine similarities between the functional RAT queries and the *Wans* is calculated using GloVe embedding. Then the model chooses the highest similarity score as the final *answer*. In the above example, a similarity score is calculated between the various *Wans: SICK, DISEASE, INFLUENZA, PERSON, VIROSE,* and the initial functional RAT query (*wa, wb,* and *wc*): *FLU, NAUSEOUS, VIRUS*. Table 7 shows the cosine similarities scores and the final answer (*answer* column name as in Table 7) that the proposed model chooses.

For example, with the functional RAT query: *FLU, NAUSEOUS, VIRUS*

connected (w_a, w_b) = *sick*

connected (w_b, w_c) = *virose*

connected (w_c, w_a) = *disease*, *influenza*, *person*

Wans = *sick*, *virose*, *disease*, *influenza*, *person*

Now, the similarity score is calculated between all the Wans and the functional RAT query *(wa, wb,* and *wc*) as in Table 7

w_a, w_b, and w_c	w_{ans}	similarity score	*answer*
flu, nauseous, virus			influenza
flu, nauseous	sick	flu - sick: 0.507 nauseous - sick: 0.489 virus - sick: 0.317	
flu, virus	disease influenza person	flu - disease: 0.568 virus - disease: 0.561 nauseous – disease: 0.137 flu - influenza: 0.824 virus - influenza: 0.669 nauseous – influenza: 0.109 flu - person: 0.214	

		virus - person: 0.209 nauseous – person: 0.201	
nauseous, virus	virose	nauseous - virose: no embedding for virose virus - virose: no embedding for virose flu – virose: no embedding for virose	

Table 7 Cosine similarity score and top embedding for two-word intersection

As it has been pointed out in Table 7, similarity score between *virus* and *sick* is calculated even though *sick* was not present in intersection between the functional RAT nodes: *flu*, *nauseous*. Similarly, the same process is followed for other *Wans.* Then the highest average of the similarity score is considered as *answer.*

That is,

(similarity(*flu*,*sick*) + similarity(*nauseous*,*sick*)+similarity(*virus*,*sick*)) / 3

In this example cosine similarity of influenza is greater than the cosine similarity score of *sick*, *disease*, *person*, *virose* and ***influenza*** is considered as the ***answer***

Experimental Results

This approach answered 25 functional RAT queries from the 48 normative functional RAT queries giving an accuracy of 62.5%. Nonetheless, this approach performed much better than the research methodology proposed in Section 5.1.1. Table 8 shows a comparative result on how the functional RAT query performed concerning three- and two-word convergence. As we can see, the research approach proposed in Section 5.1.2. performed much better, yet even with a two-word intersection, some of the functional RAT queries like *DICTIONARY, VERSE, VOCABULARY* is unable to provide any answer candidate that matches the ground solution and to try solving this limitation another research approach is proposed in Section 5.1.3.

w_a, w_b, and w_c	***answer* of three-word**	***answer* of two-word**	**ground solution**
question, reply, solution	answer	answer	answer
bud, dandelion, petals	flower	flower	flower
discuss, gossip, telephone	talk	talk	talk
bench, sofa, stool	chair	chair	chair
flu, nauseous, virus	-	influenza	sick
sensitive, sob, weep	-	cry	cry
crown, royalty, throne	-	king	king
fault, incorrect, justice	-	-	wrong
dictionary, verse, vocabulary	-	-	words
marsh, saliva, slippery	-	-	wet

Table 8 Experimental Results (Approach 2: functional RAT)

Like in section 5.1.1, when evaluated the W_{ans} (the set of answer candidates formed with two and three-word intersection) with the normative data, the proposed model provided with the same accuracy of 62.5% meaning all the W_{ans} that the model retrieved as answer candidate were correct.

5.1.3 Approach 3 – Depth 2

As seen in Sections 5.1.1 and 5.1.2, sometimes three- or two-word intersection does not provide an answer candidate (For example, from the last three queries from Table 8) or the correct ground solution (For example, for the query *FLU, NAUSEOUS, VIRUS* from Table 8). Thus, a third approach is proposed and evaluated where the model is designed to find nodes with a path of length two from the words of functional RAT query (w_a, w_b, and w_c).

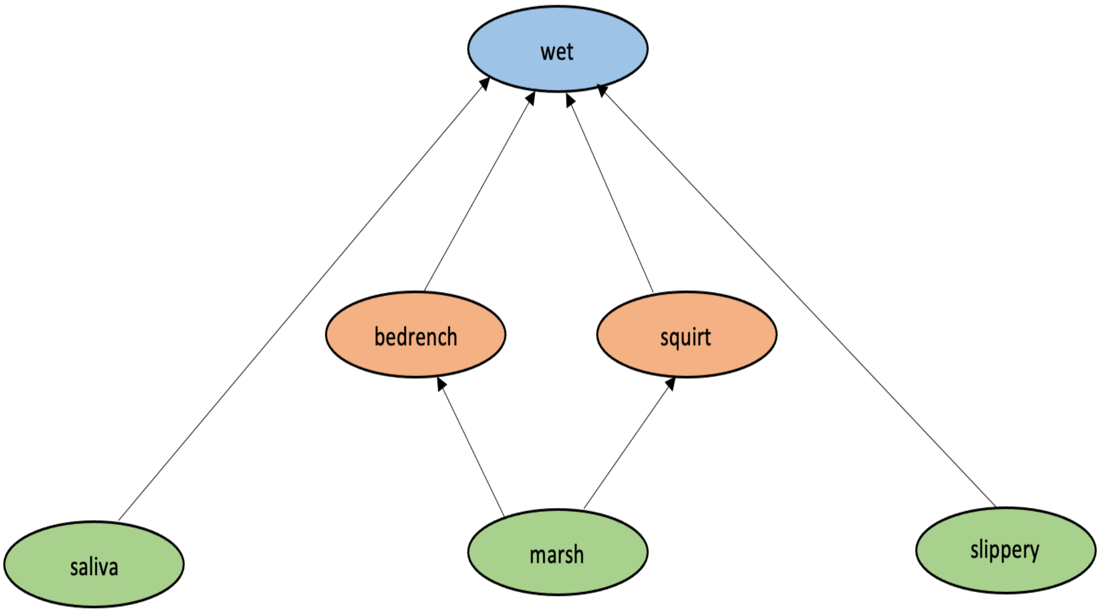

Figure 10 Representation of path length 2

From Figure 10, the green nodes *SALIVA, MARSH, SLIPPERY* is the initial functional RAT query, and the answer candidate *WET* (represented in blue) is connected to the initial two words *SALIVA* and *SLIPPERY* with path length one and *MARSH* are connected to nodes *BEDRENCH* and *SQUIRT* (represented in orange) which are then connected to *WET* (the answer candidate) with path length two or depth two. Only a few nodes are represented for the purpose of visualization.

Constructing a set of connected nodes: The functional RAT query (w_a, w_b, and w_c) is looked upon ConceptNet, and this proposed approach finds all corresponding nodes N for w_a, w_b, and w_c using ConceptNet.

Constructing a set of connected nodes with path length two: Then the proposed approach looks for each node N and determine the set of all nodes N' in ConceptNet which are connected to one of the nodes in N with a maximum of path of length 2.

Determining an intersection between the connected set of nodes: The system finds the intersections from all these retrieved nodes N and N' which has maximum of path of length 2. However, this results in thousands of possible W_{ans} due to large amount of N (nodes that are directly connected to functional RAT query - w_a, w_b, and w_c) and N' (nodes that are connected from N with path length 2). Hence, this research method is used only with three-word intersection (like in Section 5.1.1), the reason being with two-word intersection (as in Section 5.1.2), there will definitely be a solution as a single word from a functional RAT query like *QUESTION, REPLY* and *SOLUTION* has 1061 nodes connected to it with a path of length one, and with the path of length two, there are 89582 nodes.

