**Evaluation of Depression in Adolescents and Teens Using Different Search Services**

**Introduction:**

Title: Depression in Adolescents and Teens

Description: Find information on depression in adolescents and teens with prevention techniques.

Narrative: A relevant document may include the development of depression in adolescents and teens. Discussion of effective treatment strategies, such as cognitive-behavioural therapy, antidepressant medications, and family therapy, in managing adolescent depression and preventing relapse is also relevant. Documents that focus on adult depression or depression in other age groups are not relevant.

**Evaluation Policy:**

In determining which documents are relevant or not, I adopted different strategies depending on the search service. In Google, Bing and StartPage when using both query types and ProQuest using Boolean query, SocialSearch using ranking query, I read through the first 20 documents from the results and determined if it was relevant or not. However, for YouTube and Bing Video I implemented a different approach in assessing relevant videos from top to bottom and from left to right respectively due to the different presentation layouts. As videos have different durations, it would be time-consuming to watch all the videos in full length. Therefore, to determine the relevancy of these videos I only watch these videos for 5 minutes. A different approach has to be applied to Bing Images and Google Images such as following the grid layout presentation in assessing images from left to right and top to bottom.

**Facet Analysis:**

I have decided to use the predefined PICO (Problem/Patient/Population, Intervention, Comparison, Outcome) for the facet analysis of this topic. PICO is a widely used framework that is specifically designed for formulating clinical questions in the domain of healthcare and medicine. For this topic, it is suitable to use PICO rather than the other facet analysis frameworks such as PMEST or ad-hoc as it helps clarify the key components of a clinical question, thus making it easier to search for and review relevant evidence. By implementing the PICO framework, I examined the narrative field of the topic and was able to identify the terms within each facet.

Problem/Patient/Population: Adolescents, depression.

Intervention: Cognitive-behavioural therapy (CBT), medications, family therapy

Comparison:

Outcome: Reduction in depressive symptoms, prevention of relapse and improvement in overall mental well-being

I have placed ‘adolescents’ and ‘depression’ in ‘P’ as these terms describe the population and problem respectively. Individuals in the adolescent age group who are experiencing depression are the main focus of this study. I have placed ‘Cognitive-behavioural therapy (CBT)’, ‘antidepressant medications’, and ‘family therapy’ in ‘I’ as they describe different types of interventions or treatments being considered. As the narrative field does not include the information need for a comparison, therefore ‘C’ is left blank. I placed ‘Reduction in depressive symptoms’, ‘prevention of relapse’, and ‘improvement in mental well-being’ in ‘O’ as these are possible outcomes when treatments are given to adolescents.

**Search Strategies:**

Using the PICO analysis created, I started on a web search engine Google to understand my topic as I am the most familiar with this engine compared to other search services that I will go through and see what results I would achieve. Then, I looked at the words in each facet. For my preliminary search, I used words that are similar using a thesaurus which leads to a lot of words being used and along with these unsuccessful results. Therefore, I decided to use fewer words which proved to be better in my search query as using words with similar meanings does not increase the search for related documents but rather confuses the search engine. Therefore, I implemented the following strategies to simplify my Boolean and ranking queries to ensure relevant documents are being given in the result. If the same phrases/ words are being used in relevant documents from the result then I will keep them in my query such as teenager and depression. If new phrases/ words are being used in relevant documents then I will experiment using it in the query. I would use similar words to improve relevance in the search query, by removing words/phrases that are frequently found in documents that are irrelevant to our topic. To achieve the most relevant searches, I had to employ these strategies which is a long iterative approach. During this, I noticed words that introduce ambiguity led to a higher number of irrelevant documents than using particular terms.

For the ‘ranking’ and ‘Boolean’ queries I developed these queries using facet analysis along with similar word choices for both. However, I used wild card operators when creating the Boolean query and adding a few words to different search services.

