CARD: Credibility assessment model for the Google Knowledge Graph Card

Author: Muriël de Wit University of Twente P.O. Box 217, 7500AE Enschede The Netherlands

ABSTRACT

In 2012, the Google Knowledge Graph Card was introduced. This is the information block Google shows after entering a search query. The purpose of this paper is to illustrate a practical model in order to assess the credibility of the Google Knowledge Graph Card. The model is created from several credibility assessment theories and with the analysis of four Google Knowledge Graph Cards. From that analysis, five categories where distinguished and these categories all have their own credibility assessment method. This assessment method is the CARD model, which is the practical implementation model in order to assess the credibility of the Google Knowledge Graph Card. CARD stands for Comparison, Author, Results, and Doublechecking. Next to that, a demonstration is given of the CARD model and the advantages and disadvantages of the CARD model in comparison to other theories are discussed in the paper. This CARD model can be used by anyone who is interested in assessing the credibility of the Google Knowledge Graph Card.

Graduation Committee members: 1st Supervisor: dr. A.B.J.M. Wijnhoven 2nd Supervisor: dr. M. de Visser

Keywords

Google Knowledge Graph Card, credibility, search results, informative search, CARD, Comparison, Authors and Results, Double-checking

This is an open-access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



12th IBA Bachelor Thesis Conference, July 9th, 2019, Enschede, The Netherlands. Copyright 2019, University of Twente, The Faculty of Behavioral, Management and Social sciences.

1. INTRODUCTION

2012: the year wherein Google introduced the Google Knowledge Graph. A knowledge graph is a knowledge base that collects information from various sources in order to answer questions or gives information about a search query. The output that the Google Knowledge Graph gives is called the Google Knowledge Graph Card (Toonen, 2019). The Google Knowledge Graph Card is the block on the right-hand side of the web browser on a computer or laptop, and on top when searching on a phone. In this block, the answer to a question in the search query is given, or a broad informative summary is given but this depends on the search query.

The Google Knowledge Graph Card can be useful in order to collect information in a fast way. However, concluded from the research of Nakamura et al (2007), people tend to trust that the search engine shows results based on credibility. This is something remarkable and needs extra attention since the Google Knowledge Graph Card is the first thing that people see when entering a search query in Google. This would mean that people take the information in the Google Knowledge Graph Card for granted and think it is credible, without them being aware of the risk that information can be biased.

A lot of research has been done in order to clarify and conceptualize credibility. The problem with those theories is often that the theory in order to measure credibility is not applicable to the Google Knowledge Graph Card. The reason for that is, is that the existing theories are too broad and the needed elements in order to assess credibility are not available in the Google Knowledge Graph Card. The Google Knowledge Graph Card only shows a small amount of information, on a lot of different areas from that search query. Often the theories in order to assess credibility are applicable when a lot of information is available for one sort of area. For that reason, none of the available theories are applicable, but only parts of these theories and a combination of these theories can be used in order to assess the credibility of the Google Knowledge Graph Card.

Considering the power that Google has by showing the information in the Google Knowledge Graph Card, it is important for people to be able to assess the credibility of those cards. The risk that people can think that Google shows results based on credibility, and the fact that there is not a theory that describes how to assess the credibility of the Google Knowledge Graph Card led to the research question: How to assess the credibility of the Google Knowledge Graph Card? In this paper, this research question has been answered. First, several theories are discussed in order to assess credibility. After that, the search query type has been distinguished and categories in the Google Knowledge Graph Card are determined. For these categories, the CARD model has been created in order to assess the credibility of those categories. The CARD model is a combination and has parts of several credibility assessment theories. The CARD model is a practical assessment model for the Google Knowledge Graph Card and stands for Comparison, Authors, Results, and Double-checking.

2. THEORETICAL FRAMEWORK

More in-depth information about the theories that are being used is presented in this chapter. First, the conceptualization of credibility is discussed. After that, it is explained what the Google Knowledge Graph Card is, and another type of Knowledge Graph is introduced. At last, a small introduction to cognitive biases is presented.

2.1 Conceptualization of Credibility

Yamamoto and Tanaka (2005) explain five dimensions in order to assess the credibility. The dimensions are:

- *Accuracy dimension:* this clarifies whether the source is accurate for the search query.
- *Objectivity dimension:* this explains whether the results given are biased in any way.
- *Authority dimension:* this dimension measures whether the author has a good reputation and the skills of the author are good enough in order to create content.
- *Currency dimension:* this stands for the fact whether the information given by the search engine is up-to-date and if it is updated regularly.
- *Coverage dimension:* this explains whether the information is complete and comprehensive.

These five dimensions are derived from the checklist model from Kapoun (1998). In this checklist 27 questions are asked in total, divided under the five dimensions.

However, according to Meola (2004), critical thinking is required while using this checklist from Kapoun (1998) when evaluating the credibility of a search result. He states that there are questions in the checklist, that if it is possible to answer those, the credibility assessment of the search results is already done. This means that not all the questions help by evaluating the credibility of the result and that the questions are too broad. For this reason, Meola decided to create his own contextual approach to website credibility evaluation. He says that the checklist approach is an internal method of evaluation, and the contextual approach focuses more on the external part of the information given on the website. The contextual approach uses three techniques. The first technique is promoting and explaining reviewed resources. In this technique, students do not have to review the source by a checklist, but they can rely on the peer-reviewed assessments from other students. The second technique is called comparison. In this technique, students analyze the differences and similarities of two or more websites regarding their content, or compare a website to another format of information. Think about newspaper articles or scholarly books. Next to that, comparison can help to detect bias. Since the student is not only focussing on a small frame of references, the student is able to detect differences. A difference that can be noticed is the difference between language. More balanced sites use neutral language, while biased sites use more inflammatory language. By comparing this, students are better able to notice any bias. The third technique Meola (2004) describes in the contextual approach is corroboration. In this technique, the information on the website is confirmed against other sources.