Pre-processing the connected nodes: From the retrieved nodes only, nouns were kept, and this was done using WordNet

Calculating cosine similarity score between the answer candidates: To narrow down the thousands of nodes which has a path length 2 from N that the proposed system provides as W_{ans} from the above step, the average similarity score is calculated between the W_{ans} and w_a, w_b, and w_c, as in Sections 5.1.1 and 5.1.2. Then these similarity scores are arranged in descending order (greatest similarity score is more similar) and top 10, top 5, and top 3 W_{ans} are retrieved. Table 9 lists a few of the examples which are solved using Path of Length two. The column name 'answer' is the highest similar word for the functional RAT query which is final answer that the proposed model produce.

w_a, w_b, and w_c	**Top 10 w_{ans}**	**Top 5 w_{ans}**	**Top 3 w_{ans}**	**answer**	**Ground truth**
question, reply, solution	answer, statement	answer, statement	answer, statement	**answer**	answer
bud, dandelion, petals	flower, plant	flower, plant	flower, plant	**flower**	flower

discuss, gossip, telephone	talk	talk	talk	**talk**	talk
bench, sofa, stool	chair, furniture, seat	chair, furniture, seat	chair, furniture, seat	**chair**	chair
flu, nauseous, virus	fever, nauseous, sickness, nausea, illness, disease, cholera	fever, nauseous, sickness, nausea, illness	fever, nauseous, sickness	fever	sick
sensitive, sob, weep	cry, moan, sorrow, scream, sadness, sad, despair, howl, smile, regret	cry, moan, sorrow, scream, sadness	cry, moan, sorrow	**cry**	cry
crown, royalty, throne	-	-	-	-	king
fault, incorrect, justice	unfair, inaccurate, improper, mistake, flawed, faulty, correct, wrong	unfair, inaccurate, improper, mistake, flawed	unfair, inaccurate, improper,	unfair	wrong
dictionary, verse, vocabulary	word, grammar, language, thesaurus, phrase, words, translation, meaning, idiom	word, grammar, language, thesaurus, phrases	word, grammar, language	**word**	word
marsh, saliva, slippery	wet, mud, slimy, puddle, watery, grass, dry, ooze, tongue, viscous	wet, mud, slimy, puddle, watery	wet, mud, slimy	**wet**	wet

Table 9 Experimental Results (Approach 3: functional RAT)

Experimental Results

The above proposed method is implemented and evaluated with the normative data from Olteteanu, Schöttner and Schuberth (2019). This research methodology provided with a better result than the research approaches used in Section 5.1.1. and Section 5.1.2. Table 10 shows the accuracy of the proposed approach w.r.t Top 10, Top 5, Top 3 and Top 1; if the answer candidate which are present in Top 10 matches the ground solution, then it is considered as correct, similarly for Top 5, Top 3 and Top 1. The reason for a better accuracy score can be because if a single word from the initial test query has an average of 350 nodes connected to it with a path of length one, then this same node can have an average of 29000 nodes with a path of length two.

Table 10 Accuracy Score for Functional RAT of Path length 2

Accuracy	Top 10	Top 5	Top 3	Top 1
Functional RAT for path length 2	70.8%	64.6%	54.2%	47.9%

Evaluating the answer candidates W_{ans}

When evaluated the W_{ans} (the set of answer candidate that the proposed model provides after finding the intersection) that the proposed approach provided, the accuracy was 83.33%, meaning 40 out of 48 queries had the *ground solution* in the set of answer candidates W_{ans}. For example, the query, *ALGEBRA*, *CALCULUS*, *TRIGONOMETRY* provided W_{ans} as math,

5.1.4 Approach 4 – Solving Functional RAT using Word Embedding

The fourth research method which solves functional RAT is proposed by using word embeddings without ConceptNet. Two word embedding models are used in the below approach.

5.1.4.1 Global Vectors for Word Representations – GloVe word embedding

Initially, in Sections 5.1.1, 5.1.2, and 5.1.3, GloVe embedding was used to calculate the similarity scores for the functional RAT query (w_a, w_b, and w_c) and the answer candidates (W_{ans}). In this proposed approach, GloVe was used to retrieve about thousand nearest neighbours for the words in functional RAT query. The following steps are used to design the proposed approach.

Constructing a set of connected nodes: To solve the functional RAT queries using GloVe embedding, the word vectors w_a, w_b, and w_c, are looked up in GloVe and the nodes (the nearest neighbours) that are connected to these words vectors are retrieved.

Setting a threshold: Since there can be thousands of nearest neighbours, a threshold of retrieving only top 1000 nearest neighbours are set (the nearest neighbours are ordered decreasingly based on the highest to lower similarity score).

Pre-processing the connected nodes: Then, all the retrieved nearest neighbours for the word vectors w_a, w_b, and w_c are expunged from stop words, adjectives, or verbs using WordNet.

Determining the answer candidates: After removing the nodes from the previous steps which are not noun, an intersection is done to determine the answer candidates (W_{ans}). These nearest neighbours are ordered decreasingly based on the highest to lower similarity score. Few examples are listed in Table 11. The column name 'answer' is the most similar word to the final RAT query.

Table 11 Experimental Results: GloVe (Approach 4: functional RAT)

w_a, w_b, **and** w_c	**Top 10**	**Top 5**	**Top 3**	**answer**	**ground truth**
question, reply, solution	**answer**, suggestion, problem, explanation, idea, reason, comment, possible, wrong, why	**answer**, suggestion, problem, explanation, idea	**answer**, suggestion, problem	**answer**	answer
bud, dandelion, petals	leaf, **flower**, root, bloom, chrysanthemum, peony, hibiscus, lily, lavender, clover	leaf, **flower**, root, bloom, chrysanthemum	leaf, **flower**, root	leaf	flower
discuss, gossip, telephone	**talk**, conversation, chat, call, internet, interview, telling, advice, news, information	**talk**, conversation, chat, call, internet	**talk**, conversation, chat	**talk**	talk
bench, sofa, stool	**chair**, couch, recliner, furniture, chaise, ottoman, armchair, settee, desk, sit	**chair**, couch, recliner, furniture, chaise	**chair**, couch, recliner	**chair**	chair
flu, nauseous, virus	influenza, infection, swine, fever, outbreak, diarrhea, pox, cough, sickness, measles	influenza, infection, swine, fever, outbreak	influenza, infection, swine	influenza	sick
sensitive, sob, weep	**cry**, sigh, groan, sad, gasp, pity, sorrow, tremble, anguish, grief	**cry**, sigh, groan, sad, gasp	**cry**, sigh, groan	**cry**	cry
crown, royalty, throne	kingship, scepter, monarchy, empress, highness,	kingship, scepter, monarchy, empress, highness,	kingship, scepter, monarchy	kingship	king

	royalty, nobility, lordship, rulership, peerage				
fault, incorrect, justice	**wrong**, mistake, assumption, contrary, disregard, invalid, unfortunate, failing, assertion, blame	**wrong**, mistake, assumption, contrary, disregard	**wrong**, mistake, assumption	**wrong**	wrong
dictionary, verse, vocabulary	**word**, grammar, language, thesaurus, phrase, pronunciation, words, translation, spelling, meaning	**word**, grammar, language, thesaurus, phrase	**word**, grammar, language	**word**	word
marsh, saliva, slippery	mud, swamp, drip, puddle, damp, mucus, wetness, mouth, gravel, sand	mud, swamp, drip, puddle, damp	mud, swamp, drip	mud	wet

Experimental Results

The above-proposed model is implemented and evaluated with the normative data from Olteteanu, Schöttner and Schuberth (2019). At first, a threshold of similarity score of 0.5 is considered, and only 5 out of 48 queries were answered, out of which only four answers matched the ground solution. So, the threshold was set to 0.25, where the proposed approach provided a better result. Experimental results for few functional RAT queries are illustrated in Table 11, and Accuracy scores are presented in Table 12.

Table 12 Accuracy Score for solving functional RAT using GloVe

Accuracy	Top 10	Top 5	Top 3	Top 1
Functional RAT using Gensim	58.30%	54.2%	50%	42.6%

5.1.4.2 Gensim

Gensim is used for representing documents as semantic vectors. It is a free and open-source python library that is used for Natural Language processing.

