# Obtaining Query-specific Similar Concepts with BERT-based retrieval for Commonsense Knowledge Gameboard

**Martijn van Woensel***
**Supervisor(s): Gaole He, Ujwal Gadiraju†**
**EEMCS, Delft University of Technology, The Netherlands**
22-6-2022

**A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering**

*M.G.A.vanWoensel@student.tudelft.nl
†{g.he, u.k.gadiraju}@tudelft.nl

## Abstract

Commonsense knowledge based question answering is a recent topic that has seen a surge in interest. Yet most models obtain general data, this paper looks at obtaining query-specific similar concepts using first and second-order proximity together with BERT-based retrieval. Using these query-specific concepts new commonsense knowledge can be obtained using a Game with a purpose. Results show that this current implementation leaves room for improvement.


## 1 Introduction

How would you answer the question “Where do I put my coffee mug after it’s dry?”? Depending on the person you might get different answers but in general, you would get an answer like cabinet or shelf. But this is not the answer to the question. The question is how does a person answer a question? For this specific question, you would be required to know what a coffee mug is, that it becoming dry means that it has been cleaned and after being cleaned it is usually put away and stored somewhere. For a person these things are obvious, but for a computer, this requires a lot of relations between many different concepts to even have a grasp of answering the question.

These relations are known as commonsense knowledge. Commonsense knowledge roughly has two types of knowledge, generative knowledge, and discriminative knowledge. Generative knowledge is relations that are true an example of generative knowledge would be “chair is a furniture”. Discriminative knowledge is relations that can be used to differentiate between things, e.g. “chair is not a mammal”.

Current ways of obtaining this type of knowledge are either done using humans answering questions [1] or using machine learning combined with human workers[2]. In general, these works obtain data for a bunch of different and broad subjects.

But what if you want to obtain commonsense knowledge with a specific query in mind?
In this paper, a method of obtaining different concepts related to a query will be described. The intent of this is to generate a selection of different topics/concepts to be used for obtaining commonsense knowledge. Specifically using this selection for the starting board in FindItOut[3] a game with a purpose which is a “Guess Who” type of game. Where players have to guess the opponent’s concept by asking each other questions in order to learn the opponent’s concept.

Answer the research question “How to obtain commonsense knowledge for a specific purpose?” is an easier task if the question is split up into subquestions.

What are related concepts to the starting concept and do they need to have direct relations to each other? How do we evaluate the “relatedness” of these concepts to the query? How does this method compare to the previously established Greedy maximum coverage algorithm used in FindItOut?

In order to easily obtain a starting concept from a query the CommonSenseQA[1] dataset will be used. Using this starting point a set of related concepts can be obtained using a commonsense knowledge database [4]. To evaluate the quality of the obtained related concepts BERT[5] is used for scoring. BERT is a state-of-the-art Natural Language Processing model. Which is also the best at answering the multiple-choice questions from the starting dataset CommonsenseQA. And to answer the final sub-question BERT is once again used for rating the selected concepts and comparing them to the previously obtained game boards in the FindItOut game.

The paper is structured as follows: In section 2 related work, a summary of similar work will be described. In the next section, section 3 Methodology the pipeline of this method will be presented. Section 4 will show the experiments and findings. Then a brief talk about responsible research and ethics in section 5 and finally the conclusion and future work in section 6.

## 2 Related work

In the past there have been a few Game With a Purpose (GWAP) the goal of the game varies per game, but for this paper in general the goal is to obtain commonsense knowledge via a game. Some examples are the ESP game [6] and Phetch [7]. A more recent example is FindItOut [3], it compares multiple concepts at the same time while playing a ”guess who” type of game, players have to differentiate between different concepts while trying to guess the other player’s concept. This allows FindItOut to generate data about multiple concepts at the same time.

There are some databases such as ConceptNet [4] which is ”a knowledge graph that connects words and phrases of natural language with labeled edges.” (p. 1). It can be used to find topics and concepts related to the concept that is being searched for. This can be used for Natural Language Processing to understand the meaning behind words and concepts. For this research, ConceptNet will be used to obtain similar concepts to the query concept. Using first and seconder order proximity.

In LINE [8] first-order and second-order relations are defined. For this paper, these definitions will also be used.

”**(First-order Proximity)** The first-order proximity in a network is the local pairwise proximity between two vertices. For each pair of vertices linked by an edge (u, v), the weight on that edge, $w_{uv}$, indicates the first-order proximity between u and v. If no edge is observed between u and v, their first-order proximity is 0.”(p. 3)

”**(Second-order Proximity)** The second-order proximity between a pair of vertices (u, v) in a network is the similarity between their neighborhood network structures. Mathematically, let $p_u = (w_{u,1}, ..., w_u, |V|)$ denote the first-order proximity of u with all the other vertices, then the second-order proximity between u and v is determined by the similarity between $p_u$ and $p_v$. If no vertex is linked from to both u and v, the second-order proximity between u and v is 0.” (p. 3)

The first and second-order proximity is used to obtain a set of related concepts that might not be directly connected. Because the concepts share so many neighbors the concepts are likely similar, which means they might be lacking connections and that is exactly why we would want to select these concepts.