Another way of conceptualizing credibility is done by Fogg (1999). He states that credibility is a perceived quality, made up of two dimensions. The perceived quality means that credibility is not a tangible source, but it depends on how the individual perceives it. The two dimensions he discussed are trustworthiness and expertise. Trustworthiness can be defined by terms as truthful, unbiased and well-intentioned. The other dimension is the expertise, which can be defined with terms as experienced, competent and knowledgeable. So, if the credibility is perceived as high, it can be assumed that the expertise and trustworthiness have both high levels (Fogg, 1999).

2.2 Triangulation

Triangulation is another method in order to assess the credibility of internet content. Denzin (2017) describes that triangulation is a method that confirms the truth of statements by comparing data to the same sort of phenomenon, from different views and standpoints and different theories and their research methods. Wijnhoven and Brinkhuis (2015) used inquiring theories together with the related meta-requirements in order to create a prototyping study for the use of an information triangulator.

Inquiring systems have quality requirements included about the information and the systems give requirements for triangulation (Churchman, 1979). From these inquiring systems, the meta-requirements are identified. This identification leads to a set of criteria for evaluating the existing tools about triangulation, and thus credibility assessment. Churchman (1979) identifies five inquiring systems.

- The first inquiring system, called Lockean inquiring system explaining the validity, reliability, and precision of the data need to be checked.
- The second inquiring system, called Leibnizian inquiring system saying that all relevant variables need to be covered.
- The third inquiring system, called Kantian inquiring system saying that all the categories need to be identified in order to be able to assess the completeness.
- The fourth inquiring system, called Hegelian inquiring system demanding that the author is identified, together with his expertise and that from the publisher.
- The fifth inquiring system, called Singerian inquiring system saying that all other four inquiring systems need to be used effectively so that information becomes useful in order to make decisions.

2.3 Knowledge Graphs

A knowledge graph is an information tool which consists of facts about persons, organizations or other topics. This information is gained by free sources given on the internet and is put in one clear overview, the so-called Knowledge Graphs Cards (Rospocher et al., 2016).

'The Google Knowledge Graph uses standard Schema.org types and is compliant with JSON-LD specification' (Google, 2015). Schema.org is a tool to create, promote and maintain schemas for structured data on the whole world wide web, and beyond (Schema.org, 2019). This is the tool the Google Knowledge Graph uses to create Cards for in the Knowledge Graph. This tool creates Cards from structured data. But not all data is structured on the Internet. JSON-LD is the specification that structures the data. This is being done by organizing, connecting and linking the unstructured data so that it is being structured and linked to each other (JSON-LD, 2014).

This is the way the Google Knowledge Graph works, but there exists another, more extensive type of Knowledge Graph; the so-called Event-Centric Knowledge Graph designed by Rospocher et al. (2016).

An Event-Centric Knowledge Graph (ECKG) is a variation on the Knowledge Graph. An ECKG is a Knowledge Graph that takes events given by news articles also in consideration while creating the Knowledge Graph Card.

Nowadays, events tend to get lost in our memories and are not given in a Knowledge Graph Card. This can be of a disadvantage when an individual tries to find information. It might think all the important information is given in the Knowledge Graph Card, but actually, only the hard facts are in there. More fluid information is not taken into consideration, and this could bias the search result. The reason for this is, is that often, the news and events are not provided in a structured content, and thus cannot be found and used by normal Knowledge Graphs.

The difference with this method and others, such as the 'Connecting the Dots in news articles' method from Shahaf and

Guestrin (2010), is that in other 'event-centric' Knowledge Graphs, like those of 'Connecting the Dots in news articles', natural language processes is used in the news articles, and thus the knowledge graph is only collecting facts. By doing this, the chain of facts from the news articles form a structure that can be used for Knowledge Graphs. This is different from the ECKG, because deep natural language processing is used, which creates a knowledge base from the articles (Rospocher et al., 2016).

2.4 Cognitive Biases

Several kinds of biases exist. One type of bias that will be discussed is the cognitive bias. 'Cognitive biases are assumed to arise because of people's limited ability to attend to and properly process all the information that is potentially available for them.' (Kruglanski & Ajzen, 1983). Lau and Coeira (2007) discuss four types of cognitive biases. The first one is the anchoring bias. People tend to focus more on one piece of information (the anchor), and relies on that when making the decision. The second cognitive bias is the order effect bias. The order in which the information is presented will affect the decision of the individual. The third cognitive bias is the *reinforcement effect*. This bias explains that if individuals are multiple times exposed to information, this will influence the final decision. At last, the fourth cognitive bias is the *exposure effect*. In this bias, the final decision will be influenced on the fact that the individual is already familiarized with the given information, and thus will it influence the decision making (Lau & Coiera, 2007).