Creating a set of connected nodes: From Gensim, a method *Word2Vec.most_similar* is used to calculate the nearest neighbours. This method has three parameters that are used:

- positive = []
- negative = []
- restrict_vocab = None

The positive parameter supports in providing positive words based on similarity and negative parameter provides negative words based on similarity. The restrict_vocab limits the range of vector.

The functional RAT test query (w_a, w_b, and w_c) are passed as positive parameter and, the negative parameter is kept empty. The restrict_vocab is set to 1000 (meaning the first 1000-word vectors in the vocabulary list arranged decreasingly is retrieved).

Pre-processing the answer candidates: These retrieved data are then checked against WordNet, and only nouns are kept.

Determining the answer candidates: Likewise, to Section 5.1.3 and 5.1.4.1 (Using GloVe word embedding), top 10 nodes, top 5 and top 3 are retrieved. Table 13 shows the result obtained using Gensim.

Table 13 Experimental Results: Gensim (Approach 4: functional RAT)

w_a, w_b, and w_c	Top 10 w_{ans}	Top 5 w_{ans}	Top 3 w_{ans}	answer	Ground truth
question, reply, solution	**answer**, response, liquid_bleach, bleach_liquor, gram's_solution, leading_question, question_of_law, rejoinder, spirits_of_ammonia, evasive_answer	**answer**, response, liquid_bleach, bleach_liquor, gram's_solution	**answer**, response, liquid_bleach	**answer**	answer

bud, dandelion, petals	floral_leaf, **flower**, flower_bud, umbrellawort, common_dandelion, easter_daisy, ray_flower, dandelion_green, tidytips, petal	floral_leaf, **flower**, flower_bud, umbrellawort, common_dandelion	floral_leaf, **flower**, flower_bud	floral_leave	flower
discuss, gossip, telephone	**talk**, chitchat, chat, telephone_conversation, gossiping, table_talk, shmooze, shop_talk, scandalmonger, conversation	**talk**, chitchat, chat, telephone_conversation, gossip	**talk**, chitchat, chat	**talk**	talk
bench, sofa, stool	music_stool, flat_bench, chaise_longue, settee, morris_chair, couch, banquette, footstool, recliner, campstool	music_stool, flat_bench, chaise_longue, settee, morris_chair	music_stool, flat_bench, chaise_longue	music_stool	chair
flu, nauseous, virus	influenza, contagious_disease, tumor_virus, upper_respiratory_infection, viral_infection, slow_virus, swine_influenza, respiratory_syncytial_virus, communicable_disease, asian_influenza	influenza, contagious_disease, tumor_virus, upper_respiratory_infection, viral_infection	influenza, contagious_disease, tumor_virus	influenza	sick
sensitive, sob, weep	**cry**, weep, snivel, blue_murder, whimper, wailing, blubberer, lament, tears, bawler	**cry**, weep, snivel, blue_murder, whimper	**cry**, weep, snivel	**cry**	cry
crown, royalty, throne	-	-	-	-	king

fault, incorrect, justice	**wrong,** error, mistake, erroneousness, unfairness, incorrectness, injustice, wrongness, misinterpretation, wrongdoing	**wrong,** error, mistake, erroneousness, unfairness	**wrong,** error, mistake	**wrong**	wrong
dictionary, verse, vocabulary	wordbook, pocket_dictionary, desk_dictionary, etymological_dictionary, bilingual_dictionary, thesaurus, learner's_dictionary, internal_rhyme, lexis, eye_rhyme	wordbook, pocket_dictionary, desk_dictionary, etymological_dictionary, bilingual_dictionary	wordbook, pocket_dictionary, desk_dictionary	wordbook	word
marsh, saliva, slippery	tobacco_juice, swamp, gleet, mud, mucus, wetland, bog, slick, spit, salt_marsh	tobacco_juice, swamp, gleet, mud, mucus	tobacco_juice, swamp, gleet	tobocco_juice	wet

Experimental Results

The above proposed approach is implemented and evaluated with the normative data from Olteteanu, Schöttner and Schuberth (2019). Usually, Gensim did not provide as good result like GloVe.

Table 14 Accuracy score for calculating functional RAT using Gensim

Accuracy	Top 10	Top 5	Top 3	Top 1
Functional RAT with Gensim	39.6%	31.2%	29.2%	18.8%

Comparing between GloVe and Gensim

When analysed the answer candidates (W_{ans}) before data pre-processing Gensim provided with more plurals, adjectives, and verbs than GloVe word embedding. Moreover, when comparing the accuracy GloVe performed much better than Gensim. Out of 48 functional RAT query, Gensim provided no answer candidate for 2 queries:

- *CROWN, ROYALTY* and *THRONE*
- *ADULTS, DEVELOPMENT* and *YO-YO*

However, GloVe was not able to answer only the later query, this is because GloVe did not find the word Yo-Yo. Gensim provided a creative answer (*KINGSHIP* - top 1) than the ground solution (*KING)* for the functional RAT query *CROWN, ROYALTY* and *THRONE.* It can also be a plausible solution since *KINGSHIP* refers to a 'position, office, or dignity of a king[21]' and the query words refers to the title or positions.

[21] https://www.merriam-webster.com/dictionary/kingship

5.1.5 Conclusion

Sections 5.1, 5.2, 5.3 and 5.4, explored the various research methodologies used to solve functional RAT queries. By analysing the various research approaches that were adopted, knowledge acquisition done using ConceptNet with 2-word intersection performed much better than the knowledge acquisition done using Word Embedding and depth 2. Table 15 shows a quick glimpse of various accuracy scores achieved using these different research methodologies. The answer candidates that were provided by the designed methodology is evaluated with the 48 normative functional RAT queries of Olteteanu, Schöttner and Schuberth (2019). From Table 15, it can be deduced that Path of length 2 from Section 5.1.3 performs better than other approach.

<table>
<tr><td colspan="2">3-word intersection</td><td>25%</td></tr>
<tr><td colspan="2">2-word intersection</td><td>62.5%</td></tr>
<tr><td rowspan="4">Depth 2</td><td>Top 10</td><td>70.8%</td></tr>
<tr><td>Top 5</td><td>64.6%</td></tr>
<tr><td>Top 3</td><td>54.2%</td></tr>
<tr><td>Top 1</td><td>47.9%</td></tr>
<tr><td rowspan="4">GloVe</td><td>Top 10</td><td>58.30%</td></tr>
<tr><td>Top 3</td><td>54.2%</td></tr>
<tr><td>Top 5</td><td>50%</td></tr>
<tr><td>Top 1</td><td>42.6%</td></tr>
<tr><td rowspan="4">Gensim</td><td>Top 10</td><td>39.6%</td></tr>
<tr><td>Top 5</td><td>31.2%</td></tr>
<tr><td>Top 3</td><td>29.2%</td></tr>
<tr><td>Top 1</td><td>18.8%</td></tr>
</table>

Table 15 Comparison of Accuracy Scores

Table 16 illustrates a comparison of output using various approach. Bolded words in the table represents that the answer matches the ground solution of the normative data. When analysed the results in depth, there were cases when the words were synonyms, yet it was considered as FALSE (not matching the ground solution). A brief explanation is provided in Chapter 6. Appendix A shows the results with the 48 normative data.

Table 16 Comparison of answer form various approach

w_a, w_b, and w_c	Ground truth	3-word intersection	2-word intersection	Depth 2	Glove	Gensim
question, reply, solution	answer	**answer**	**answer**	**answer**	**answer**	**answer**
bud, dandelion, petals	flower	**flower**	**flower**	**flower**	leaf	floral_leave
discuss, gossip, telephone	talk	**talk**	**talk**	**talk**	**talk**	**talk**
bench, sofa, stool	chair	**chair**	**chair**	**chair**	chair	music_stool
flu, nauseous, virus	sick	-	influenza	fever	influenza	influenza
sensitive, sob, weep	cry	-	**cry**	**cry**	**cry**	**cry**
crown, royalty, throne	king	-	**king**	**king**	kingship	-
fault, incorrect, justice	wrong	-	-	unfair	**wrong**	wrong
dictionary, verse, vocabulary	word	-	-	**word**	**word**	word_book
marsh, saliva, slippery	wet	-	-	**wet**	mud	tobacco_juive

5.2 Research Question 2: "Build an AI system which solves structural RAT"

The second research question that this master thesis focuses on is to build a computational solver which solves Structural RAT. Structural RAT is when words occur together, forming compound words. To implement a model which solves structural RAT, the following steps were developed.