For formulating the ranking query, I decided to use fewer words which would later become successful. I expected that plurals and the preposition ‘in’ would result in fewer results but relevant documents, however, this did not affect the number of results or relevancy of the documents retrieved. The most successful search was "depression" “adolescent” "teenager" Medications Family Therapy Cognitive-behavioural therapy Symptom Improve Better Well-being where the top 10 of the documents retrieved were relevant to this topic and no duplicates. Other searches that were not successful were "depression" “adolescent” Medications Family Therapy Cognitive-behavioural therapy Symptoms Improved Better Well-being which included depression in children which is not relevant. I decided to include “teenagers”, the present tense of “improved” and the non-plural version of symptom to widen the search net. Depending on the search service I will use the final ranking ‘”depression" “adolescent” "teenager" Medications Family Therapy Cognitive-behavioural therapy Symptom Improve Better Well-being.’

For formulating the Boolean query, similar to the ranking query I used fewer words in the query. When using the ‘ranking’ approach for non-media searches, it proved to be effective, whereas the ‘Boolean’ approach needed additional adjustments for various search engines. I experimented using wild card operators which proved to be difficult to use on searches on the web.

To widen the search net, I used the following wildcard operators along with the Boolean operators:

Boolean: AND, OR

Proximity: Adjacency “ ”

Wildcard: Truncation \*, to include plurals

The initial Boolean query is below:

(Depression) AND (Teen\* OR Adolescen\*) AND (Medications OR Family Therap\* OR Cognitive-behavioural therapy (CBT)) AND (“Symptom improve\*” OR “Better Well-being”)

I experimented using various wildcard operators in different places to see if I could achieve relevant results. Below is one of the experimented Boolean queries:

(Depression) AND (Teen\* OR Adolescen\*) AND (Medication\* OR Family NEAR/ 1 Therap\* OR Cognitive-behavioural therapy OR CBT) AND (“Symptom improv\*” OR “Better well-being”)

However, the use of wildcard operators is not difficult as all search engines do not use the same operators thus making it hard to convert in various search engines. Using the experimented Boolean query did not widen our search net and included the overall topic of depression. Therefore, I will try to simplify this to achieve better results by using Boolean operators (AND, OR), proximity operator (adjacency) and only one wildcard operator (truncation). Our final Boolean query is below:

(Depression) AND (Teens OR Adolescents) AND (Medications OR Family Therapy OR Cognitive-behavioural therapy (CBT)) AND (“Symptom improve\*” OR “Better Well-being”)

I implemented the Successive Fractions strategy for our final Boolean below, this will change depending on each search service:

{Set 1} = {Depression}

{Set2} = {Set1} AND { Teens OR Adolescents}

{Set 3} = {Set2} AND { Medications OR Family Therapy}

{Set 4} = {Set 3} AND {(“Symptom improv\*” OR “Better well-being”}

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Search Service** | **Query Type** | **P@5** | **P@10** | **AveP** | **RT-Dup** | **LB** | **NT** | **SPAM** |
| Google Web Search | Boolean  (2Dsearch) | 0.60 | 0.80 | 0.58 | 0 | 0 | 0 | 0 |
| Ranking | 0.80 | 0.80 | 0.58 | 0 | 0 | 0 | 0 |
| Bing Web Search | Boolean  (2Dsearch) | 0.80 | 0.70 | 0.55 | 0 | 0 | 0 | 0 |
| Ranking | 1.00 | 0.80 | 0.77 | 1 | 0 | 0 | 0 |
| Google Images | Ranking | 0.40 | 0.50 | 0.31 | 1 | 0 | 0 | 0 |
| Bing Images | Ranking | 0.20 | 0.10 | 0.03 | 0 | 0 | 0 | 0 |
| YouTube | Ranking | 0.80 | 0.50 | 0.48 | 1 | 0 | 0 | 0 |
| Bing Video | Ranking | 0.00 | 0.00 | 0.00 | 0 | 0 | 0 | 0 |
| ProQuest Dialog | Boolean | 0.60 | 0.60 | 0.34 | 0 | 0 | 0 | 0 |
| Social Searcher | Ranking | 0.80 | 0.70 | 0.51 | 0 | 0 | 0 | 0 |
| Startpage | Boolean | 0.60 | 0.50 | 0.43 | 0 | 0 | 0 | 0 |
| Ranking | 0.60 | 0.80 | 0.51 | 0 | 0 | 0 | 2 |
| ChatGPT | Other | 0.40 | 0.50 | 0.22 | 0 | 4 | 0 | 0 |