CommonsenseQA[1] is a dataset of multiple choice questions that is created using ConceptNet[4] and human workers. The questions are made up of a question, question concept, correct answer, and 4 incorrect answers. This dataset is used to have a starting concept together with a question.

Once a set of related concepts has been obtained it has to be evaluated on how well the concepts relate to the starting question. To do this a rating is needed to rank the concepts. For this BERT[5] can be used, BERT is a state-of-the-art Natural Language Processing model as mentioned in the introduction.

## 3 Methodology

Starting with the query in this paper the commonsenseQA questions are used so that there is a quick and easy way to question the concept. So the actual starting point would be the question concept. From here we start the process as shown in figure 1.

Using ConceptNet a first and second-order relation can be obtained by taking every concept looking at the neighbors and then comparing it with every other concept and looking for the similarities in the concept neighbors. If two concepts share a lot of neighbors they have high first-order similarity. Now there is a ranking of similar concepts.

From here a BERT[5] implementation is used in the form of [9] it implements BERT but in an easier-to-use way. Using the rankings from before a syntax can be set up where we can create sentences in the form Concept1 relation Concept2 relation Concept3. Using this together with the query a score can be obtained in the range of [-1, 1] higher meaning the query and sentence are very related to each other.

## 4 Discussion about Experiments

In this section, a brief discussion will be held about the produced results. In table 1 a few examples of the generated gameboard are shown. There are some asterisks to these generated examples. For example, the generation that was used in the FindItOut game was dependent selecting a bunch of questions and trying to search up different questions with similar concepts. It searches up the question with the most overlap in these concepts and puts them together. With this, a bunch of questions is selected together to create a size of 16 concepts. Thus the game board does not generate specifically to that question but rather a bunch of questions together. Whereas the solution shown in this paper generates depending on one specific query/question. Taking a look at these generated boards one might conclude that the concepts that are generated do not always seem very fitting to the question. Perhaps if multiple key concepts were taken from the query rather than just one key concept the output might have been better.

In figure 2 a boxplot is shown showing how well BERT rates the concepts that are generated using the maximum coverage algorithm used in FindItOut and the one that is shown in this paper. The y axis shows the BERT scoring. Overall FindItOut does better in its best-generated concept compared to the solution in this paper. But this outcome was expected as FindItOut uses the multiple-choice answers that come with the CommonSenseQA dataset. And the average of the FindItOut performing much lower also makes sense, as these gameboards are generated with a bunch of questions rather than just the specific question. So a lot of unrelated concepts appear.

## 5 Responsible Research

In this section the ethical aspects of this research are discussed and the reproducibility of the methods is shown.
The code that has been created for this paper will be available on a github repository that is available on the TU Delft repository site. Here links will also be made to the datasets that are used in this paper. Every dataset's paper has already been referenced in this paper. But for the exact versions of the datasets used in created code will be linked on github. The overall reproducibility of results with comparing to the FindItOut boards is not 100%, as the FindItOut generation uses a lot of randomnesses. It has been reduced a lot from the original code that can be found on github[1]. But it still is not 100% the same every time it is run.

## 6 Conclusions and Future Work

In this paper, a process of obtaining related commonsense knowledge specific to a query has been shown.
A simple solution by using First and second-order proximity is used in order to find unconnected neighbors. Afterward, BERT scoring is used to create a selection of concepts.

Follow-up work could be done by either using higher-order proximity, using natural language models in order to obtain more query-related concepts.

## References


[1] A. Talmor, J. Herzig, N. Lourie, and J. Berant, "Commonsenseqa: A question answering challenge targeting commonsense knowledge," *arXiv preprint arXiv:1811.00937*, 2018.

[2] Y. Gil, J. Honaker, S. Gupta, Y. Ma, V. D. Orazio, D. Garijo, S. Gadewar, Q. Yang, and N. Jahanshad, "Towards human-guided machine learning," in *Proceedings of the 24th ACM International Conference on Intelligent User Interfaces (IUI)*, (Marina del Rey, CA), 2019.

[3] A. Balayn, G. He, A. Hu, J. Yang, and U. Gadiraju, "Ready player one! eliciting diverse knowledge using a configurable game," *Virtual Event*, 2022.

[4] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge," 2017.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.


[1] https://github.com/delftcrowd/FindItOut

[6] L. Von Ahn and L. Dabbish, “Esp: Labeling images with a computer game.,” in *AAAI spring symposium: Knowledge collection from volunteer contributors*, vol. 2, 2005.

[7] L. Von Ahn, S. Ginosar, M. Kedia, R. Liu, and M. Blum, “Improving accessibility of the web with a computer game,” in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 79–82, 2006.