5.2.1 Approach 1 – The three-word intersection

Creating a set of connected Nodes: Unlike functional RAT, where all the nodes that the initial test query are connected to are retrieved, the proposed model to solve structural RAT retrieves only compound words for the initial words in RAT query (*wa, wb,* and *wc*).

Determining the answer candidates: After retrieving the compound words for the words in the test query, an intersection operation is performed to find the W_{ans}. Figure 11 shows how this intersection operation is performed, for example, query *COTTAGE, SWISS, CAKE.*

$$(X\_cottage \cup cottage\_X) \cap (X\_swiss \cup swiss\_X) \cap (X\_cake \cup cake\_X)$$

Figure 11 Intersection operation for the test query COTTAGE, SWISS, CAKE

Cleaning the connected nodes: The nodes that are found as answer candidates are expunged from stop words, adjectives, or verbs. After this operation is performed, the answer that the proposed model provides is *Wans.* Table 17 shows some initial results with the three-word intersection approach.

w_a, w_b, and w_c	w_{ans}
cottage, swiss, cake	cheese
cane, daddy, plum	sugar
loser, throat, spot	-
right, cat, carbon	-
water, mine, shaker	-

Table 17 Examples of structural RAT

Experimental Results

When implemented and analysing the research approach, which uses the three-word intersection, unlike functional RAT (Section 5.1.1.), no query had more than one $W_{ans.}$ However, fewer structural RAT query provided with answer candidates that matched the ground solution, this limitation is tried to be solved in Section 5.2.2. Table 18 shows the Experimental results of few structural RAT queries.

w_a, w_b, and w_c	w_{ans}	Ground Solution
cottage, swiss, cake	cheese	cheese
cane, daddy, plum	sugar	sugar
loser, throat, spot	-	sore
right, cat, carbon	-	copy
water, mine, shaker	-	salt

Table 18 Experimental Results (Approach 1: Structural RAT)

These W_{ans} were evaluated with the 144 normative data of Bowden and Jung. The accuracy of the model was just 13.29%. To solve more structural RAT and improve the accuracy, a second approach was proposed and discussed in Section 5.2.2.

5.2.2 Approach 2 – Two-word intersection

Some of the structural RAT queries solved by the human participants in the 144 normative data will not be known to a computational solver since different agents can have a different knowledge base. This can be a reason why the approach adopted in Section 5.2.1 did not provide a more satisfying accuracy even though structural RATs are compound words. Hence, when the proposed model does not find a two-word intersection, an approach like Section 5.1.2 is implemented when the proposed system looks for a two-word intersection.

Creating a set of connected nodes: The structural RAT query is looked up in ConceptNet. A set of compound words that are connected to the structural RAT query are retrieved.

Determining the answer candidates: In case, the initial test query does not have a three-word intersection, it looks for intersections with all possible combinations as in Section 5.1.2.

Cleaning the answer candidates: After retrieving the intersection as answer candidates, stop words, verbs, adjectives, or pronouns are removed. For example, the query *DUCK, DOLLAR, FOLD* returns *ONE'S* as one of its W_{ans}. In this case, *ONE'S* is removed as it is a pronoun[22]. For example, in some cases, for a query *DREAM, BREAK, LIGHT,* '-ing' was considered a compound word; these were neglected too. Table 19 shows the results after data cleaning.

w_a, w_b, and w_c	w_{ans}
right, cat, carbon	
right, cat	
right, carbon	copy
cat, carbon	

Table 19 Example of structural RAT queries with two word intersection

In most of the cases, the model provided only one W_{ans}. However, fewer queries that had more than one W_{ans} similarity score were calculated as in Section 5.1. Examples of such query is given in Table 20.

[22] https://www.merriam-webster.com/dictionary/ones

w_a, w_b, and w_c	w_{ans}	*answer*
opera, hand, dish	-	towel
opera, hand	-	
opera, dish	copy	
hand, dish	out, tower, side	
cat, carbon, right	-	copy
cat, carbon	black	
cat, right	foot, animal	
carbon, right	copy	

Table 20 Examples of structural RAT queries with top_embedding

Experimental Results:

With the approach followed in Section 5.2.2. the proposed system was able to answer better than the approach followed in Section 5.2.1. After evaluating the column *answer* with the 144-ground solution of Bowden and Jung, the approach presented with an accuracy of 37.25%. Table 21 shows some experimental results using the proposed approach.

w_a, w_b, and w_c	w_{ans}	Ground Solution
cottage, swiss, cake	cheese	cheese
cane, daddy, plum	sugar	sugar
loser, throat, spot	sore	sore
right, cat, carbon	copy	copy
water, mine, shaker	salt	salt
opera, hand, dish	towel	soap

Table 21 Experimental Results (Approach 2: structural RAT)

5.2.3 Conclusion

The approach discussed in Section 5.1.3, where functional RAT was solved using a path of length two, was not employed to solve structural RAT as Structural RATs are compound words and the path of length two is not coherent.

In brief, section 5.2.2, which solved structural RAT using the two-word intersection, performed much better than the approach from section 5.2.1. Sometimes a computational solver cannot present an answer candidate, while a human participant can because of different knowledge organisations. This knowledge organisation can also be a reason why functional RAT performed much better than structural RAT with knowledge acquitted from ConceptNet as ConceptNet represents common-sense knowledge, where according to AI-community, common sense knowledge refers to millions of basic facts. Appendix B shows the result for 144 structural RAT query.

5.3 Research Question 3: “Constructing explanations for functional RAT”

The third part of this master thesis is to construct natural language explanations for the final answer (*answer*) that the designed model delivers for solving functional RAT. Unlike structural RAT, where queries are generally associated together (same syntactic structure) in a language forming compound words with no functional meaning, functional RAT has a functional relationship with the queries rather than a language relationship. Hence explanations for functional RAT can be achieved by understanding the semantic relation between the nodes. To answer the research question for constructing explanations on why an answer candidate is connected to the initial functional RAT test query, ConceptNet, a knowledge base represented as graphs is used to study the “why”.

As illustrated in Figure 12, nodes depict entities like the initial test queries in yellow, and edges illustrates the relationship between these entities like *TypeOf*, *LocatedAt,* etc. For example, if the initial test query is *DAISY, TULIP, VASE,* as shown in Figure 10, the proposed system from Section 5.1. can provide *top_embedding* to be *FLOWER,* and to answer the research question three to explain “why” *FLOWER* is the answer candidate, the proposed system is expected to provide an explanation like: “*Daisy* and *Tulip* are type of flowers, and *Flowers* are located in *Vase*”.