Table 1. Evaluation metrics for all search services with two different query types (Boolean and Ranking)

**Evaluation:**

Table 1 shows the performance of each search service along with its query type (either ranking or Boolean) in terms of P@5, P@10, AveP and if duplicates documents are found (RT), if the link is broken (LB), if the search does not retrieve any documents (NT), and if a document is spam (SPAM). I will be going through all the findings for each search service and the corresponding query type. Firstly, I noticed that the web search services that allowed both query types were 2D search using Bing and Google and StartPage. From this, we see that in Bing using ranking is higher in terms of AveP (0.77), P@5 (1.00) and P@10 (0.88) than Bing using Boolean query in 2D search. Whereas, in Google using ranking is only a higher value of P@5 (0.80) while AveP and P@10 are identical (0.80 and 0.58 respecitvely). The opposite occurred in Startpage where the P@5 (0.60) is identical in both queries, but the AveP (0.80) and P@10 (0.51) are higher in ranking query type. These results portray that the rank-based list of words performed better than Boolean queries. This indicates that our Boolean queries need more adjustments to perform better or that the ranking query used in these search services was optimal. Another reason is Google limits the number of terms in the search to 32 leading to a poorer performance in the Boolean query type. We expect day-to-day users that use search services that provide the two different query types to use ranked-based list of words method as it is easy to understand making companies spend more time and resources enhancing their algorithms to optimise the searches for its users.

Secondly, From Table 1, there are duplicate documents in visual search service, Google Images and the web search service, Bing when using the rank-based method. The duplicate present in Bing is an academic article which is published on at least one website such as ResearchGate or that is not the website where the article is published such as PubMed. ResearchGate is a professional networking site website where researchers and scientists share their published work. Duplicates were not present in ProQuest due to full-text indexing by narrowing down the search net and giving relevant documents. Also, ProQuest only allows the search based on its own databases, making it proficient and well-organized than commonly used search services like Bing and Google which are web search services. We expected similar results on the video search services as users upload their content on YouTube making this search service have its own database, whereas Bing Video gets its results from YouTube, Vimeo, Dailymotion and many others. However, a duplicate was present in YouTube and not Bing due to the differences in content indexing.

Thirdly, A strategy was implemented for video search engines to determine the relevancy as videos have varying durations. From Table 1, Bing Video using the ranking-based method was the worst performing search service overall with 0 relevant videos, also performing worse than YouTube using the ranking-based method in terms of the evaluation measures. The videos that came from this search only involved Cognitive Behavioural Therapy (CBT). Both video search services did not have any SPAM, repeated, link broken, and not retrieved documents. The reason for no SPAM documents in both video search services is due to the strict moderation policies and community guidelines present resulting in cleaner and only relevant documents being retrieved. We expected the video search services to perform poorer than ProQuest as the area of interest for our topic is public health, social work, and neuroscience which would be more present in academic papers, however, YouTube performed better than ProQuest. This could be due to our Boolean query; therefore, fine-tuning is required to enable the full capabilities of the query.

Fourthly, I employed a systematic approach, assessing images from left to right and top to bottom due to the grid layout presentation. Both image search services from Google and Bing indicate that these search services are not appropriate for our topic. Google Images achieved a higher precision at 5 and 10 images and a higher average precision than Bing Images as it is the second lowest performing search service after Bing Video. Most of these images from both search services covered depression overall but not in teens and adolescents. Google Images had a higher precision and resulted in more relevant images because Bing Images only gave images that were related to CBT overall rather than specifically for our target age group. The only relevant image from Bing Images is an image from an academic blog on CBT on teens to treat depression. The remaining images from Google Images involved depression in kids and children along with intervention techniques such as CBT and medication which were taken from academic articles and blog posts which I decided to not be relevant. Also, we noticed majority of the documents retrieved in Google using the ranking method and Boolean query were academic research papers while the rest were blog posts (80% for both).