3. METHODOLOGY

This chapter describes the chosen research method and demonstrates how the decisions are made in the analysis of the Google Knowledge Graph Cards. Two types of Google Knowledge Graphs Cards are analyzed in order to distinguish between categories (see Appendices 1 and 2). These categories are necessary in order to combine those to the literature about credibility assessment. These categories together with the theory on how to assess them, are put together in the CARD model. The CARD model is a practical model in order to assess the credibility of the Google Knowledge Graph Card with the distinguished categories.

3.1 Research Method

The chosen research method is a qualitative research method by which literature is reviewed and used in order to create a new model to be able to assess the credibility of a Google Knowledge Graph Card. First, it is distinguished which format of the Google Knowledge Graph Card is analyzed, and which search query type. After that, two specific search queries are chosen and the Google Knowledge Graph Cards are analyzed on their elements and components. From these observations, categories are distinguished, and they are put together in the CARD model.

3.2 Format and Search Query Type

Before starting to create a model to assess the credibility of the Google Knowledge Graph Card, it is important to distinguish and make clear which type of the Google Knowledge Graph Card will be analyzed. The way the Google Knowledge Graph Card will be presented is depending on a lot of factors. First of all, there needs to be decided whether the research will be done on a computer web browser, a tablet web browser or a phone web browser. The Google Knowledge Graph Cards shows different results depending on the device, while the same search query is being used. The results on the phone, for example, may include hyperlinks to a phone number and leave out other information that is less relevant on the phone.

For this research, it is decided to analyze the Google Knowledge Graph Card from a web browser on a computer. The reason for this is, that this Google Knowledge Graph Card is the most expanded, and thus have more components to check the credibility.

Next to that, it is important to clarify what type of search queries will be used by analyzing the credibility of the Google Knowledge Graph Card. According to Gabbert (2018), there are three types of search queries. She defines a search query as: 'the words and phrases that people type into a search box in order to pull up a list of results.' (Gabbert, 2018). The three types of search queries are the navigational search query, the informational search query, and the transactional search query. First, the three types will be explained briefly, and after that, it will be clarified which type will be used in the research. The navigational search query is a set of words which are filled in, in the search engine, with the intention to find a website or to be directed to a certain website. An example is 'twitter' as the search query. The intention of this is to be navigated to the website of Twitter, <u>www.twitter.com</u>.

The transactional search query is a search query that is being used when someone has the intention to do a transaction or to make a purchase on the internet. For example, when clicking on the 'shopping' hyperlink underneath the search bar, with the intention to purchase something.

The last type of search query is the informational search query. This search query is very broad and the reason for this is, that one search query can come up with a bunch of relevant results. (Gabbert,2018).

The informational search query is the type of search query where this research is focused on. If people put in an informational search query, with the intention to obtain information, the Google Knowledge Graph Card is one of the first results the individual sees, on the right-hand side of the computer screen.

3.2.1 Person and Location Search Query

Now that it is decided to focus on the informational search query, it is important to clarify the types of informational search queries. First of all, people can search for information about a company. This can be a big company, such as Coca Cola, but also smaller, less known companies. Next to that, an informational search query can be about a brand, a person, nutrition, locations, restaurants, news, furniture and so on.

For this research, it is important to use informational search queries that are not too complicated and dependent on other factors. For example, applying Coca Cola as a search query, Google decides what the person will see. Google can give a Knowledge Graph Card of the company Coca Cola as a first result, but also the nutritional values of Coca Cola can be given as the first Google Knowledge Graph Card. Google decides this on the basis of the individual's personalization settings. This means that the search results depend on the specific individual, and that is something that needs to be avoided when doing research. That is why it is decided to analyze two types of Google Knowledge Graph Cards which will have several similarities when using the informational search query. The first one is about a person and the second one about a location. The Google Knowledge Graph Cards of a person will have the same types of components and elements, regarding who the person is that is being searched for. Next to that, the structure will have similarities as well and that is why it is a good option to use in this research. The Google Knowledge Graph Card from a location will also have the same type of components and structure, regarding the location that is being searched for. Not a lot of research has been done yet in order to assess the credibility of the Google Knowledge Graph, so it is important to use clear search queries that do not differ that much when creating a credibility assessment model.

4. **RESULTS**

4.1 Categories from Analysis

From the analysis of the four Google Knowledge Graph Cards, five general categories are distinguished from all the information in the Google Knowledge Graph Card. See Appendices 1 and 2 for the Google Knowledge Graph Cards that have been used in order to distinguish the categories. Since the focus of this research is on the informational search query, all other categories in the Google Knowledge Graph Card are left out of consideration. The main categories that were often in the Google Knowledge Graph Card and which are left out of consideration are the links to websites in order to be navigated to, or the 'People' also search for' category. The reason why the links to websites are being left out of consideration in this research to the credibility assessment of the Google Knowledge Graph Card is because by clicking on the link, the informational search query will turn into a navigational search query (Gabbert, 2018). Next to that, the 'People also search for' category is left out of consideration as well, because this is not relevant for the informational search query of that subject that is being googled. This can be interesting when someone would like to know more about relatives of that person, or maybe surrounded cities, but this is not where the focus is on.

The five categories that did arise from the analysis of the Google Knowledge Graph Card are the picture, a text from Wikipedia with details, a category named launched products, a map and, weather information. See table 1 for an overview of the categories.