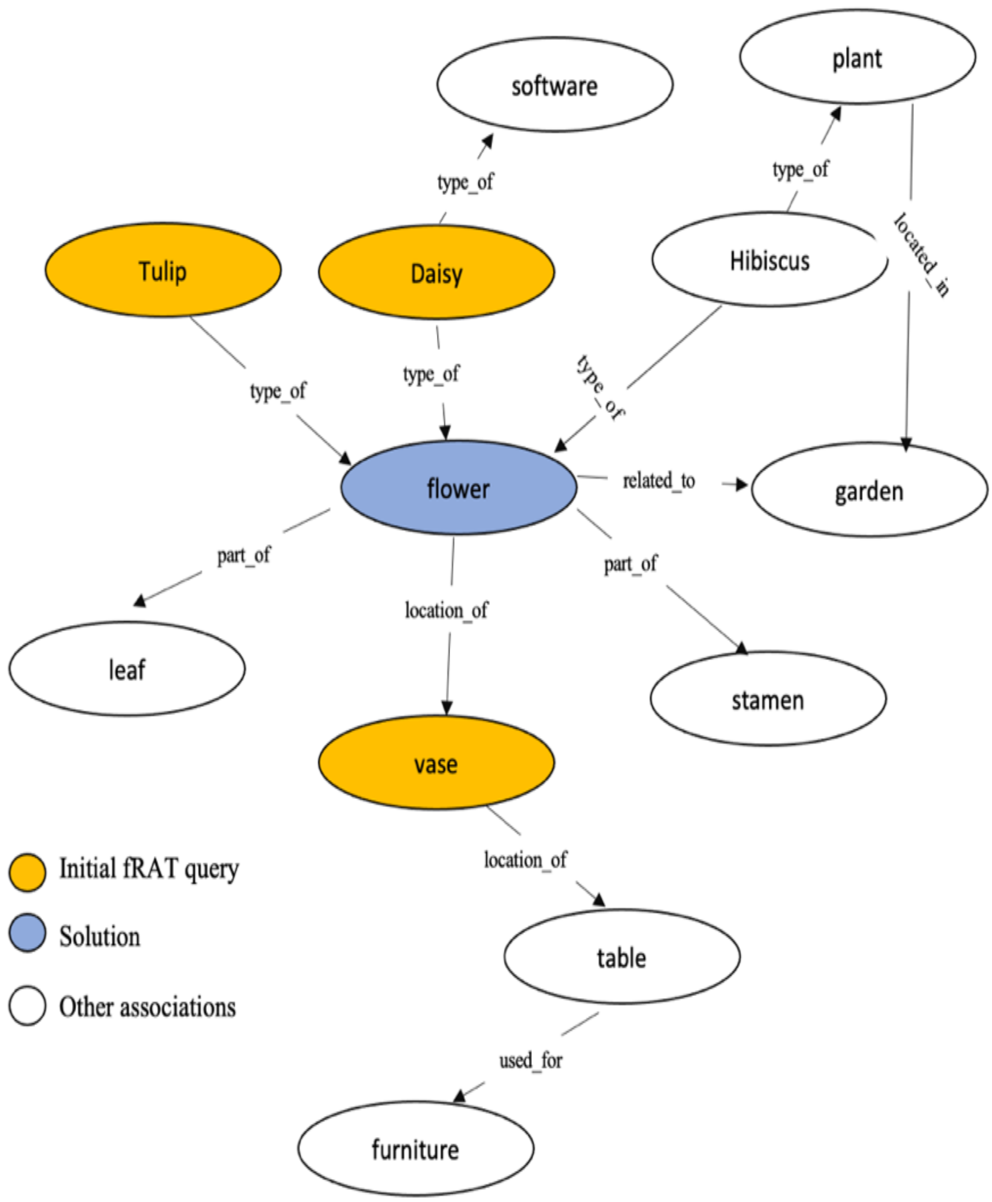

Figure 12 Small snippet of ConceptNet, for the initial query word DAISY, TULIP, VASE

5.3.1 Research Approach

To build a model to answer this research question, below steps are followed:

Retrieving edge names: The relationship between the functional RAT query (w_a, w_b, *and* w_c) and *answer* from Section 5.1.3(since this approach provided better accuracy), is looked up (The direction in which the entities are connected are considered since they are crucial for forming explanations). The experimental results are as in Table 22. However, these explanations that the proposed model provides are not in perfect natural language as they follow a basic template of {0} *edge_names* {1} where {0} and {1} represents the entities or the initial test queries (w_a, w_b, *and* w_c). For example: "solution *synonym* answer", instead a perfect natural language that the proposed model aims to provide is *"solution and answer are synonyms"* another example, *"question antonym answer"* the model can provide "question is the opposite of answer."

Table 22 Initial results for the query question, reply, solution

w_a, w_b, and w_c	Top_embedding	Explanation
question, reply, solution	answer	question "*desires, distinct_from, antonym, related_to*" answer reply "*related_to, is_a, synonym, antonym*" answer solution "*synonym, related_to*" answer

Natural Language explanation: To achieve this natural language explanation, hand crafted template proposed by Feldman, Davison, Rush (2019) is used. Some examples for these hand-crafted templates can be found in Table 23.

Edge_names	Template
related_to	{0} is like {1} / {1} is like {0} {0} is related to {1} / {1} is related to {0)
at_location	You are likely to find {0} in {1}
desires	{0} wants {1} {1} wants {0}
synonyms	{0} and {1} have similar meaning
antonyms	{0} is the opposite of {1}

Table 23 Handcrafted templates for explanations

In case the edge name is *'part_of',* then the proposed model checks for the direction of relationship, that is, whether {0} is connected to {1} or {1} is connected to {0} and choose the order of entity accordingly. So, the initial results from Table 24 will be modified as in Table 25.

w_a, w_b, and w_c	question, reply, solution
answer	answer
Initial Results	question "desires, distinct_from, antonym, related_to" answer reply "related_to, is_a, synonym, antonym" answer solution "synonym, related_to" answer
Results with template	question wants answer question is related to answer reply is an answer | reply has a similar meaning to answer solution has a similar meaning to answer answer is the opposite of reply

Table 24 Initial Results: "Why" and entity is connected to an answer candidate

After obtaining results as in Table 25 with hand crafted templates, then these results are closely examined. To make it better, if an entity has more than one relationship, then they are combined. For example, from Table 25, “reply is an answer | reply has a similar meaning to answer” is combined to “reply is an answer and reply has a similar meaning to answer”. Example of an updated explanation is listed in Table 25.

Table 25 "Why" an answer candidate is related to Functional RAT

w_a, w_b, and w_c	question, reply, solution
answer	answer
Initial Results	question "desires, distinct_from, antonym, related_to" answer reply "related_to, is_a, synonym, antonym" answer solution "synonym, related_to" answer
Results with template	question is related to answer question wants answer reply is an answer reply has a similar meaning to answer solution has a similar meaning to answer answer is the opposite of reply
Explanations	question is related to answer and question wants answer reply is an answer and reply has a similar meaning to answer and answer is opposite to reply. solution has a similar meaning to answer

Explanation for structural RAT

The focus of this master thesis is to construct explanations for functional RAT queries, however when tried to construct explanations for structural RAT which are compound words most nodes were connected to each other with the edge names 'is_a', 'related_to' or 'derived_from'. These structural RAT queries also do not require many explanations since they have only language association between them.

Evaluation of the explanations

These explanations can be provided to human participants to evaluate the feasibility of the approach.

6 Limitations

6.1 Research Question 1: “Build a system which solves functional RAT”.

While examining the research approach proposed in Sections 5.1 (Using three-word intersection) and 5.2 (Using two-word intersection), there were some plausible answer candidates that the proposed system provided. Sometimes these plausible answers can be considered even more “creative” or unpredictable from a human perspective. One such example is *BENCH, SOFA, STOOL*; the model provided *FURNITURE* as one of its $W_{ans.}$

Further studying the answer candidates provided with the approach proposed in Section 5.3 using the path of length 2, some outputs were synonyms of the ground solution, yet the model considered that it does not match the ground solution. These plausible synonyms can result due to data regularities. Examples of such queries are *FLU, NAUSEOUS, VIRUS* provided with the following W_{ans}: *SICKNESS, NAUSEOUS, NAUSEA, ILLNESS, DISEASE, CHOLERA* as answer candidates; and the ground solution *SICK* is a synonym of S*ICKNESS*[23]. Similarly, when the ground solution is *MATH,* but if the proposed model provides *MATHEMATICS* as its answer candidate, the designed model still considers the answer candidate as wrong.

Similarly, when investigating the results from the approach used using Word Embedding in Section 5.4, some answer candidates that the model retrieves as final output are also synonyms to the ground solution. For example, for the query *ARREST, BADGE, DEPUTY,* the model provides *POLICEMAN* or *OFFICER* as answer candidates, and the ground solution is *COP. COP* and *POLICEMAN* are synonyms[24]. Some other examples of this case are listed in the below Table 26.

