Skincare Chatbot

Vivian Xia

Northwestern University MSDS453: Natural Language Processing

Syamala Srinivasan

March 12, 2022

Introduction & Problem Statement

The skincare industry has and will continue to grow significantly. It is predicted that the market will expand 4.8% each year for the next seven years. Consumers have focused more on skincare and their physical appearance in the last few years especially in skin disorders and malignancies. Skincare products with natural ingredients are likely to see significant demand over the years from the growing consumer inclination toward natural beauty products ("Skin Care Products Market," 2021). With a more focused interest in the ingredients that are in their products, it would be beneficial to build a chatbot to answer skincare ingredient questions. These questions would include what each ingredient is and what it treats.

The objective is to build a chatbot that answers questions on skincare ingredients similar to a customer service agent recommending certain ingredients to customers based on skincare concerns. In order to create a chatbot, text from four web articles were collected. The four sources include Refinery29, Good Housekeeping, Byrdie, and Vogue. These texts were selected with the objective in mind, so the documents contain information on skincare ingredients and the skin concerns that they treat. An ontology will be constructed based on the collected skincare documents to gain an understanding of the document contents and the type of questions that the chatbot will be able to answer. Knowledge graphs will also be built to visualize the relationships between entity pairs in the documents. The chatbot will then be constructed using NLTK by vectorizing the text from the documents and computing the similarity between the question and sentence tokens. With the use of cosine similarity, the chatbot will be able to answer the question with the most similar sentence to that question. Another chatbot will be built with the same objective of answering skincare ingredient questions. It, instead, will only be able to provide a binary yes/no response to questions. This chatbot will be built using keras with an attention-LSTM model.

Data

The library os was imported to allow access to the datasets. The libraries pandas and numpy are used for formatting the data. The library spacy is used to extract entity pairs and relations to build the knowledge graphs. The library network is used to create the graphs. Matplotlib is used to visualize graphs including the directed graph and performance metric plots for the models. The tqdm package shows the progress of a process using a progress bar. Tensorflow, keras, and sklearn are used to build the model. The re, string, and nltk library are used to clean and preprocess the text.

The data is sourced from four web articles including Refinery29, Good Housekeeping, Vogue, and Byrdie. These four sources include information on different skincare ingredients. Bydie, in particular, defines each ingredient in more scientific terms. The other three sources reference different ingredients and its effect on various skin concerns. Due to the formatting of the articles, the text was copied and pasted into a spreadsheet. The ontology referenced the document text in its original form. For the knowledge graph, the data is imported with all four texts combined and split at periods and added to the corpus. For the analysis of the knowledge graphs, the data is later normalized to lower case and the entity pairings that only have one or none entities are filtered out of the dataframe.

The chatbot built using NLTK is constructed from the data from the four sources. The four texts are joined into one text. It is then preprocessed by tokenizing the text into sentence tokens and work tokens. The text is further cleaned by converting all tokens to lowercase, removing punctuation, and applying lemmatization. The text is then vectorized to allow for the computation of cosine similarity.

For the chatbot built using keras, a dataset of fact, question, and answer is manually constructed from the document contents. Each fact in the fact column is compiled of one or many sentences from the documents. The question column contains questions relating to its corresponding fact, and the answer column contains the answer to that corresponding question.

Experiment 1's dataset contains 50 rows with 22 "No" and 28 "Yes" answers. The dataset was cleaned by removing punctuation and other non-alphabetic tokens from the documents. The training set is made up of 17 "No" and 20 "Yes" answers, so the training set is about balanced. The training and test set are created from the clean data so that the training set has 37 and the test set has 13 instances. To try and improve model results, Experiment 2 uses a larger dataset of 95 instances with 48 "Yes" answers and 47 "No" answers. The text for this experiment had non-alphabetic tokens removed as well as lemmatization, resulting in 650 tokens. The training set consisted of 71 instances with 37 "Yes" and 34 "No" answers.

The vocabulary was created from the unique tokens from the facts and questions as well as the addition of "no," "yes," "?," and ".". The vocabulary for Experiment 1 had a total of 723 unique tokens and Experiment 2 had a total of 650 unique tokens. The facts, questions, and answers are then vectorized by tokenizing the vocabulary so that each token is encoded with an index.

Research Design and Modeling Method(s)

Document Ontology

An ontology was created to organize and represent the document and its concepts. It is used to identify the documents on skincare to allow for a better understanding of the document contents. With the number of ingredients in each document, it is important to organize them into their respective classes. From these classes, it will make it easier to comprehend the context behind each document as well as generate the type of questions that the contents of these documents can answer. A manual ontology is first created as a draft to gain an understanding of the classes and sub-classes that should be created for the domain of the documents on skincare. The final ontology is created using the ontology editor Protégé. This software allows for a more organized view of the concepts in order to analyze the relationships.

Knowledge Graph

A knowledge graph is constructed from the document contents and Python algorithms to observe the entities of the sentences and their relationship with one another as well as with other sentences in the text. The data is imported and split at the periods before adding it to the corpus. The entity pairs are found using an algorithm that looks for punctuation, compound words, modifiers, and subjects. The pairs extracted represent the subject and object in each sentence. The relations are extracted to be used as the identifying relationship between the entity pairings.

A data frame of the associated source, target, and edge tokens is created to observe the relationship between the three. To ensure that the same spelled entities and relations can be identified in the knowledge graphs, the data frame is normalized into lowercase. There are also orphan entity pairs where only one or no entities are extracted from that sentence. Because those nodes do not provide much information on the relationships between entities, those rows in the data frame are filtered out before using the data frame to create the directed graphs. The data frame decreases from 369 rows to 316.

Building Chatbot using NLTK

This chatbot was created using the document text. The tokens were vectorized using tf-idf to transfer the text into a numerical representation that takes the number of times each term is in a document and how common the term is in the document. The tf-idf tokenizer also removes the stop words from the sentences when it converts the tokens. Using cosine similarity, the similarity between the question asked by the user and the words in the corpus is evaluated to return a sentence from the corpus that is most similar to the tokens in the question asked (Dass, 2018).

Building Chatbot using Keras

With the use of text, a type of sequential data, an LSTM model has a useful architecture that can capture those sequence patterns. It is able to memorize these correlations and use it as context when predicting. The LSTM model is able to memorize short-term and long-term sequence correlations so that it remembers the relevant correlations but forgets the irrelevant ones. To improve on the memory even further, an attention-LSTM is used to build the chatbot. An attention model gives a higher weight to the relevant parts by paying attention to what is more important. In this case, the attention model will give a higher weight to the matching words of the input fact and question (z_ai, 2019).