The picture and the text from Wikipedia are in every Google Knowledge Graph Card when using an informational search query. These are the so-called joint categories. The picture always shows the search query, but then on image. If a city is the search query, the most typical picture from that city should be shown by the Google Knowledge Graph Card. The text in the Google Knowledge Graph Card is always retrieved from Wikipedia, and is always short and includes the most important terms regarding that search query. The detailed information depends on what the search query is. If it is a person, it could be the date of birth or the number of children. But if it is a location, it could be the population or the area.

If the search query is a person, different things can be shown in the Google Knowledge Graph Card. If the person is an artist, the songs and albums are given in the Google Knowledge Graph Card. But if the person is a writer, books can be given. That is why the category for these types of products or amusement that the person had created is called the launched products.

The fourth category distinguished from the analysis of the Google Knowledge Graph Card is the map. When the search query is a location, the Google Knowledge Graph Card shows this location on a map, so you can see where it is located on the globe.

The last category derived from the analysis of the Google Knowledge Graph Card is the weather information. This is also only shown in the Google Knowledge Graph Card when a location is the search query.

Google Knowledge Graph Card	Category	Category
Joint	Picture	Text
Person	Launched products	
Location	Мар	Weather

Table 1. Overview categories

4.2 CARD Model

For these five categories, a model is created in order to assess the credibility of them in the Google Knowledge Graph Card. This model is called the CARD model. CARD stands for Comparison, Authors, Results, and Double-checking. See figure1. These words arose from a combination of the theory in the theoretical framework and the practical implementation.

First, the four words are explained and a practical implementation tool is explained by this word. After that, in the next section, the model is validated with the help of a Google Knowledge Graph Card. This is done using a location search query.

4.2.1 Comparison

The assessment method of comparison is used in almost every credibility measurement theory. Meola (2004) describes this in his contextual approach as well. By comparing, it is possible to notice differences in the information given. If there are differences being noticed, you can say that the given information in the Google Knowledge Graph Card is not credible. Comparing helps also to assess one of the five dimensions that Yamamoto and Tanaka describe (2005). This dimension is the objectivity dimension. This dimension says that the results need to unbiased and true. By comparing, differences can be noticed, and differences mean that the content is not the same and one of them is biased.

A practical implementation for comparison is suitable for four of the five categories. These categories are the picture, the launched products, the map, and the weather. By comparing these four categories, it is possible to say something about the credibility of those categories in the Google Knowledge Graph Card. Only one implementation method is given here, and that is for the picture category. The other three categories are discussed in section Double-checking.

In order to compare the picture given in the Google Knowledge Graph Card with other pictures, the fastest way is to use the 'images' heading right underneath the Google search bar. By clicking on 'images', all related images to this search query are shown. Now it is possible to compare the picture from the Google Knowledge Graph Card with the other pictures. If the same picture is shown, or pictures that look like the one in the Google Knowledge Graph Card, it can be said that the picture in the Google Knowledge Graph Card is credible.

4.2.2 Author and Results

Author and Results are two words, that stands for one category. This is the text category and is in every Google Knowledge Graph Card. Yamamoto and Tanaka (2005) describe the five dimensions in order to assess the credibility. The two dimensions applicable for Author and Results are authority and accuracy. In the authority dimension, it is checked whether the author has a good reputation. And in the accuracy dimension, it is checked whether the source is accurate for the search query. In order to assess the credibility of that text, two practical implementation practices are explained. The first one is Authors. The text in the Google Knowledge Graph Card has Wikipedia as a source. In order to assess the credibility of the Wikipedia page where the text is coming from, it is possible to look at the number of authors from that Wikipedia article. The more authors that have worked on that article, the more likely it is that the article is credible, and thus the text in the Google Knowledge Graph Card is credible as well.

The second practical implementation is Results. For this tool, it is important to choose a term in the text. The best term would be something that relates to the search query. For example, if the search query is Enschede, a term in that text is 'twents'. Twents is the dialect the people speak in this city and area of The Netherlands. This would be a good term to use in the Results tool. Now that the term is decided, the original search query together with the term can be typed into the search bar. The more results show up where both the term and the original search query are used, the more likely it is that the text in the Google Knowledge Graph Card is credible.

The text in the Google Knowledge Graph Card has always Wikipedia as a source. Denning et al. (2005) addresses a few risks of Wikipedia in their paper: 'Wikipedia Risks'. One of those risks is the fact that anyone is able to modify the information given on the Wikipedia site. Sometimes this will be checked by others, and people will find out that it is not correct, but this is not always the case (Denning et al., 2005). On the other hand. Niederer and Van Dijck (2010) say that not anyone is able to adjust the content of a Wikipedia article and that there is a hierarchy of those who are allowed to edit. Blocked users have the least permission to edit, and the administrators have the most rights. In between from less to more permission, are the anonymous users, registered users, bots and administrators at last, which is just small group of 10 people. There are control protocols and systems that prevent that people without permission can edit the Wikipedia article. Next to that, there are bots who have permission to edit content without the need of human-decision making. They can be recognized on the authors list when they have 'bot' in their username. These bots are created by people who create Wikipedia articles and once the bots are approved, they gain rights to edit and do administrative work such as preventing spam and detect abuse of Wikipedia.