Table 26 Examples of few plausible solutions

w_a, w_b, and w_c	answer candidate	ground solution	Plausibility
arrest, badge, deputy	deputy	cop	Synonym
exam, scare, terror	anxiety	fear	Synonym
fierce, steel, warrior	tough	strong	Synonym

[23] https://www.thesaurus.com/browse/sick
[24] https://www.thesaurus.com/browse/cop

Considering Table 16 Comparison of answer form various approach, which compares the answers provided by the proposed system upon using various research methodology, it is easy to see that the some of the answers for functional RAT query can be considered plausible like for query:

- FLU, NAUSEOUS, VIRUS – research approach from Section 5.1.3 provides *FEVER* as it's answer which can be plausible to the ground solution *SICK*.
- FAULT, INCORRECT, JUSTICE – research approach from Section 5.1.3 provides *UNFAIR* as it's answer, which can be plausible to the ground solution *WRONG*.
- CROWN, ROYALTY, THRONE – research approach from Section 5.1.4 (GloVe) provides *KINGSHIP* as it's most similar nodes which can be considered as plausible to the ground solution *KING*.

Cases like BANDAID -BANDAGE or DEAD – DEATH were also considered false. One of the functional RAT queries with the word 'yo-yo' was not present in ConceptNet hence for this query, the proposed approach did not provide any intersection or any similar node to this word.

In some cases, words like *MATHEMATICS* or *WORDS* were provided as answer candidates and the ground solution were *MATH* or *WORD*. These limitations are solved by looking for parts of words in the complete word. But there are some case which this fails like when the word is *DESCRY* and the system considers as *CRY*.

These limitations can be solved by looking up synonyms for the answer candidates and matching with the ground solutions using WordNet.

6.2 Research Question 2: "Build a system which solves structural RAT".

As proposed in Section 5.2, Structural RAT did not perform better in ConceptNet. However, WordNet could be used to solve these limitations and WordNet is common-sense knowledge base which is a collection of 200 000 distinct words of primary nouns, verbs, and adjectives and used by computational linguistic community. As also seen in Table 13 , using WordNet provided more compound words than other word embeddings like GloVe. However, the best way approach to solve structural RAT would be to use bi-gram model as in bi-gram model the words occur together.

While solving structural RAT queries the runtime was more compared to solving functional RAT queries. Figure 13 shows the snippet of the run-time, and 144 structural RAT query took 22 hours to execute using Approach 1.

```
144it [22:20:30, 558.54s/it]
```

Figure 13 Runtime for Structural RAT

6.3 Research Question 3: "Constructing explanations for functional RAT"

In some cases, these handcrafted explanations are irrational. Implementing these hand-crafted templates, the relationship type *'related_to'* is replaced with *{0} is like {1}* or *{1} is like {0}*, for example, *"twinkle is like stars"*, *"moon is like star"*, *"fish is like pond"*, or *"question is like answer"* however if these explanations are replaced with *{0} is related to {1}* or *{1} is related to {0}* is logical that *is "twinkle is related to star"*, *"moon is related to stars"*, *"fish is related to pond"* or *"question is related to answer"*. Nevertheless, there are some cases like *"brawl is like fight"* where using *{0} is like {1}* sounds reasonable.

Table 25 "Why" an answer candidate is related to Functional RAT, also shows an unreasonable explanation "answer is opposite to reply". Other unreasonable explanation that the model retrieved was *"you are likely to find figure in hand"*. This is because ConceptNet provided with a relationship between answer and reply as antonyms of. A reason for this absurd explanation can be because ConceptNet knowledge mainly from crowdsourced resources. Other ontologies like DBpedia can be used to get better explanations.

7 Conclusion and Future Works

Remote Association Test is an empirical test to measure the associative ability of a human participant and depending upon this associative ability creativity is measured in humans. In this master thesis, Remote Association Test was tried to be solved computationally using knowledge from common-sense knowledge bases like ConceptNet and word embeddings like GloVe and Gensim. In this master thesis three research questions were answered: ‘To solve functional RAT’, ‘To solve structural RAT’ and ‘To explain why an answer candidate is selected’.

In brief, to answer the first research question, four approaches were proposed, implemented, and evaluated with the 48 normative data from Olteţeanu, Schöttner and Schuberth (2019). The first approach, which used 'three-word intersection', performed with an accuracy score of 25%. This was improved with the second approach on using 'two-word intersection, which produced an accuracy of 62.5%. A third approach which looked for 'path of length 2', provided an accuracy of 70.8% when retrieved the top 10 nodes most similar. The fourth approach that was proposed used word embedding for knowledge acquisition and provided an accuracy of 58.30% while using GloVe and 39.6% while using Gensim. To answer the research question, using ConceptNet for knowledge acquisition provided better results than Word embeddings and approach that used two-word intersection provided with better accuracy.

To sum up the second research question, two approaches were proposed, implemented, and evaluated with the 144 normative data of Bowden and Jung (2003). The first approach, ‘three-word intersection,’ answered few queries and provided an accuracy of 13.29% and the second approach, ‘two-word intersection,’ provided an accuracy of 37.25%. Both these approaches did not perform well as compared to functional RAT queries. The reason can be because ConceptNet is a common-sense knowledge base, and structural RAT is a collection of compound words. This limitation can be solved by using WordNet for knowledge acquisition.

Finally, the third research question was answered by looking upon the nodes (answer candidate and the RAT query) and edge connections in ConceptNet. A method proposed by Feldman, Davison, Rush (2019) was used to provide a natural language explanation. These explanations can be evaluated using human participants as future works. This study can also be implemented with other ontologies like DBpedia.

As future work on this research regarding computational creativity,

- Building a computational solver for structural RAT using WordNet.
- Explanations on "why" an answer candidate is related to can be produced using other ontologies
- More functional RAT can be proposed by looking up which node (W_{ans}) in ConceptNet has three w_a, w_b, *and* w_c

8 Reference

Akbari Chermahini, S., Hickendorff, M. and Hommel, B. (2012) ‘Development and validity of a Dutch version of the Remote Associates Task: An item-response theory approach’, *Thinking Skills and Creativity*. doi: 10.1016/j.tsc.2012.02.003.

Bowden, E. M. and Jung-Beeman, M. (2003) ‘Normative data for 144 compound remorte’, *Behavioral Research Methods, instruments, and Computers*.

Colton, S. (2012a) ‘Automated theory formation in pure mathematics.’, in *Springer Science & Business Media*.

Colton, S. (2012b) ‘The painting fool: Stories from building an automated painter. In Computers and creativity’, *Springer*, pp. 3–38.

Colton, S., Charnley, J. and Alison Pease (2012) ‘Computational Creativity Theory: The FACE and IDEA Descriptive Models’, in *Proceedings of the Second International Conference on Computational Creativity*.

Colton, S., Goodwin, J. and Veale, T. (2012) ‘Full-FACE poetry generation’, in *In Proceedings of the Third International Conference on Computational Creativity*, pp. 95–102.

Cook, M. and Colton, S. (2014) ‘Ludus ex machina: Building a 3D game designer that competes alongside humans’, in *Proceedings of the 5th International Conference on Computational Creativity.*

Davies, L. M. and Lynn Newton (2018) *Creativity is a human quality that exists in every single one of us*. Available at: https://theconversation.com/creativity-is-a-human-quality-that-exists-in-every-single-one-of-us-92053.

Dippo, C. (2013) ‘Evaluating The Alternative Uses Test of Creativity’, in *Proceedings of the National Conference, On Undergraduate Research (NCUR) 2013 University of Wisconsin La Crosse, W*.

Dow, Gayle T. and Mayer, R. E. (2004) ‘Teaching students to solve insight problems: Evidence for domain specificity in creativity training’, *Creativity Research Journal*. doi: 10.1080/10400410409534550.

Dow, G. T and Mayer, R. E. (2004) ‘Teaching students to solve insight problems: Evidence for domain specificity in creativity training’, *Creativity Research Journal*, pp. 389–398.

Duncker, K. (1945) ‘On Problem Solving: Psychological Monographs’, *American Psychological Assiciation Psychological Bulletin*.

Ebersold, S. (2008) ‘Organization Workshop organizers’, (June 2015).

Guilford, J. P. (1967) ‘The Nature of Human Intelligence. McGraw-Hill, New York, 1967. xiv

+ 538 pp., illus.', *Science*. doi: 10.1126/science.162.3857.990-a.