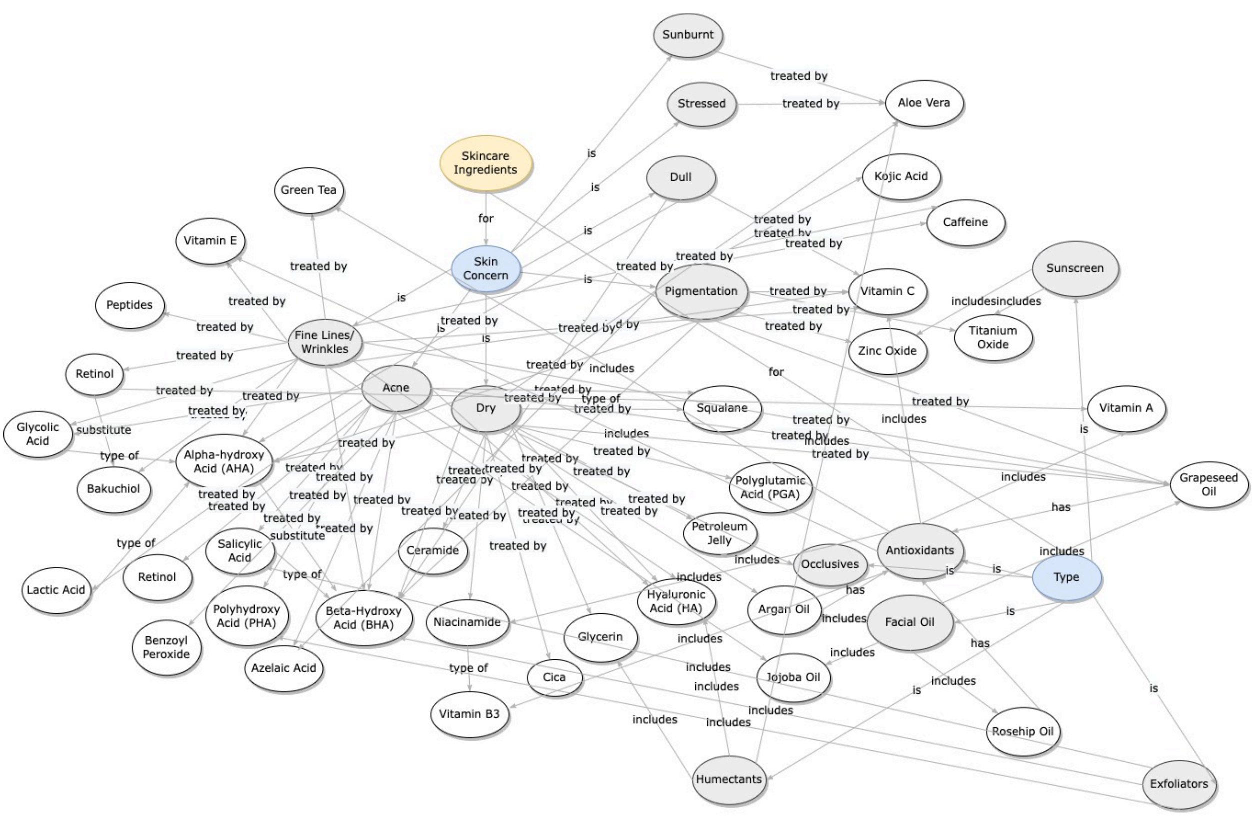
This attention-LSTM model is built using a Sequential class. The inputs sentences are encoded two times, once with an output dimension of 64 and another with an output dimension of the longest question length. The input questions are also encoded using a question encoder. The attention to the relevant tokens is found by taking the dot product of the encoded sentence that had same 64 output dimensions as the encoded questions. By computing the dot product and doing the softmax, the similarity of the sentence and question is evaluated so that the model can give a higher weight to those matched words. The matches are added to the other set of encoded sentences to evaluate the output vector. With this output response vector, it is concatenated with the encoded questions to be used as an input to the LSTM layer. A dropout layer is added to help with overfitting. The output layer is a Dense layer with the size of the vocabulary and activation function softmax (z_ai, 2019).

Each model is compiled using the rmsprop optimizer, categorical crossentropy loss function, and accuracy metric. Both models will be trained with a fixed batch size of 15 and 150 epochs. The model performance is recorded in a table in Figure 17. The performance will also be visualized using performance metric plots for loss and accuracy as well as a confusion matrix.

Results

Document Ontology

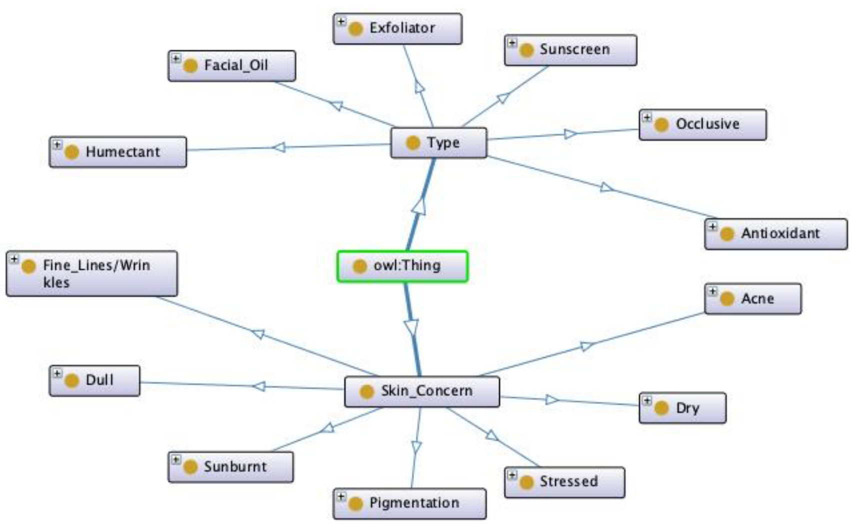
Figure 1. Manually constructed ontology for skincare ingredients.



The manual ontology draft has the classes based on skincare ingredients. The domain is represented in yellow. The classes are in blue, and the sub-classes are in gray. The class Skin Concerns include fine lines/wrinkles, acne, dry skin, pigmentation, and dull skin. The ingredients that can treat each concern are identified.

The manually created ontology is noisy and not organized, so it will be translated into Protégé, an ontology editor, for better visualization and comprehensive of the relationships. Figure 2 shows the classes and sub-classes of the ontology.

Figure 2. Protege ontology with classes and sub-classes.



With the ontology, the document text includes information on fine lines/wrinkles, dull, sunburnt, pigmentation, stressed, dry, and acne skin concerns and how they can be treated using specific ingredients. There are also many different types of skincare ingredients that can be used to classify the ingredients. The questions that can be answered from the text includes, for example, “What is retinol?” where the user can ask for a more technical definition of different skincare ingredients. The text also contains information on which ingredients can treat different skin concerns and vice versa, which skin concerns does a specific ingredient treat such as “What skin concern does grapeseed oil treat?” and “What ingredient improves dry skin?” Figure 8 also shows that there are skincare ingredients that are a type of another ingredient as well as substitutes for one another, enabling questions such as “What ingredient can substitute tea tree oil” or “Is salicylic acid a type of BHA?” These questions will be answered by the chatbot using NLTK.

Knowledge Graph

The entity pairings and relations are extracted from the text. The relations “is,” “use,” “are,” “says” are the most common for the dataframe. The knowledge graph visualizes the relationships between the entity pairings found using the Python algorithms. After normalizing and removing orphan nodes, the most common relations are "is," “says,” “are,” and “helps.” A subset filtering for each of these relations is visualized into a knowledge graph. Subsets filtering for different ingredients, skincare concerns, and skincare ingredient types are also visualized to analyze relationships. With the knowledge graph, there is coreference where ingredients are referred to as “it.” By using “it,” there is less context into which ingredient the target has a relationship to.

Building Chatbot using NLTK

The chatbot is created using tf-idf vectorization and cosine similarity. With the questions generated from analyzing the ontology as the input, the chatbot successfully answers the questions as shown in Figure 16. However, there are formatting issues in the answer because the texts that were extracted from the web articles used articles to refer to the ingredients, so the answers are not grammatically correct. The answer for “What ingredient can substitute tea tree oil?” is correct but does not mention the context of being able to substitute tea tree oil in the answer itself.

Figure 3. Chatbot using NLTK results.

Question	Answer
What is retinol?	retinol what it does: a fat-soluble derivative of vitamin a, retinol is the reliable otc anti-aging ingredient.
What skin concern does grapeseed oil treat?	grapeseed oil according to green, “grapeseed oil is packed with antioxidants such as vitamin e, a, and c.” she says that “using grapeseed oil can improve the texture, tone, and elasticity of the skin.
What ingredient improves dry skin?	“squalane improves the skin’s elasticity and moisture retention as well as protects the skin against sun damage.
What ingredient can substitute tea tree oil?	argan oil: a moisturizing, fragrance-free plant oil made from the kernels of argan trees.
Is salicylic acid a type of BHA?	salicylic acid is the most common bha in skincare.

Building Chatbot using Keras

The chatbot was built using an attention-LSTM model to predict the correct answer for each question by capturing the sequential correlations between the facts, questions, and answers. Experiment 2’s model had a better validation accuracy score compared to Experiment 1’s model. The larger dataset and smaller number of nodes in LSTM helped to improve the performance of the model. However, the validation accuracy was still bad for Experiment 2 and was barely better than randomly guessing. Both experiments also showed overfitting. Overall, the chatbot using the attention-LSTM model did not perform well.