The research of Niederer and Van Dijck (2010) implicates that Wikipedia is not such a bad source, but even though there are the bots, and the users, and the content creators, adjustments to the text is still possible by anyone who is not blocked. Wikipedia does detect big changes, but small adjustments are less recognized. So, people who would like to harm the Wikipedia article have still the opportunity to do so, even though the change is much smaller that it remains like that. For this reason, only looking at the authors is not enough. That is why looking at the results is needed as well.

Thus, for the text category, two practical tools are given in order to assess the credibility of that text. If a Wikipedia article has, for example, only a few authors, this does not mean that the information in Wikipedia is not true. That is why the Results tool is there as well. If both the number of authors and the number of results is high, it can be concluded that the text in the Google Knowledge Graph Card is credible. If the number of authors is low, but the results are high, it is also possible to say that the text in the Google Knowledge Graph Card is credible. But if both the number of authors and the number of results are low, it can be concluded that the text in the Google Knowledge Graph Card is not credible.

4.2.3 Double-checking

Double-checking is additionally to the comparison tool from the CARD model. Meola (2004) describes next to the comparison technique, also the corroboration technique in the contextual approach. By corroboration, he means that the information is being confirmed by other sources. This is what is done by Double-checking. Only comparing from the contextual approach is not enough for the three leftover categories. These categories are the launched products, the map and the weather category from the Google Knowledge Graph Card. In order to assess the credibility of these categories are being compared with well-

known sources. That is why it is called double-checking: you compare and check with a well-known website. For each of the categories, a practical implementation will be explained.

First, the launched products category. The best way to check this category is to compare the information given by the Google Knowledge Graph Cards with other websites that provide this sort of information. For a music artist, it is possible to use YouTube as a source and search whether the songs the Google Knowledge Graph Card shows, are also available on YouTube.

By doing this, the Google Knowledge Graph Card will be compared with another source to see if they give the same output. Next to only comparing the information, it is recommended to compare the Google Knowledge Graph Card with a well-known website or service. If the song or album that needs comparison, cannot be found by a well-known website like YouTube or Spotify, it can be said that the information given by the Google Knowledge Graph Card is not credible. If the launched products given by the Google Knowledge Graph Card are easily found by the well-known other sources, it can be said that the launched products in the Google Knowledge Graph Card are credible.

Second, the practical assessment method to assess the credibility of the map category from the Google Knowledge Graph Card is discussed. Google uses Google Maps when showing the location on the map in the Google Knowledge Graph Card, thus using the heading 'maps' underneath the search bar in Google is not an option now. In order to assess whether Google shows the location on the right place on the map, another mapping website can be used. One of the options would be Apple Maps, Bing Maps or OpenStreetMap. If one of the options shows the location on the same location on the map, as Google Maps does, it can be said that the location category form the Google Knowledge Graph Card can be assessed as credible. Third, the last category for this practical model is the weather category from the Google Knowledge Graph Card. For this category, it is important that it is being updated regularly. A way to assess this is to verify and compare the information given in the Google Knowledge Graph Card with another source. As well as for the assessment of the launched products, as for the map category and the weather category, it is important to use a wellknown source to check with. By doing this, it is possible to verify the information given in the Google Knowledge Graph Card with another source. For the weather category, a good site to use could be www.accuweather.com or www.weather.com. These are wellknown weather websites and update their content very regularly. This is important as well because one of the five dimensions from Yamamoto and Tanaka (2005) is applicable here. The dimension is the currency dimension, which says that the content needs to be updated regularly. For weather, this is all that matters. Weather information from a day before is not what is in the present. If the weather information in the Google Knowledge Graph Card is showing the same results as the well-known website, it can be concluded that the weather information given in the Google Knowledge Graph Card can be assessed as credible.

Now that the model is created, it is important to come back to the theory from Fogg (1999). His theory describes credibility as a perceived quality with the dimensions of expertise and trustworthiness. The method of comparing contributes to the level of trustworthiness. By comparing, trustworthiness will be increased due to the fact that if similarities are noticed, it is easier to trust the content and to think it is truthful. Expertise is also being covered in this model.



4.3 Demonstration of the CARD Model

Now that it is explained how the model works and which practical implementations can be used in order to assess the credibility of the Google Knowledge Graph Card, the model is verified with the help of one example Google Knowledge Graph Card. See figure 2 for that example. A Google Knowledge Graph Card with a location search query is used for this section of the paper. The reason for this is because this includes four of the five categories described in section 4.1. The only category that is not being analyzed and verified in this example is the launched products category, but the same method of comparing and double-checking is required here as the map category.





Frankfurt Figure 2. Example Google Knowledge Graph Card

Prague

I ondon

Germany

Munich

First of all, it is important to see which categories are in the Google Knowledge Graph Card. The picture and the text are always in the Google Knowledge Graph Card, as you can see as well in the 'Berlin' example. Next to that, it can be observed that the location is shown on a map and the actual weather circumstances are given. Those are the four categories that can be assessed on credibility with the CARD model.

The CARD model starts with the C of Comparing. The picture that is shown in the Google Knowledge Graph Card, should look like the ones that Google shows when 'images' is being clicked on. See Appendix 3 for the pictures that Google gave after clicking on 'Images'. The first picture that is shown is the same as the picture in the Google Knowledge Graph Card and the other pictures show the same distinctive elements of Berlin in the pictures. Thus, it can be said that de picture in the Google Knowledge Graph Card with the informational search query of Berlin is credible.