[8] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “Line: Large-scale information network embedding,” in *Proceedings of the 24th international conference on world wide web*, pp. 1067–1077, 2015.

[9] T. Gao, X. Yao, and D. Chen, “Simcse: Simple contrastive learning of sentence embeddings,” *arXiv preprint arXiv:2104.08821*, 2021.

| Query | First and Second order based concepts | FindItOut Concepts |
|---|---|---|
| Glass is used to build these so you can see outside. | door , paint, plate, rock, tree, water, window, wood | plate, vase, restaurant, window, glass, cup, floor, bathroom, backpack, cabinet, dishwasher, windowsill , shelf, sink, books, table, |
| Where could you find a library that people of any age can visit? | apartment, books, college, dictionary, property, school, shelf, warehouse | notebook, inside, hotel, notes, bathroom, school, closet, house, desk, library, meeting, person, hospital, college, bedroom, palace, |
| Where does light come from? | current, lamp, motion, sun, sunshine | human, submarine, floor, ocean, closet, workplace, house, desk, universe, cupboard, attic, sun, chair, hope, light, table, |
| Where would my son place his paper plate? | bowl, clock, cut, food, glass, paper, spoon, table | plate, vase, restaurant, window, glass, cup, floor, bathroom, backpack, cabinet, dishwasher, windowsill, shelf, sink, books, table, |

Table 1: Board generation using First order and second order grouping versus the FindItOut gameboards

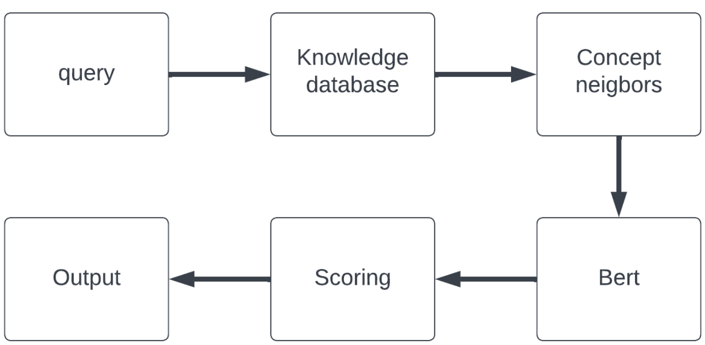


Figure 1: The pipeline of the methodology

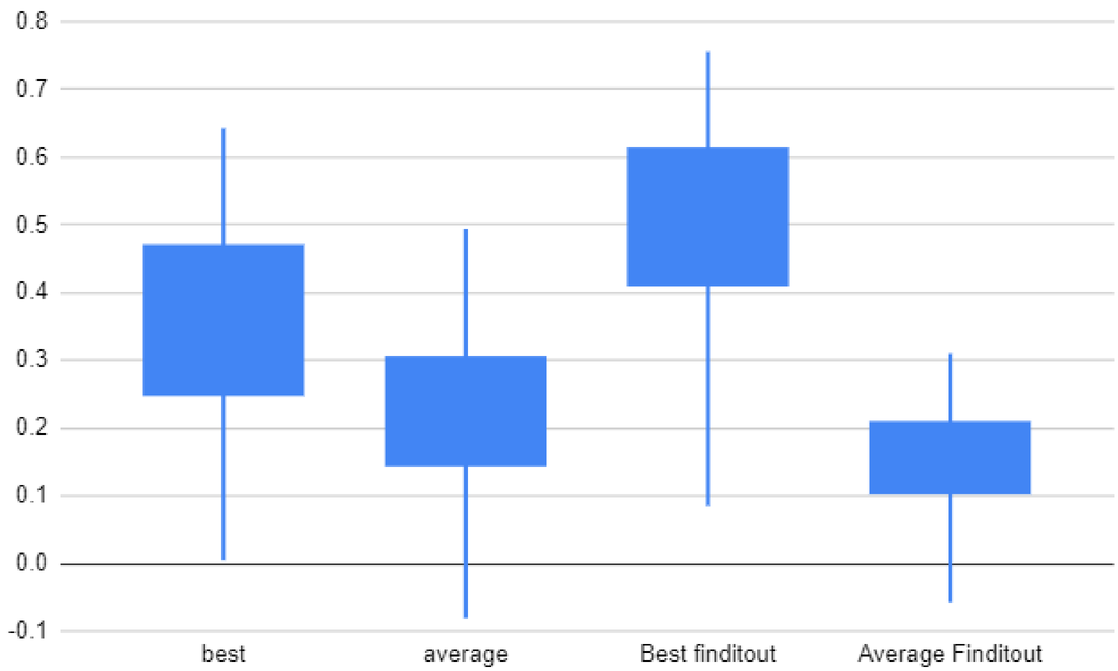


Figure 2: A boxplot showing BERT scoring in the y axis with on the x axis the best case and average case for this paper's method and the FindItOut method