Analysis and Interpretation

Document Ontology

While creating the manual ontology, AHA and BHA were commonly referred to as a treatment ingredient for multiple skin concerns including acne-prone, dry, pigmentation, dull, stressed, sunburnt skin. The classes in skin concerns with the most individuals include acne, dry, fine lines/wrinkles, and pigmentation. One observation that is noted is that all the ingredients containing acid such as azelaic acid and kojic acid do not all belong in the exfoliators sub-class of Type. The ingredients that the documents note as substitutes for the other are identified such as retinol and bakuchiol. There are ingredients mentioned that are a type of another ingredient such as glycolic acid and AHA, so these relationships are identified. Most ingredients address multiple skin concerns. Not all ingredients were addressed as part of a Type class. Some classes in Type address specific skin concerns such as the instances of humectants can all address dry skin.

While recreating the ontology in Protégé, all the types of facial oils contain antioxidants, so that relationship between the two sub-classes is identified rather than each instance of the facial oil sub-class being related back to antioxidants.

Figure 4. Protege ontology humectant sub-class.

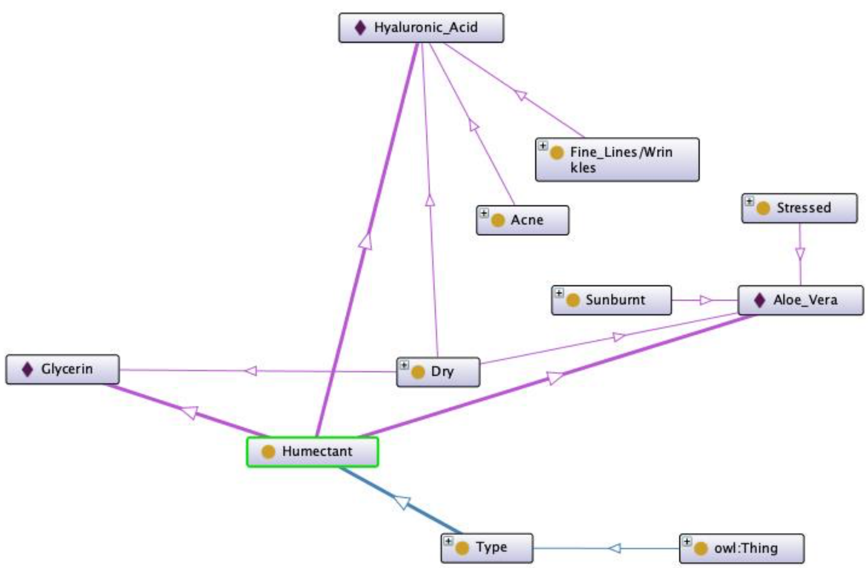
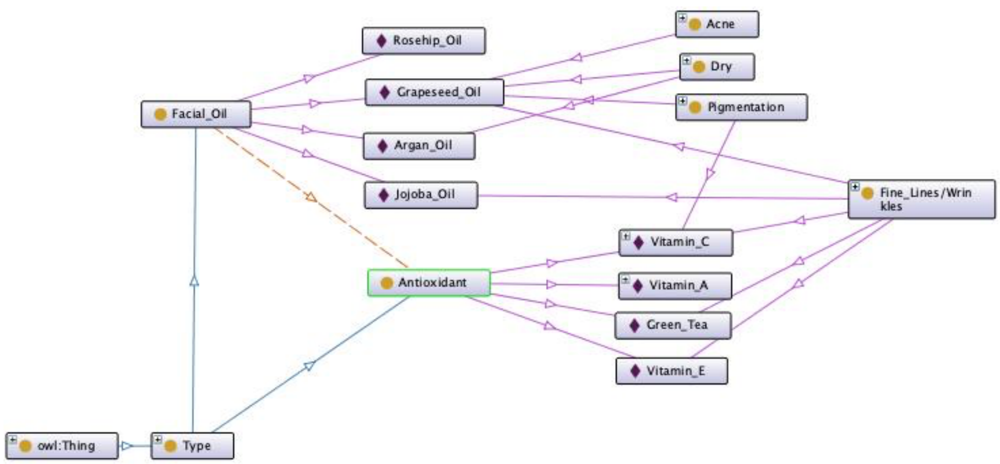


Figure 3 looks at the Humectant sub-class. Its instances of glycerin, hyaluronic acid, and aloe vera all have a relationship with the skin concern sub-class Dry. This visual shows the relationship between humectants and dry skin.

Figure 5. Protege ontology facial oil and antioxidant sub-classes.



Looking at Facial Oil sub-class in Figure 4, there is a relationship between facial oils containing antioxidants. According to the documents, there are no specific skin concerns all facial oils address. Grapeseed Oil seems to be the most diverse ingredient among the other facial oils, addressing four skin concerns. The Antioxidant sub-class in Figure 4 has ingredients that all address fine lines/wrinkles. It is noted that the documents did not include information on Vitamin A benefits. As one or many of these antioxidants are in facial oils, it can be assumed that all facial oils can address fine lines/wrinkles as well. However, that skin concern may not be all facial oils' most beneficial application area.

Figure 6. Protege ontology sunscreen sub-class.

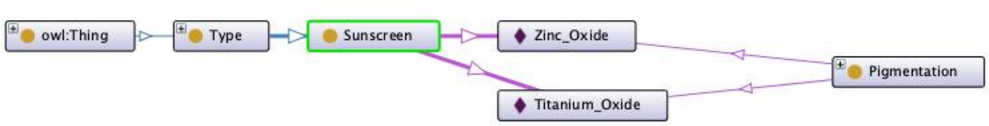


Figure 7. Protege ontology occlusive sub-class.

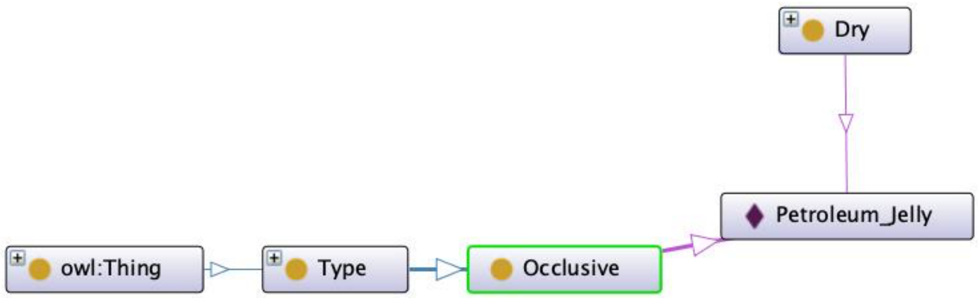


Figure 5 shows that the sunscreens instances both address pigmentation. In Figure 6, the occlusive, petroleum jelly, is identified and it addresses dry skin. As mentioned in the documents, it can be assumed that occlusives, as a sub-class, address dry skin.

Figure 8. Protege ontology exfoliator sub-class.

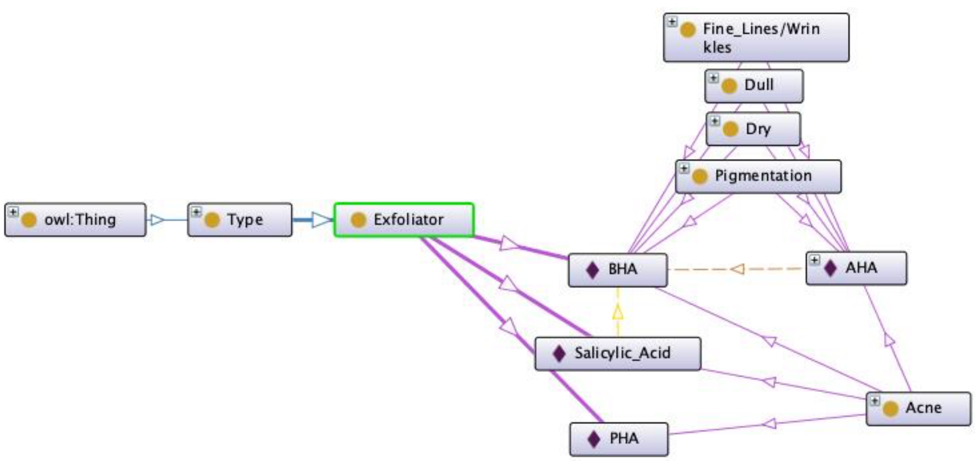


Figure 9. Protege ontology expanded AHA relationships.