The second letter from the CARD model is A from Authors. The first step is to go to the Wikipedia article from Berlin and click on 'view history' on the right top side. Here it can be seen how many amendments have been made and when. For Berlin, there are so many amendments that it is only possible to see the last 500 made, or there is the option to see the oldest 500 changes. There are so many adjustments made, that it is possible to conclude that the text on the basis of the number of authors is credible. See Appendix 4 for the last changes made to this article.

The third letter is R from Results. This stands for the number of results that appear when picking a term from the text and combine that with the original informational search query which is Berlin. The term picked from this example is Cold War. The amount of search results with the search query 'Cold War Berlin 'is about 59.800.00 results. See Appendix 5. There are so many results that confirm that Berlin and Cold War are linked together, it is possible to say that the text in the Google Knowledge Graph Card with the search query 'Berlin' is credible.

The last and fourth letter from the model is the D from Doublechecking. For the informational search query 'Berlin', two categories are applicable here. The first one is the map, and the second one is the weather information category. For the map, another well-known website that shows locations on a map can be used. Now it is decided to use the OpenStreetMap. On the OpenStreetMap, Berlin is shown on the exact same location as Google Maps did. Next to that, Berlin has the same shape in comparison to the location given in the Google Knowledge Graph Card. See Appendix 6. Thus, it can be concluded that the Google Knowledge Graph Card with the search query 'Berlin', shows Berlin on the right spot of the map.

The weather information needs to be Double-checked as well with a well-known website. The chosen website for this comparison is <u>www.weather.com</u>. It can be seen that the temperature, the wind direction and speed, and the humidity is all exactly the same around the same time. See Appendix 7. It can be concluded that the weather information given in the Google Knowledge Graph Card with the informational search query 'Berlin' can be assessed as credible.

5. ADVANTAGES AND DISADVANTAGES OF THE CARD MODEL COMPARED TO OTHER MODELS

In this chapter, the advantages and disadvantages of the CARD model are given. Next to that, each theory in the theoretical framework is being discussed with their advantages and disadvantages in comparison to the CARD model.

First of all, the biggest advantage of the CARD model is that it is fast and easy to use. Everyone should be able to follow all the letters from the CARD model in order to assess the most important categories from the Google Knowledge Graph Card with informational person and location search queries. Next to the fact that the CARD model is fast and easy, it is a model that took the most applicable parts of existing theories and combined that in order to create a model, especially for the Google Knowledge Graph Card.

Not only has the CARD model advantages, but also disadvantages exist. The biggest one is that it is only applicable for a location or person search query. the AR form Authors and Results can be used for every Google Knowledge Graph Card, as long as it is an informational search query. Comparing can be used for the picture as well, but when the informational search query changes, meaning that it is no longer a person or location, it could be that the CARD model is incomplete or non-applicable at all for the new arising categories.

In comparison to the credibility assessment method with the five dimensions of Yamamoto and Tanaka (2005) and the CARD model, it can be said that the CARD model is a lot faster than checking the credibility based on their five dimensions. The advantage of the five dimensions is that it is more extensive, but the problem with that is as well that it is time-consuming and some dimensions could be combined into one. For example, accuracy and objectivity could be combined into one dimension. If something is biased, it could mean that the results are less accurate.

The five dimensions are derived from the checklist approach from Kapoun (1998). This is a list full of checkboxes with questions in order to assess the credibility. The advantage of this is, is that it is clear. The 27 questions are direct and nothing is left over for imagination. In contradiction to the five dimensions from Yamamoto and Tanaka (2005). Here the five dimensions have a name and a small description, but for example, the term skills in the authority dimensions is very broad. What are skills? Which skills? The CARD model had four words, and each word stands for a clear description in order to implement the practical assessment method. This looks like the checklist approach, but way shorter. The checklist approach is more detailed and extensive, but more time taking as well. So, the checklist approach is clear and direct. The disadvantage of this credibility assessment method is that if it is impossible to answer a few of those questions, the assessment of credibility becomes less reliable.

The third theory that is described in the theoretical framework is the contextual approach from Meola (2004). This approach contains three techniques in order to assess the credibility of the Google Knowledge Graph Card. Two of the three techniques are processed in the CARD model. These are the comparing technique and corroboration technique. The third technique that is not incorporated in the CARD model is the promoting and explaining reviewed sources. The reason for this is that this is not applicable to the Google Knowledge Graph Card. The disadvantage in this technique is that it is time-consuming in order to review a source, but once it is reviewed multiple times, it can be very useful in order to assess the credibility.

The fourth theory is the conceptualization of credibility as a perceived quality by Fogg (1999). He stated that credibility is the perceived quality on two dimensions: trustworthiness and expertise. These two dimensions are both covered in the CARD model, but the disadvantage of this conceptualization is that it is not extensive enough. Credibility is not only expertise and trustworthiness but needs to be up-to-date and the content needs to be clear, for example.

The fifth and last credibility assessment method is triangulation with the inquiring systems from Churchman (1979). The advantages of triangulation are that it collects information from different standpoints and views and present that to the person. Next to that, the meta-requirements from the inquiring systems used in triangulation are very broad. This theory is the most extensive, and complete credibility assessment method. The problem with this is, is that in theory it sounds good, but it can only be used with software systems. Wijnhoven and Brinkhuis (2015) created a prototype taking all the inquiring systems into consideration with the most important meta-requirements. The conclusion from this prototype was that triangulation can help form opinions by searchers when the tool does not only check facts but also gives information from different views and standpoints. The other disadvantage of this assessment method is that it creates a whole new output, and the Google Knowledge Graph Card will not be relevant at all. But the CARD model has some meta-requirements which can be recognized in the model, such as that the author needs to be identified with his expertise, as in the Author word from the CARD model.