Hamilton, M. A. (1982) '"Jamaicanizing" the Mednick Remote Associates Test of Creativity', *Perceptual and Motor Skills*. doi: 10.2466/pms.1982.55.1.321.

Jiménez, G., Fernández-cosials, K. and Mínguez, E. (2017) 'Could Creativity Be Taught and Evaluated in a Nuclear Could Creativity Be Taught and Evaluated in a Nuclear Engineering Course ?', *NESTet2016*, (May).

Kilgarriff, A. and Fellbaum, C. (2000) 'WordNet: An Electronic Lexical Database', *Language*. doi: 10.2307/417141.

Kim, K. H. (2006) 'Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT)', *Creativity Research Journal*. doi: 10.1207/s15326934crj1801_2.

Lenat, D. B. (1976) 'AM: An artificial intelligence approach to discovery in mathematics as heuristic search', *DTIC Document*.

Liu, H. and Singh, P. (2004) 'ConceptNet - a practical commonsense reasoning tool-kit', *BT Technology Journal*. doi: 10.1023/B:BTTJ.0000047600.45421.6d.

Maier, N. R. F. (1931) 'Reasoning in humans. II. The solution of a problem and its appearance in consciousness', *Journal of Comparative Psychology*. doi: 10.1037/h0071361.

Mednick, S.A., Mednick, M. (1971) 'Remote Associates Test: Examiner's Manual.', *Houghton Mifflin, Boston*.

Mednick, S. (1962) 'The associative basis of the creative process', *Psychological Review*. doi: 10.1037/h0048850.

Mednick, S. A. (1968) 'Remote Associates Test. J, 2, 213-214.', *Journal of Creative Behavior*, (2), pp. 213–214.

Mednick, S. A. and Mednick, M. P. (1967) 'Examiner's manual: Remote Associates Test.', *Boston: Houghton Mifflin.*

Nelson, D. L., McEvoy, C. L. and Schreiber, T. A. (2004) 'The University of South Florida free association, rhyme, and word fragment norms', *Behavior Research Methods, Instruments, and Computers*. doi: 10.3758/BF03195588.

Nevo, B. and Levin, I. (1978) 'Remote associates test: assessment of creativity in Hebrew.', *Megamot*.

Olteţeanu, A.-M. (2020) *Cognition and the Creative Machine*, *Cognition and the Creative Machine*. doi: 10.1007/978-3-030-30322-8.

Olteteanu, A.-M., Schöttner, M. and Schuberth, S. (2019) 'Computationally resurrecting the functional Remote Associates Test using cognitive word associates and principles from a computational solver', *Knowledge-Based Systems*.

Olteţeanu, A. M. and Falomir, Z. (2015) ‘ComRAT-C: A computational compound Remote Associates Test solver based on language data and its comparison to human performance’, *Pattern Recognition Letters*. doi: 10.1016/j.patrec.2015.05.015.

Olteţeanu, A. M., Gautam, B. and Falomir, Z. (2015) ‘Towards a visual remote associates test and its computational solver’, in *CEUR Workshop Proceedings*.

Oltețeanu, A. M. and Schultheis, H. (2019) ‘What determines creative association? Revealing two factors which separately influence the creative process when solving the Remote Associates Test’, *Journal of Creative Behavior*. doi: 10.1002/jocb.177.

Olteţeanu, A. M., Schultheis, H. and Dyer, J. B. (2018) ‘Computationally constructing a repository of compound remote associates test items in American English with comRAT-G’, *Behavior Research Methods*. doi: 10.3758/s13428-017-0965-8.

Pearce, M. and Wiggins, G. (2004) ‘Improved methods for statistical modelling of monophonic music’, *Journal of New Music Research, 33(4)*.

Pease, A. *et al.* (2001) ‘Evaluating Machine Creativity’, *Workshop on Creative Systems, 4th International Conference on Case Based Reasoning*, (November), pp. 129–137.

Pereira, F. C. *et al.* (2015) ‘Experiments With Assessment of Creative Systems: An Application of Ritchie’s Criteria’, in *Second Computational Creativity Workshop, IJCAI2005*.

Qiu, J. *et al.* (2008) ‘The neural basis of insight problem solving: An event-related potential study.’, *Brain and Cognition, 68(1)*, pp. 100–106.

Schneider, S., Fischer, J. . and König, R. (2011) ‘Rethinking automated layout design: developing a creative evolutionary design method for the layout problems in architecture and urban design’, *In Design Computing and Cognition’10*.

Schooler J.W., M. J. (1995) ‘The ineffability of insight’, in *The Creative Cognition Approach*, pp. 97–133.

Singh, P., Barry, B. and Liu, H. (2004) ‘Teaching machines about everyday life’, *BT Technology Journal*. doi: 10.1023/B:BTTJ.0000047601.53388.74.

Torrance, E. P (1998) ‘Torrance tests of creative thinking: Norms-technical manual: Figural (streamlined) forms a & b. Princeton’, *NJ: Personnel Press.*

Torrance, E.P. (1998) ‘Torrance tests of creative thinking: verbal forms A and B’, *Bensenville, IL: Scholastic Testing Service.*

Torrance, E. P. and Ball, O. (1998) ‘Torrance tests of creative thinking: Norms- technical manual: Figural (streamlined) forms a & b’, *Bensenville, IL: Scholastic Testing Service.*

Whitt, J. K. and Prentice, N. M. (1977) ‘Cognitive processes in the devel- opment of children’s enjoyment and comprehension of joking riddles. Developmental Psychology’, *Developmental*

Psychology, p. 129.

Wiggins, G. A. (2006) 'A preliminary framework for description, analysis and comparison of creative systems', *Knowledge-Based Systems, 19(7)*, pp. 449–458.

Worthen, B. R. and Clark, P. M. (1971) 'TOWARD AN IMPROVED MEASURE OF REMOTE ASSOCIATIONAL ABILITY', *Journal of Educational Measurement*. doi: 10.1111/j.1745-3984.1971.tb00914.x.

Zunjani, F. H. and Olteţeanu, A. M. (2019) 'Cognitive AI Systems Contribute to Improving Creativity Modeling and Measuring Tools', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-030-19591-5_11.

Appendix A

Functional RAT (w_a, w_b, and w_c)	Ground solution	3 intersection	3,2 intersection	Depth 2	GloVe	Gensim
question ,reply , solution	answer	answer	answer	answer	answer	answer
sensitive ,sob, weep	cry		cry	cry	cry	cry
antlers ,doe , fawn	deer	deer	deer	deer	deer	deer
bud ,dandelion , petals	flower	flower	petals	flower	leaf	floral_leaf
colt ,mare , unicorn	horse	horse	stallion	horse	stallion	male_horse
crown ,royaly , throne	king		king		kingship	
algebra ,calculus , trigonometry	math	mathematics	mathematics	algebra	mathematics	pure_mathematics
pedal ,pull , shove	push		push	pull	push	push
clockwise ,left , wrong	right		right	left	right	
flu ,nauseous , virus	sick		influenza	flu	influenza	influenza
astronomy ,moon , twinkle	star	star	sky	star	sky	
bait ,pond , tuna	fish	fish	fish	fish	fish	rough_fish
bandaid ,trim , wound	cut		cut	wound	bandage	raw_wound
gravity ,low , up	down	down	down	down	down	down
emergency ,rapid, slow	fast		fast	rapid	fast	fast
brawl ,debate , soldier	fight	fight	fight	fight	battle	pitched_battle
birds ,frog , kite	fly		bird	bird	turtle	bird
finger ,glove , palm	hand	hand	finger	hand	hand	thumb
bed ,darkness , sedative	sleep		sleep	bed	sleep	sleeping_pill
discuss ,gossip , telephone	talk	talk	talk	talk	talk	talk
fangs ,gums , wolf	teeth		fang	teeth	tooth	fang
marsh ,saliva , slippery	wet			wet	mud	tobacco_juice
dictionary ,verse , vocabulary	words		word	dictionary	word	wordbook
fault ,incorrect , unjust	wrong		wrong	incorrect	wrong	wrong
murder,operate, vein	blood		crime	crime	suspect	stylomastoid_vein
empire,moat, princess	castle		castle	princess	castle	crown_princess
bench,sofa, stool	chair	chair	chair	chair	chair	music_stool
beaker,flask, science	chemistry	laboratory	beakers	beaker	chemistry	chemistry_lab
adults,development, yo-yo	children				learning	
cemetery,coma, noose	dead		human	grave	grave	
exam,scare, terror	fear		fear	scary	fear	reign_of_terror
hand,toe, trigger	finger	finger	finger	finger	finger	finger
angel,church, faith	god	religious	pray	faith	prayer	lay_reader