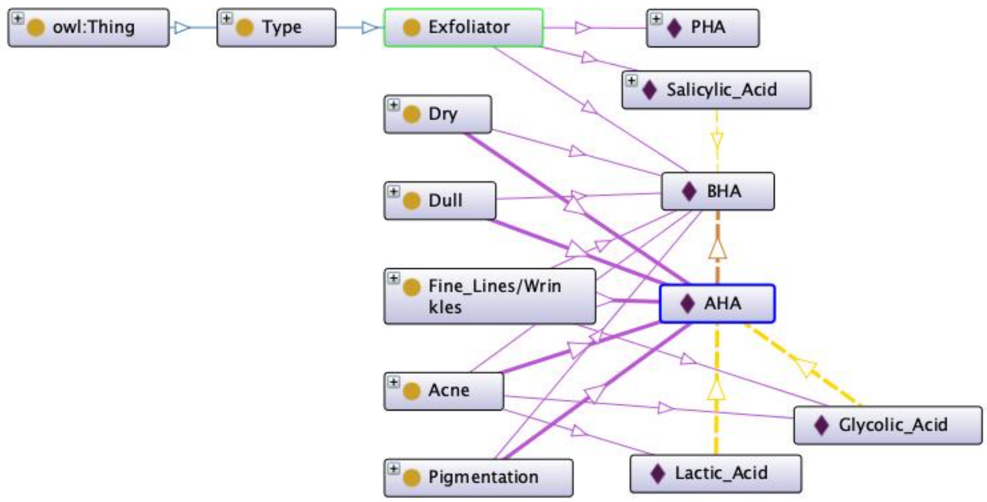
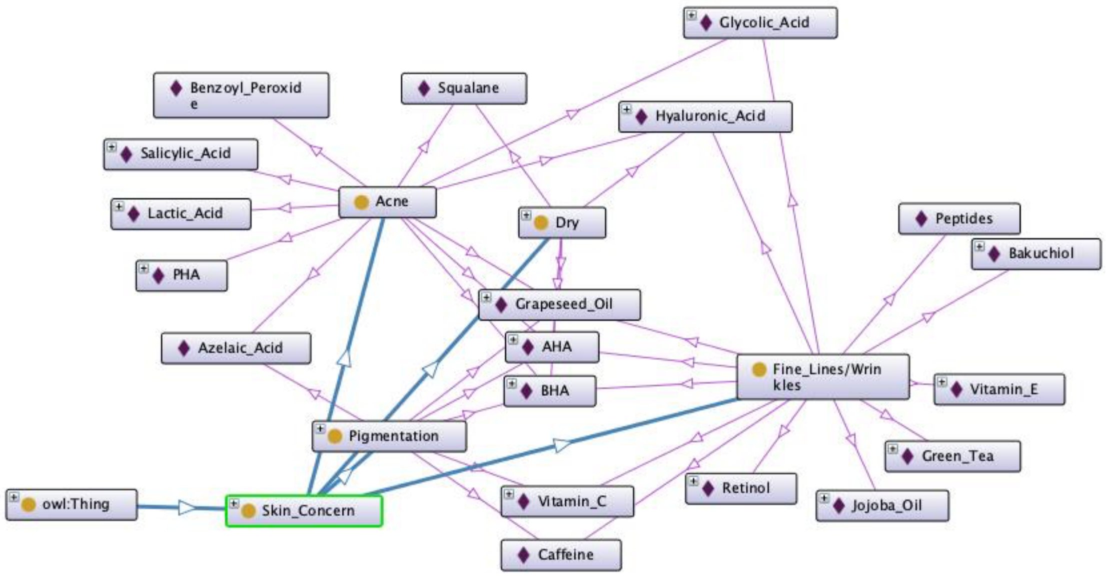


Figure 7 shows the exfoliators, BHA, Salicylic Acid, and PHA. The orange dashed line represents the substitute relationship between BHA and AHA. BHA and AHA address all the same skin concerns, so they can be substituted for one another to address those following concerns. Salicylic acid is identified as a type of BHA. Like BHA, it treats acne. All exfoliators treat acne. The skin concern of acne is most likely the most benefited area of using exfoliators. Figure 8 shows an expanded view on AHA relationships. Lactic acid and glycolic acid are types of AHA. Glycolic acid addresses acne and fine lines/wrinkles while lactic acid addresses only acne, so AHA seems to be mainly used as an acne treatment.

Figure 10. Protege ontology acne, pigmentation, and fine lines/wrinkles sub-classes.



In Figure 9, the sub-classes with the most ingredients, acne, pigmentation, fine lines/wrinkles, dry, have three overlapping ingredients of AHA, BHA, grapeseed oil. Acne, dry, and fine lines/wrinkles have hyaluronic acid as an overlapping ingredient. The other ingredients overlap with none or one other skin concern among the four displayed.

Knowledge Graph

Entity pairs are extracted from the documents. A sample of the entity pairs is outputted. Some entity pairs such as “I” and “this” are ambiguous. There are many subjects that contain “it.” However, “it” is vague with the large number of ingredients that the subject could be referring to. The relations are also extracted to describe the relationship between the entity parings. The relations “is,” “use,” “are,” “says” are the most common. There are some relations such as “skin,” “Oxide,” “dermatologist,” that should be treated as entities rather than relations.

A data frame of the source, target, and edge tokens are used to observe the relationships between the entities. There are some pairings that do make sense such as “dark vitamin C,” “dark skin spots," and "be" where vitamin C can address dark skin spots as well as "AHA BHAs," "looking dirt build skin," and "help" where AHA and BHA can help with dirt build-up in skin.

With the dataframe that is further preprocessed, the most used relations for this dataframe are "is," "says," "are," and "helps." The entire data frame is used to create a directed graph. However, it is very noisy. A subset of the data frame is taken. The most extracted relation is "is," so a subset of the data frame with only "is" relations is used to create a knowledge graph in Figure 10. This graph shows that "it" is used as a subject three times, but as mentioned above, it is difficult to make assumptions about what "it" refers to. Most entities do not refer to any specific products, but there are some like "jojoba jojoba oil" and "soften inflammation" that do provide context into jojoba oil's uses as well as "kojic acid" and "exfoliating."

Figure 11. Knowledge graph of "is" relation subset.