6. **DISCUSSION**

There are also some limitations to this CARD model. The first limitation is a limitation that people need to be aware of in any case when looking at a Google Knowledge Graph Card. Humans have cognitive biases and one of them is the order bias. This consist of the fact that people's judgments are being affected by the order in which information is exposed to them. The Google Knowledge Graph Card is always one of the first things people see when they google something. This means that people tend to trust it more since this is the first thing they see. Another limitation is the fact that the analysis of the Google Knowledge Graph Card has only been done with person and location informational search queries. The next limitation of this research is, is that it has only been done on a computer web browser. The Google Knowledge Graph on a phone or tablet looks different, and the CARD model does not fit that well then.

7. CONCLUSIONS AND RECOMMENDATIONS

The research problem discussed in this paper is the fact that no model exists with which it is possible to assess the credibility of the Google Knowledge Graph Card. There exists a lot of theories on how to conceptualize credibility, but none of these theories are applicable to the Google Knowledge Graph Card. The reason for this is that the Google Knowledge Graph Card did not have much information about one category, but the Google Knowledge Graph Card has more categories with limited information. For that reason, the CARD model is created in this paper.

The CARD model is a practical model in order to assess the following categories in the Google Knowledge Graph Card; text, picture, map, weather and launched products. CARD stands for Comparison, Authors, Results, and Double-checking. This means that for four of the five categories from the Google Knowledge Graph Card comparison can be used as an assessment method in order to assess the credibility. These categories are the picture, launched products, map and weather categories. The more similarities that come up during the comparison, the more credible it is.

However, often only comparison is not enough, and doublechecking is required for the weather, map and launched products category. This double-checking can be done by verifying the information with another well-known source. If the information is the same in the well-known source or has a lot of similarities, the more credible it is. This is not necessary when the credibility of the picture in the Google Knowledge Graph Card is being assessed, but for the map, launched products and weather categories it is. The A and R in CARD are there in order to assess the credibility of the text given in the Google Knowledge Graph Card. The text is from Wikipedia. A way to assess the credibility of the Wikipedia article is to look at the number of authors on that article. Next to that, it is possible to pick a term from the text given in the Google Knowledge Graph Card and put that in the search bar, together with the original search query. The number of results that pop up after googling this new search query says something about the credibility as well. The more results, the more credible the text in the Google Knowledge Graph is.

The biggest advantages of the CARD model are that is it very fast, clear, and easy to use in practice, but the disadvantage is that it is less detailed. In comparison to the existing theories, the CARD model is less time consuming, but still covers a lot of the other theories that are applicable to the Google Knowledge Graph Card. The other theories are credibility assessment methods, but CARD is especially for the credibility assessment of the Google Knowledge Graph Card. This can be seen as a disadvantage, but as an advantage as well, since there is finally a model now specially created for this need.

The CARD model can be used by all people who ever use the Google Knowledge Graph Card as a source of information and are wondering how they can assess the credibility of this Google Knowledge Graph Card.

Recommendations for further research are to create a credibility assessment method for the Google Knowledge Graph Card on a different format. This could be a Google Knowledge Graph Card on a phone or tablet. Another option for further research could be to test this model in a survey study or interview experts on the area of credibility assessment. Their opinion could be asked when assessing the CARD model. At last, the analysis of the Google Knowledge Graph can be done with more types of informational search queries, or even with other types of search queries. These could be the transactional search query or the navigational search query.

8. ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor dr. A.B.J.M. Wijnhoven for his time and supervision during the bachelor thesis. Furthermore, I would like to thank my peer students from the bachelor circle for the effective and inspirational meetings we had. At last, I want to give a special thanks to dr. ir. J.F. Broenink with helping me start off this thesis module, and for the tips he gave me during the finishing period of this thesis.

9. REFERENCES

Churchman, C. (1979). *The systems approach and its enemies*. Retrieved from http://epubquickaccess.info/the-systemsapproach-and-its-enemies-finding-page-select-c-westchurchman.pdf