body,commander, skull	head	head	head		head	death's_head
cello,scalpel, trumpet	instrument		instrument	sax	clarinet	violin
desk,quill, stapler	pen		staples	desk	typewriter	
arrest,badge, deputy	cop	police	officer	deputy	officer	officer
electron,inertia, zest	energy			spin	magnetism	delta_ray
diet,strain, sweat	exercise		stress	sweat	stomach	
assault,cop, murder	gun		crime	murder	crime	homicide
drill,grave, spike	hole		hole	spike	hammer	dentist's_drill
care,tactful, willing	kind			polite	willingness	
midnight,saturn, wolf	moon		fang	moon	moon	
bloom,opportunity, split	open		opening	opportunity	will	chance
accomplished,dolphin, sly	smart			well	cunning	cunning
duck,sardine, sinker	swim		food	duck	tuna	clupeid_fish
europe,mushroom, pack	trip		pick	pack	european	european_country
fierce,steel, warrior	strong		sword	steel	sword	

Appendix B

Structural RAT	wans	3 intersection	3,2 intersection
Cottage,Swiss, Cake	Cheese	cheese	cheese
Cream,Skate, Water	Ice	salt	salt
Loser,Throat, Spot	Sore		sore
Show,Life, Row	Boat		long
Night,Wrist, Stop	Watch		
Duck,Fold, Dollar	Bill		down
Rocking,Wheel, High	Chair		gear
Dew,Comb, Bee	Honey		
Fountain,Baking, Pop	Soda		soda
Preserve,Ranger, Tropical	Forest		
Aid,Rubber, Wagon	Band		wheel
Flake,Mobile, Cone	Snow		
Cracker,Fly, Fighter	Fire		ass
Safety,Cushion, Point	Pin		seat
Cane,Daddy, Plum	Sugar	sugar	sugar
Dream,Break, Light	Day	up	up
Fish,Mine, Rush	Gold	salt	out
Political,Surprise, Line	Party	up	end
Measure,Worm, Video	Tape		
High,District, House	School	field	
Sense,Courtesy, Place	Common		
Worm,Shelf, End	Book	up	up
Piece,Mind, Dating	Game		
Flower,Friend, Scout	Girl	girl	girl
River,Note, Account	Bank	of	take
Print,Berry, Bird	Blue		blue
Pie,Luck, Belly	Pot		pork
Date,Alley, Fold	Blind		up
Opera,Hand, Dish	Soap		out
Cadet,Capsule, Ship	Space		war
Fur,Rack, Tail	Coat		light
Stick,Maker, Point	Match		up
Hound,Pressure, Shot	Blood		
Fox,Man, Peep	Hole		
Sleeping,Bean, Trash	Bag		bags
Dust,Cereal, Fish	Bowl		bowl
Light,Birthday, Stick	Candle	up	up
Food,Forward, Break	Fast	up	out
Shine,Beam, Struck	Moon		light

Peach,Arm, Tar	Pit		
Water,Mine, Shaker	Salt	salt	salt
Palm,Shoe, House	Tree	field	tree
Basket,Eight, Snow	Ball		ball
Wheel,Hand, Shopping	Cart		out
Right,Cat, Carbon	Copy		animal
Home,Sea, Bed	Sick		day
Nuclear,Feud, Album	Family		family
Sandwich,House, Golf	Club	field	club
Cross,Rain, Tie	Bow		down
Sage,Paint, Hair	Brush		brush
French,Car, Shoe	Horn		polish
Boot,summer, Ground	Camp	high	winter
Chamber,Mask, Natural	Gas		face
Mill,Tooth, Dust	Saw		out
Main,Sweeper, Light	Street	up	electric
Pike,Coat, Signal	Turn		turn
Office,Mail, Sand	Box		box
Fly,Clip, Wall	Paper		front
Age,Mile, Sand	Stone		of
Catcher,Food, Hot	Dog		fast
Wagon,Break, Radio	Station	up	up
Tank,Hill, Secret	Top		like
Health,Taker, Less	Care		
Lift,Card, Mask	Face		face
Dress,Dial, Flower	Sun	girl	girl
Force,Line, Mail	Air	up	up
Guy,Rain, Down	Fall		down
Eight,Skate, Stick	Figure		up
Down,Question, Check	Mark		out
Animal,Back, Rat	Pack		
Officer,Cash, Larceny	Petty		petty
Pine,Crab, Sauce	Apple	white	lobster
House,Thumb, Pepper	Green	field	green
Carpet,Alert, Ink	Red		
Master,Toss, Finger	Ring		up
Hammer,Gear, Hunter	Head		
Knife,Light, Pal	Pen	up	up
Foul,Ground, Mate	Play	high	ball
Change,Circuit, Cake	Short	cheese	make
Blank,List, Mate	Check		
Tail,Water, Flood	Gate	salt	light

Way,Board, Sleep	Walk		go
Marshal,Child, Piano	Grand		up
Cover,Arm, Water	Under	salt	heavy
Rain,Test, Stomach	Acid		down
Time,Blown, Nelson	Full	full	full
Oile,Market, Room	Stock		single
Mouse,Bear, Sand	Trap		red
Cat,Number, Phone	Call		animal
Keg,Puff, Room	Powder		powder
Trip,House, Goal	Field	field	own
Fork,Dark, Man	Pitch		in
Fence,Card, Master	Post		security
Test,Runner, Map	Road		to
Dive,Light, Rocket	Sky	up	out
Man,Glue, Star	Super		in
Tooth,Potato, Heart	Sweet		and
Illness,Bus, Computer	Terminal		terminal
Type,Ghost, Screen	Writer		computer
Mail,Board, Lung	Black		box
Teeth,Arrest, Start	FALSE		
Iron,Shovel, Engine	Steam	steam	steam
Rope,Truck, Line	Tow	up	tow
Wet,Law, Business	Suit		practice
Off,Military, First	Base		head
Spoon,Cloth, Card	Table		table
Cut,Cream, War	Cold		up
Note,Chain, Master	Key		take
Shock,Shave, Taste	After		
Wise,Work, Tower	Clock		
Grass,King, Meat	Crab		chicken
Baby,Spring, Cap	Shower		up
Break,Bean, Cake	Coffee	cheese	cheese
Cry,Front, Ship	Battle		war
Hold,Print, Stool	Foot		up
Roll,Bean, Fish	Jelly		sauce
Horse,Human, Drag	Race	race	race
Oil,Bar, Tuna	Salad		fish
Bottom,Curver, Hop	Bell	rock	rock
Tomato,Bomb, Picker	Cherry		
Pea,Shell, Chest	Nut		rice
Line,Fruit, Drunk	Punch	up	up
Bump,Egg, Step	Goose	on	up

Fight,Control, Machine	Gun		
Home,Arm, Room	Rest		single
Child,Scan, Wash	Brain		up
Nose,Stone, Bear	Brown		red
End,Line, Lock	Dead	up	end
Control,Place, Rate	Birth		
Lounge,Hour, Napkin	Cocktail		dinner
Artist,Hatch, Route	Escape		escape
Pet,Bottom, Garden	Rock	rock	or
Mate,Shoe, Total	Running		white
Self,Attorney, Spending	Defense		
Board,Blade, Back	Switch		go
Land,Hand, House	Farm	field	out
Hungry,Order, Belt	Money		out
Forward,Flush, Razor	Straight		straight
Shadow,Chart, Drop	Eye		in
Way,Ground, Weather	Fair	high	out
Cast,Side, Jump	Broad		off
Back,Step, Screen	Door	on	up
Reading,Service, Stick	Lip		up
Over,Plant, Horse	Power	race	other