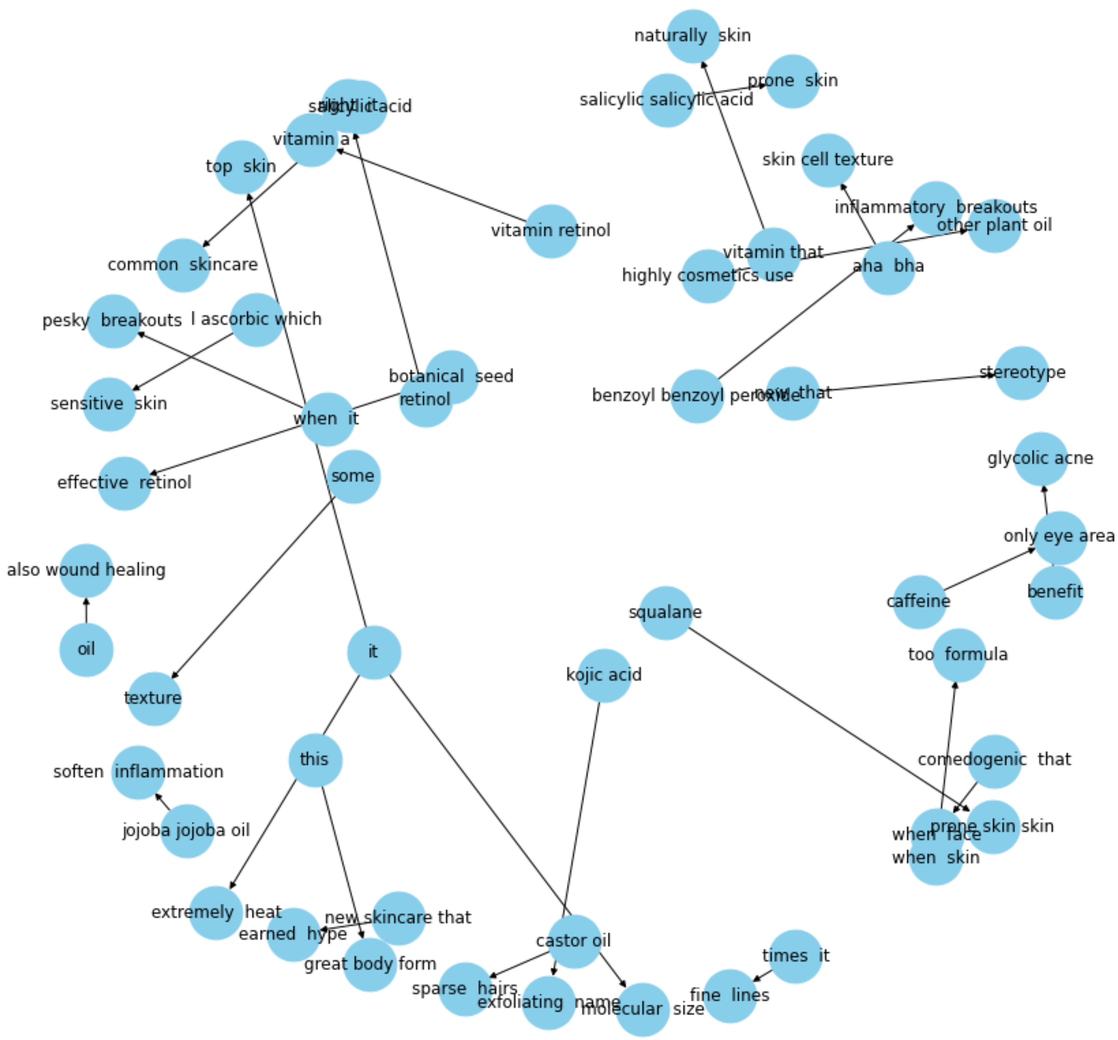
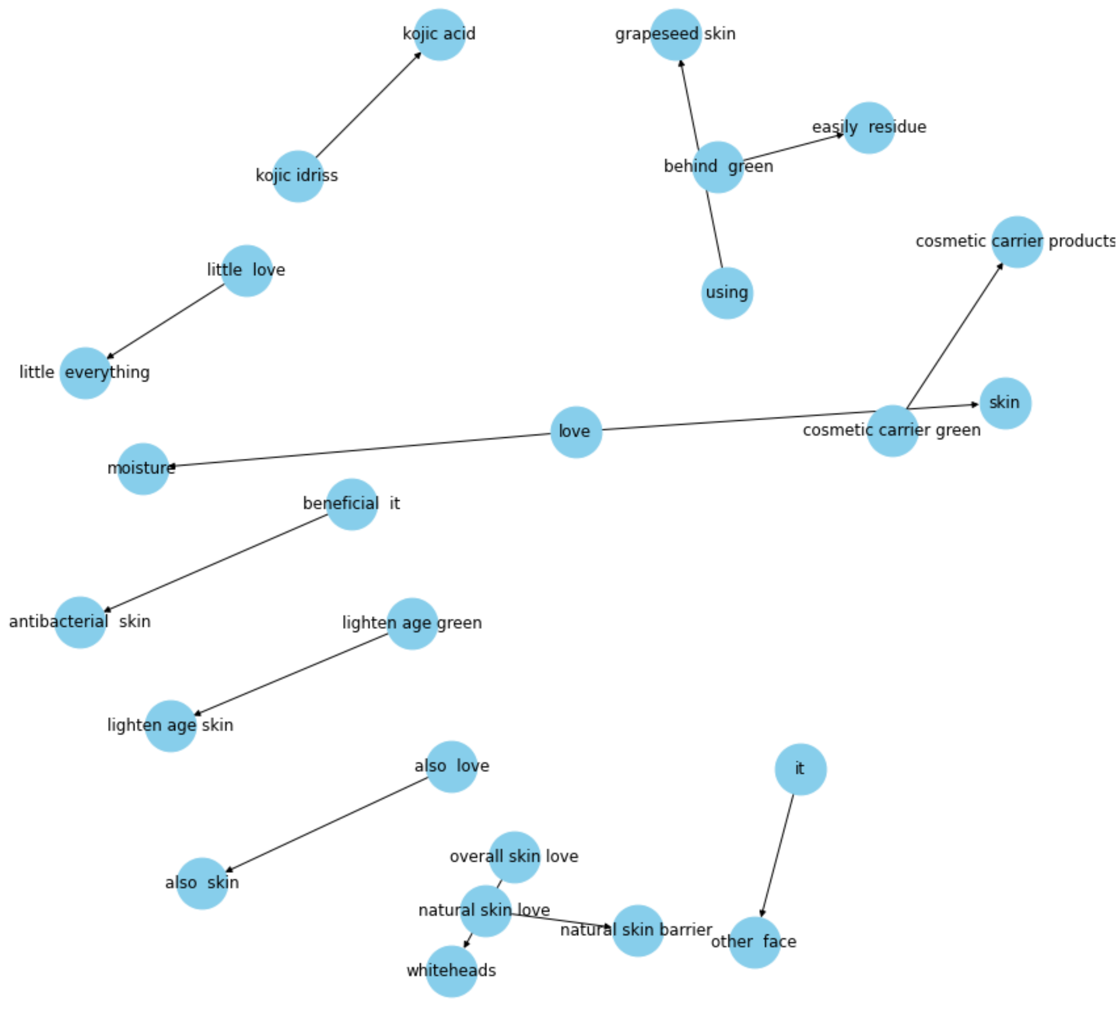
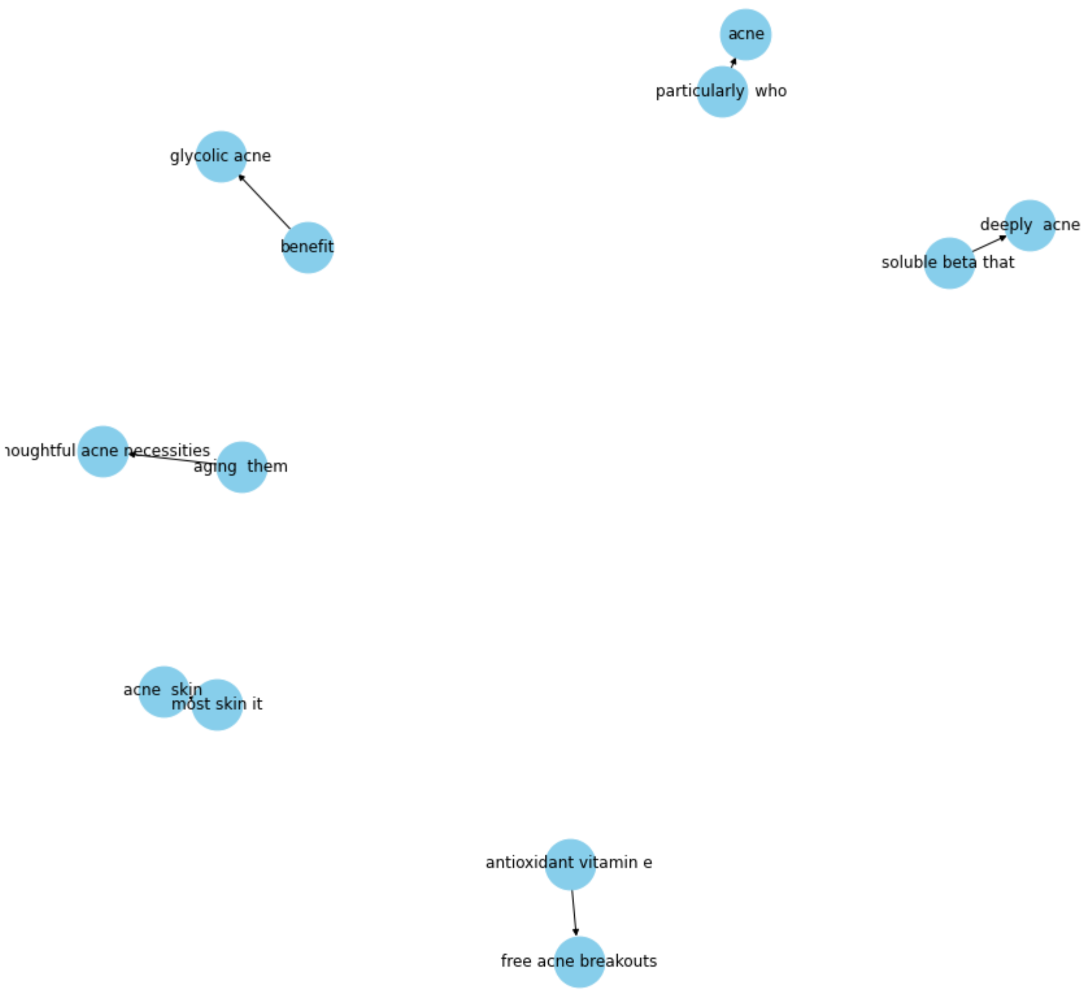


Figure 12. Knowledge graph of "says" relation subset.