- Denning, P., Horning, J., Parnas, D., & Weinstein, L. (2005). Wikipedia risks. Communications of the ACM, 48(12), 152. https://doi.org/10.1145/1101779.1101804
- Denzin, N. K. (2017). *The Research Act.* https://doi.org/10.4324/9781315134543
- Fogg, B. J. (1999). *The Elements of Computer Credibility*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1. 1.83.8354&rep=rep1&type=pdf
- Gabbert, E. (2018). The 3 Types of Search Queries; How You Should Target Them. Retrieved from WordStream website: https://www.wordstream.com/blog/ws/2012/12/10/threetypes-of-search-queries
- Google. (2015). Google Knowledge Graph Search API | Knowledge Graph Search API | Google Developers. Retrieved from developers.google.com website: https://developers.google.com/knowledgegraph/
- JSON-LD. (2014). JSON-LD JSON for Linking Data. Retrieved from https://json-ld.org/
- KAPOUN, J. (1998). Teaching undergrads WEB evaluation : A guide for library instruction. *College and Research Libraries News*, 59(7), 522–525. Retrieved from https://ci.nii.ac.jp/naid/10012693055/
- Kruglanski, A. W., and Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology*, 13(1), 1–44. https://doi.org/10.1002/ejsp.2420130102
- Lau, A. Y. S., and Coiera, E. W. (2007). Do People Experience Cognitive Biases while Searching for Information? *Journal of the American Medical Informatics Association*, 14(5), 599–608. https://doi.org/10.1197/jamia.M2411
- Meola, M. (2004). Chucking the Checklist: A Contextual Approach to Teaching Undergraduates Web-Site Evaluation. 4(3), 331–344. https://doi.org/10.1353/pla.2004.0055
- Niederer, S., and Van Dijck, J. (n.d.). Article Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. https://doi.org/10.1177/1461444810365297
- Rospocher, M., van Erp, M., Vossen, P., Fokkens, A., Aldabe, I., Rigau, G., Bogaard, T. (2016). Building event-centric knowledge graphs from news. *Journal of Web Semantics*, 37–38, 132–151. https://doi.org/10.1016/J.WEBSEM.2015.12.004
- Schema.org. (2019). Home schema.org. Retrieved from Schema.org website: https://schema.org/
- Shahaf, D., and Guestrin, C. (2010). Connecting the Dots Between News Articles. Retrieved from https://www.cse.iitk.ac.in/users/cs365/2013/hw2/shahafguestrin-11ijc_connecting-between-news-articles.pdf
- Toonen, E. (2019). What is Google's Knowledge Graph? Retrieved from Yoast website: https://yoast.com/googleknowledge-graph/
- Wijnhoven, F., and Brinkhuis, M. (2015). Internet information triangulation: Design theory and prototype evaluation. *Journal of the Association for Information Science and Technology*, 66(4), 684–701. https://doi.org/10.1002/asi.23203

APPENDICES

Appendix 1: Google Knowledge Graph Cards to distinguish category – Person





Appendix 2: Google Knowledge Graph Cards to distinguish category – Location





Appendix 3: Pictures given by Google after clicking on 'images'



Appendix 4: authors Berlin Wikipedia page

Berlin: Revision history	
View logs for this page (view filter log)	
Filter revisions	
External tools: Find addition/removal (Alternate) • Find edits by user • Page statistics • Pageviews • Fix dead links	
For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary. (cur) = difference from current version,	
(prev) = difference from preceding version, \mathbf{m} = minor edit, \rightarrow = section edit, \leftarrow = automatic edit summary	
(newest oldest) View (newer 50 older 50) (20 50 100 250 500)	
 (curl prev) © 12:10, 1 July 2019 PPEscientist (talk I contribs) m (164,616 bytes) (+500) (Added that not just on city but also on district level there are city and dis (undo) (fag: Visual edit) 	strcit partnerships.)
• (curl prev) 21:43, 28 June 2019 203:194.51.42 (talk) (164,116 bytes) (+33) (undo)	
• (cur l prev) 21:36, 28 June 2019 203.194.51.42 (talk) (164,083 bytes) (+12) (undo)	
• (cur I prev) 21:34, 28 June 2019 203.194.51.42 (talk) (164,071 bytes) (+4) (→Galleries and museums) (undo)	
• (cur l prev) 21:32, 28 June 2019 203.194.51.42 (talk) (164,067 bytes) (+24) (<i>→History</i>) (undo)	
• (cur l prev) 21:25, 28 June 2019 203.194.51.42 (talk) (164,043 bytes) (+4) (-Architecture) (undo)	
• (cur l prev) 21:22, 28 June 2019 203.194.51.42 (talk) (164,039 bytes) (+60) (undo) (Tag: references removed)	
• (cur l prev) 21:12, 28 June 2019 203.194.51.42 (talk) (163,979 bytes) (+63) (→References) (undo)	
• (cur l prev) 21:57, 26 June 2019 Coldcreation (talk l contribs) (163,916 bytes) (-1,003) (Undid revision 903625549 by 2601:183:C602:4020:1010:F5BE:FD3B:+	1BBB (talk) Unsourced
and no need to mention so many) (undo) (Tags: Undo, PHP7)	
• (cur l prev) 20:39, 26 June 2019 2601:183:c602:4020:1010:f5be:fd3b:4bbb (talk) (164,919 bytes) (+1,003) (→Nationalities) (undo)	
• (cur l prev) 18:20, 26 June 2019 2003:d4:bf16:8f00:f1e4:4d68:90b2:42e9 (talk) (163,916 bytes) (-147) (<i>Clean up</i>) (undo)	
• (cur l prev) 15:27, 26 June 2019 TiltuM (talk l contribs) (164,063 bytes) (-569) (undo) (Tag: PHP7)	
• (cur l prev) 15:09, 25 June 2019 Monkbot (talk l contribs) m (164,632 bytes) (+5) (Task 14: cs1 template fixes: misused lpublisher= (0×/6×);) (undo) (Tag: AWB)
• (cur I prev) O8:10, 25 June 2019 Xaiver0510 (talk I contribs) (164,627 bytes) (+32) (→20th to 21st centuries: added wiki link) (undo)	

Appendix 5: amount of search results with search query 'Cold war Berlin'



About 59.800.000 results (0,56 seconds)

Appendix 6: Berlin located on OpenStreetMap



Appendix 7: weather information Berlin