Fifthly, SocialSearcher with ranking query was the second highest performing search service along with YouTube with ranking query and Bing with Boolean query. For the ranking search query in this search service, I experimented using various words that were in our PICO facet analysis such as ‘teen’, ‘symptom improvement’ and ‘better well-being’. However, from this iterative process, I realised that fewer words would provide better results in terms of relevancy. SocialSearcher retrieved 7 relevant documents with similar results from the web search engines that contained most of the terms in our PICO facet analysis. The performance of the SocialSearcher was expected because of the high level of activity from the academic computer science users on Twitter.

Sixthly, ChatGPT performed poorly with AveP of 0.22, however retrieved 5 relevant documents that also did not lead to a broken link. Compared to the rest of the search services ChatGPT has the highest number of link broken documents. This is due to ChatGPT’s training data being static, the links were most likely working at the time of the data collection but have since may resulted in being broken or outdated. ChatGPT provided documents where the link was broken and no redirection was given thus being irrelevant. For example, this research paper “Depression in Teens” gave an Error Message (404 not found) as the content was not available, I classed this as irrelevant. Also, ChatGPT does not have the capability to check the present validity of the sources it creates which creates inaccuracies.

Seventhly, I came across two spam documents in our investigation which were from web search service, StartPage using the ranking query. For example, ‘Bessel van der Kolk, MD - Course with Dr. Van der Kolk’ is a trauma treatment course which contains the search term ‘depression’ in the meta tags, but inspecting it further had no relation to the search results page this leads to user confusion. Therefore, it was classified as irrelevant to our topic. However, the rest of the 8 documents were relevant where most of them were academic research papers such as ‘CBT for Teens: How Cognitive Behavioral Therapy Works’ and ‘Teen Depression’.

Lastly, in Figure 2, we see the query used to produce our search results. With the command line search service, ProQuest, we achieved the results P@5 and P@ 10 at 0.6 and the Average precision at 0.34. I expected ProQuest to be the top performing search service for our topic. Upon more investigation, a majority of these documents were dissertations and theses. These documents mentioned all the terms mentioned in our PICO facet analysis apart from CBT because I noticed the relevancy of documents would decrease when I included this term, therefore I decided to remove it. For example, documents that only discussed depression in the majority of adults and none of the intervention techniques were mentioned. Before document position 14, I retrieved relevant documents however after this I found irrelevant topics that only included adult depression which was expected. To resolve this issue, I could narrow the search net by applying a filter to exclude dissertations and theses which can be done in ProQuest. However, implementing filters in this service search service would not give us true results of our Boolean query.

**Summary:**

Overall, from our evaluation, we see that the searches using the ranking method performed better than searches with the Boolean query. Boolean query has the capability Also, the best search service for this topic is Bing with the ranking query ‘ ”depression" “adolescent” "teenager" Medications Family Therapy Cognitive-behavioural therapy Symptom Improve Better Well-being’. The worst search service is Bing Video using the ranking query. It is also important to take into consideration of spam, duplicates and link broken documents as it affects the effectiveness of the search results. We have noticed that visual search services such as video and images are not suitable for our topic, whereas web search services are. We expected the command line search service ProQuest to perform better than the web search services but the results showed that this was not the case. From this assignment, I noticed that visual search services and web search services are more user friendly and provide quick results than command line search service as precise search capabilities. Issues faced in this coursework involved tweaking the Boolean query with different operators which led the search service to give irrelevant results. For future work, I could fine-tune the final Boolean query by implementing the Boolean operator ‘NOT’ to solve the issues of adult depression being involved in the search results, this will allow the search net to be narrowed down further. Also, apply filters so only relevant documents can be retrieved, this will help solve the issue of retrieving dissertations and theses being given in the search results.

A screenshot of a computer

Description automatically generated

Figure 1. Search query from SocialSearch (Ranking) along with a few search results

A screenshot of a search engine

Description automatically generated

Figure 2. Search Query in ProQuest using Successive Fractions