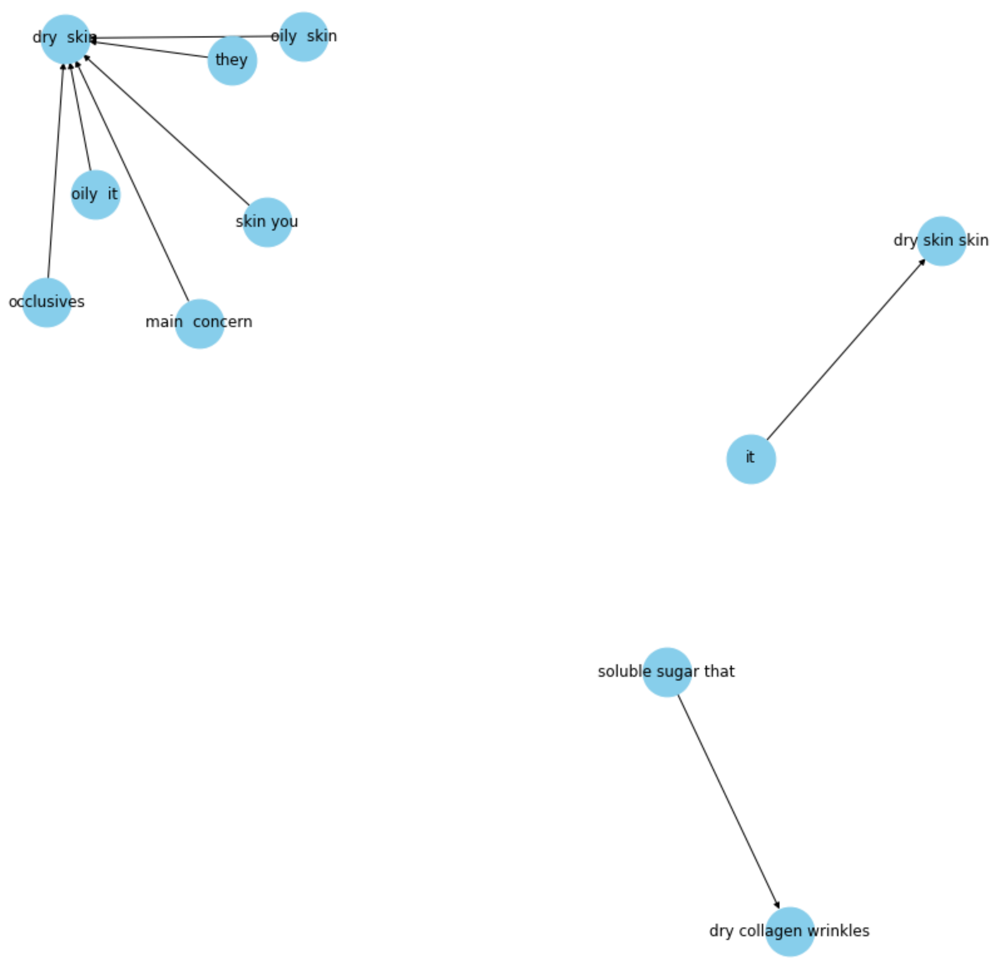
Using the relation "says," the graph in Figure 11 shows that there is a relationship between love with moisture and skin. This relationship does make sense because one of the bigger skin concern classes was dry skin, so moisture is important for the skin. This graph shows other entities that refer to "skin love" including "overall skin love" and "natural skin love." A subset using "are" as the relation is visualized. There is an entity pair that also enforces the love of moisture. There are a few entity-pairs with ingredients as its entity that gives context into a similar product as well as its uses for skin concerns. Another subset using "helps" is filtered to create a graph. This graph has one pair, “zinc” and “skin cancer,” that gives relevant context of the ingredient.

Figure 13. Knowledge graph of "acne” target subset.



A subset of target words with a partial string of “acne” was visualized into a graph, Figure 12. In this graph, there is a relationship between “antioxidant vitamin e” and “free acne breakouts.” This pair, however, does not provide the most representative context from the sentence it was extracted from. Looking back at the article, the Vitamin E is referred to in the topic of grapeseed oil, which can be used to clear up acne. Therefore, Vitamin E does help with acne despite it not being its most beneficial application which is fine lines/wrinkles, as categorized in the ontologies.

Figure 14. Knowledge graph of "dry" target subset.



In Figure 13, a knowledge graph of the subset using "dry" in the target word is visualized. The subjects related to "dry skin" show that it is a relevant skin concern category as perceived by "main concern." It also mentions "oily skin" and "oily it" which represent there is a relationship between having oily skin and having dry skin. The documents do include precautions as to which skin types are most suitable for certain products, so this may be the reason that these nodes are connected. There is also a node "occlusives" that is connected to "dry skin." From the documents and the ontologies, occlusives are ingredients to treat dry skin. Similarly, a subset where the target word contains the partial string "lines" is built as a knowledge graph. This graph shows there is a relationship between "amino peptides" and "fine lines," which enforces that peptide does treat fine lines.

Figure 15. Knowledge graph of "aha" source subset.

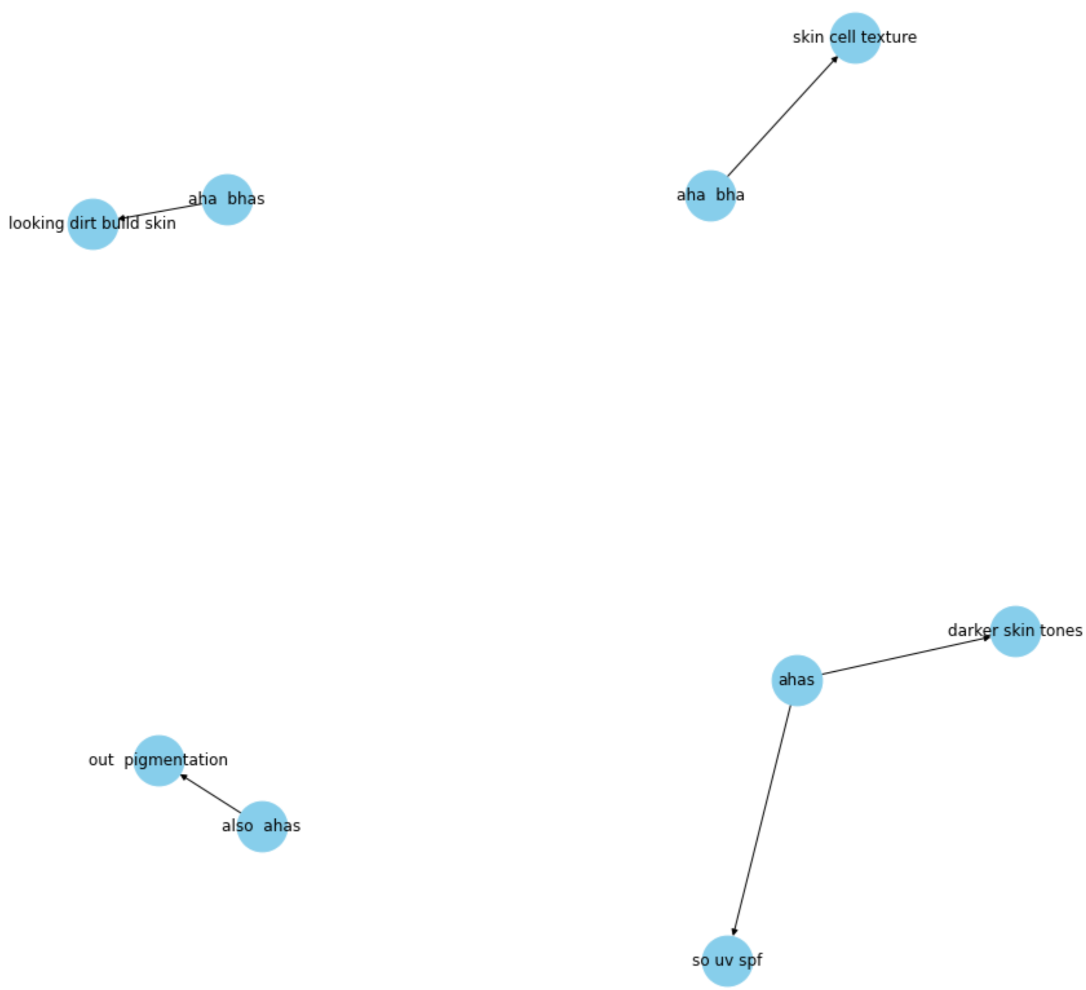
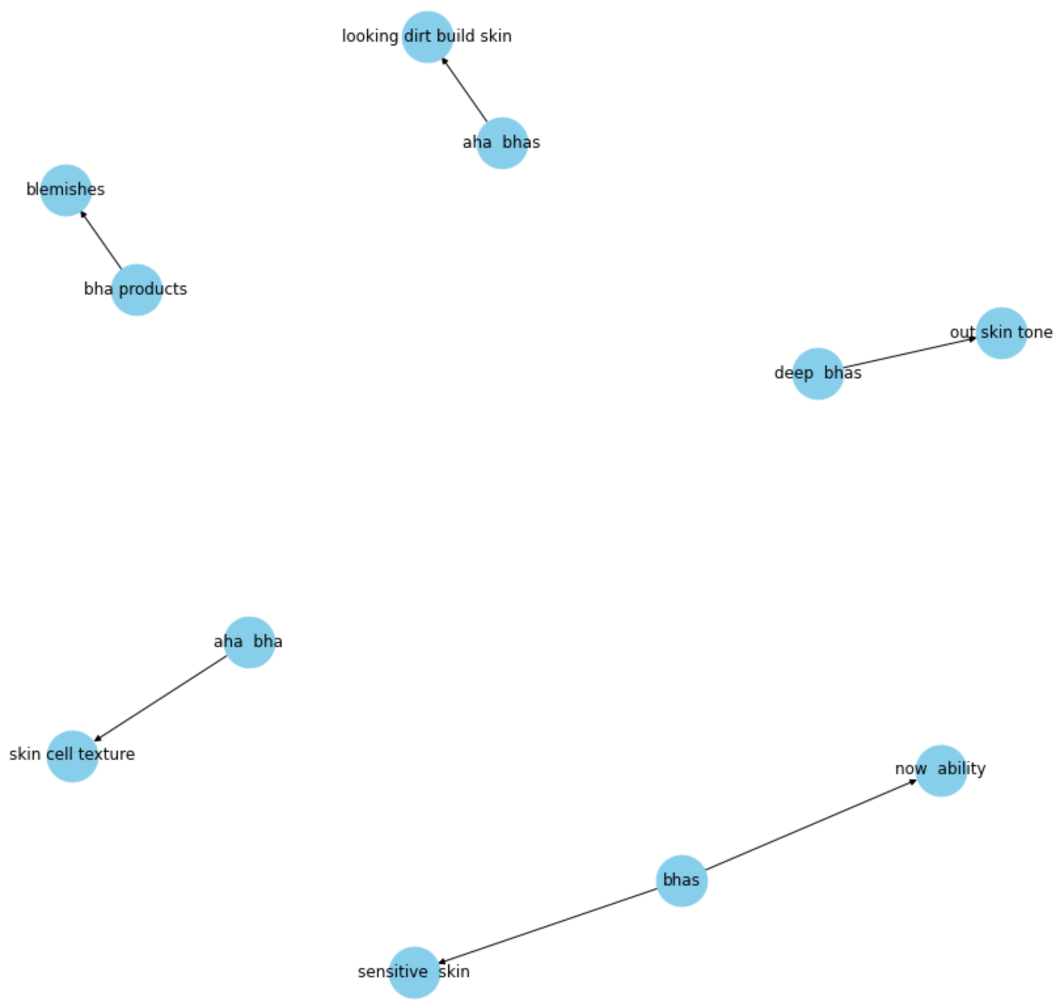


Figure 16. Knowledge graph of "bha" source subset.



A subset of entity pairs with a partial string "aha" as its source is visualized in Figure 14. The subjects often include BHA with AHA as one node, showing that these are related and are almost interchangeable. The objects are meaningful and represent what AHA targets such as dirt build-up, pigmentation or darker skin tones, and skin texture. Figure 15 shows a graph with the subset "bha" as its source. Similarly, the objects that relate to BHA offer context into its uses such as in blemishes, evening out skin tone and texture.

Building Chatbot using NLTK

The chatbot results as seen in Figure 16 show that there are formatting issues with the answers. The first letter of the answers is not capitalized and the answer for "What ingredient improves dry skin?" has quotation marks at the beginning of the sentence. In addition, the text sources had the ingredient names as headers, so the resulting answers that mention specific ingredients are not grammatically correct.

The question "What is a type of antioxidant?" was also tested as an input. The chatbot gave an out-of-context answer that did not actually answer the question. But "What are popular antioxidants?" gave a good answer. The phrasing of the question, though the same meaning, give different answers. As seen from the knowledge graphs, the text often uses "it" to refer to ingredients. Because there are so many ingredients mentioned in each article, it is difficult to know which ingredient "it" is referring to. This issue impacts the chatbot answers when asked which ingredient helps with which skin concerns. For example, when asked "What ingredient addresses both acne and pigmentation?" the chatbot answers with "it helps dry out acne, remove dead skin and oil, and even [helps] pigmentation," explains rabach." The answer includes the token "acne" and "pigmentation" but not the specific ingredient itself. Overall, the chatbot did a good job giving relevant answers to questions concerning skincare ingredients.

Building Chatbot using Keras

Experiment 1

This experiment uses a dataset of 50 instances. The LSTM layer uses 32 nodes. The training accuracy becomes 1.00 at 143 epochs while the validation accuracy only achieves 0.54 after 150 epochs. This validation accuracy is not much more accurate than randomly guessing the binary response at 0.50. The performance metric plots in Figure 18 show that the validation loss and accuracy start going in the opposite direction of the training loss and accuracy at around 100 epochs, respectively. The plots show that the model is being overfitted by the training data.

To observe the performance, one fact and question from the test set and its corresponding predicted answer from the model is viewed. The model accurately predicts the answer to the question, but the probability assigned to that answer is only 0.59. This low probability represents that the model does not discriminate between the two answers very well. A confusion matrix is visualized with the predicted and actual answers for the test set questions. The confusion matrix in Figure 19 shows that the model does a good job predicting the questions that have the answer "no" at a 0.80 classification rate. However, the model does a bad job predicting the questions that have the answer "yes" at a 0.38 classification rate.

The model is then tested using an input fact, question, and answer set where the fact is "BHA is irritating to those with sensitive skin ." and the question is "Is BHA irritating to sensitive skin ?" The model correctly predicted the answer with a probability of 0.55, which is not very high and discriminative. Another input set of "Glycolic acid is great for acneprone skin because it reduces excess oil production ." as the fact, "Does glycolic acid promote excess oil production ?" as the question, and "no" answer was used to test the model. The model predicted the answer "no" correctly at 0.65, which is not a high probability.

Experiment 2

This experiment uses a larger dataset of 95 instances. Because Experiment 1's model overfit the data, this experiment uses 25 nodes instead of 32. And the Dense layer has 650 for the smaller vocabulary. The validation accuracy is better than Experiment 1's, but is still not a good score at 0.58. The training accuracy is at 1.00, which indicates that the model overfit. The performance plot, in Figure 20, supports that there is overfitting as there is a gap between the two training and validation graphs in both loss and accuracy plots. The confusion matrix, in Figure 21, shows that the classification rates for both "yes" and "no" answers are not great. The classification rate for "yes" answers is better than Experiment 1's but is still lower than the rate of randomly guessing. The classification rate for "no" answers is worse than "Experiment 1's at 0.69, which is not a good classification rate.

A set of inputs is used to test the model performance. The fact input is "BHA is irritating to those with sensitive skin ." The corresponding question input is "Is BHA irritating to sensitive skin ?" and the answer is "yes." The model predicted incorrectly, predicting "no" with a probability of 0.79. Another set of inputs of "Glycolic acid is reduces oil production ." as the fact, "Is glycolic acid used to promote oil production ?" as the question, and "no" as the answer. The model predicted correctly with the probability of 0.81.

Conclusions

Using NLTK and keras, a chatbot was created to answer skincare questions, specifically on skincare ingredients and the skin concerns they treat. The ontology helped to identify the type of questions that can be asked using the document text. These questions included asking what a specific ingredient is, what skin concern the ingredient treats, what ingredient treats which skin concern, what group the ingredient belongs to, and what ingredient can be substituted for the other. These types of questions were successfully answered by the chatbot using NLTK. The chatbot gave relevant answers to those four types of questions. The formatting of the answer was a little off, but otherwise, did a good job.

The keras chatbot was created to answer similar skincare ingredient questions with a binary response yes/no. This model needed to be trained on a dataset. However, the dataset was not readily available, so it had to be manually constructed. The number of documents also limited the amount of facts and questions that could be constructed. The keras chatbot was created from an attention-LSTM model which gave matched words between the questions and facts a higher weight as well as memorized the short term and long term sequential patterns to give context in predicting the answer. The two models that were built did a bad job predicting the answers correctly with 0.58 accuracy for the best keras model. From the two methods, the chatbot created using NLTK yielded the best results in answering skincare ingredient related questions with relevant and correct responses. The chatbot could use some more preprocessing before implementation for formatting and performance purposes, but overall does a good job in providing quality answers.

Directions for Future Work

With the NLTK chatbot, the data could be further preprocessed and cleaned. As seen in the knowledge graphs, the source was often "it" where "it" represented a skincare ingredient but does not specify the exact one. This coreference reduces the lack of context behind the sentences. If the "it" tokens were converted to its corresponding ingredient name, the chatbot would be able to respond with sentences that are even more relevant or similar to the question. The sentence tokens can also be further preprocessed to fix the formatting of the answers. The articles often had the ingredients as headers, so that ingredient would often be in the beginning of the sentence before the actual sentence providing the answer. To give more semantic understanding to the tokens, the tokens can be vectorized with Word2Vec instead of tf-idf. This vectorization may help with computing answers that are more similar to the questions and yield more relevant answers. The keras model can be improved by using a larger corpus. A larger corpus will provide more text and facts to be used in the dataset to train the model.

References

Dass, R. (2018, September 24). *Create your chatbot using Python NLTK*. Medium. Retrieved 2022, from https://medium.com/@ritidass29/create-your-chatbot-using-python-nltk-88809fa621d1

Grant, B. (2021, July 22). *Choose Your Skincare Ingredients Based On Your Skin Concerns*. Refinery29. Retrieved 2022, from https://www.refinery29.com/en-au/best-skincare-ingredients-for-every-skin-concern

Montell, A. (2020, May 19). *The Essential Dictionary of Skincare Ingredients*. Byrdie. Retrieved 2022, from https://www.byrdie.com/skincare-ingredients-glossary

Murden, K. (2022, February 14). *An A to Z Guide to the Most Popular Skin-Care Ingredients*. Vogue. Retrieved 2022, from https://www.vogue.com/article/skin-care-ingredients-a-to-z-guide

Skin Care Products Market Size Report, 2021-2028. Grand View Research. (2021). Retrieved 2022, from https://www.grandviewresearch.com/industry-analysis/skin-care-products-market

Teich, J. (2021, October 18). *The Skincare Dictionary: Every Ingredient Explained*. Good Housekeeping. Retrieved 2022, from https://www.goodhousekeeping.com/beauty/anti-aging/a34102257/skincare-ingredient-dictionary/

z_ai. (2019, July 17). *Deep Learning for NLP: Creating a Chatbot with Keras!* Medium. Retrieved 2022, from https://towardsdatascience.com/deep-learning-for-nlp-creating-a-chatbot-with-keras-da5ca051e051

Appendix

Figure 17. Performance of models.

Experiment	Training	Validation	Process Time (seconds)	Epochs	Total Parameters	Hyperparameters
1	0.95	0.54	21.30	150	147,885	LSTM - 32 Dropout - 0.5 Dense - 666
2	1.00	0.58	25.90	150	124,550	LSTM - 25 Dropout - 0.5 Dense - 650

Figure 18. Build chatbot with keras Experiment 1 performance metrics plot.

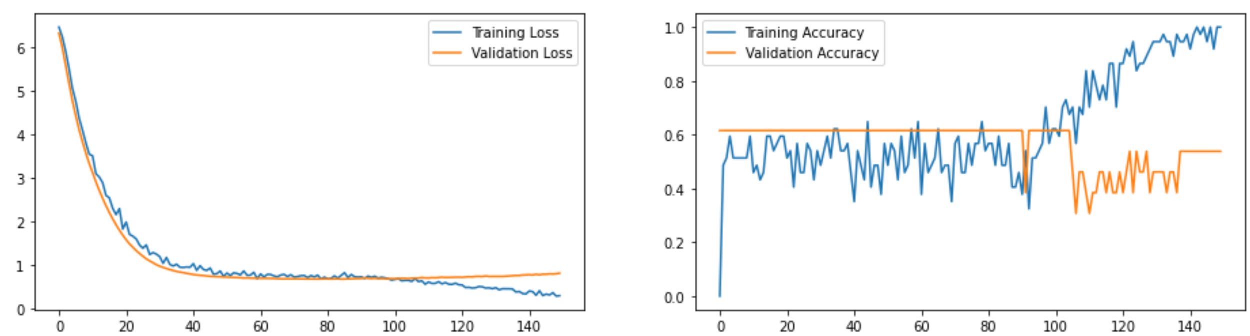


Figure 19. Build chatbot with keras Experiment 1 confusion matrix.

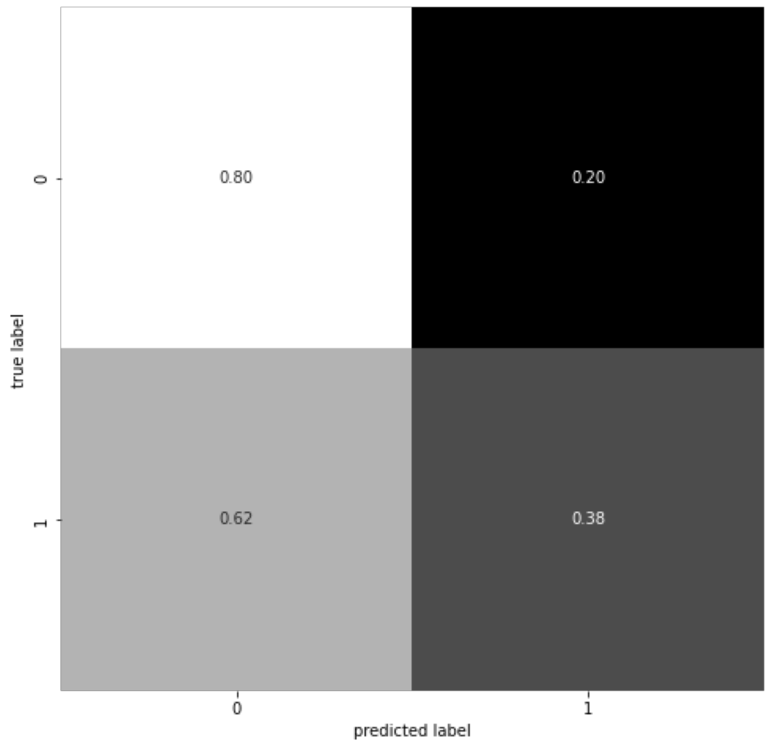


Figure 20. Build chatbot with keras Experiment 2 performance metrics plot.

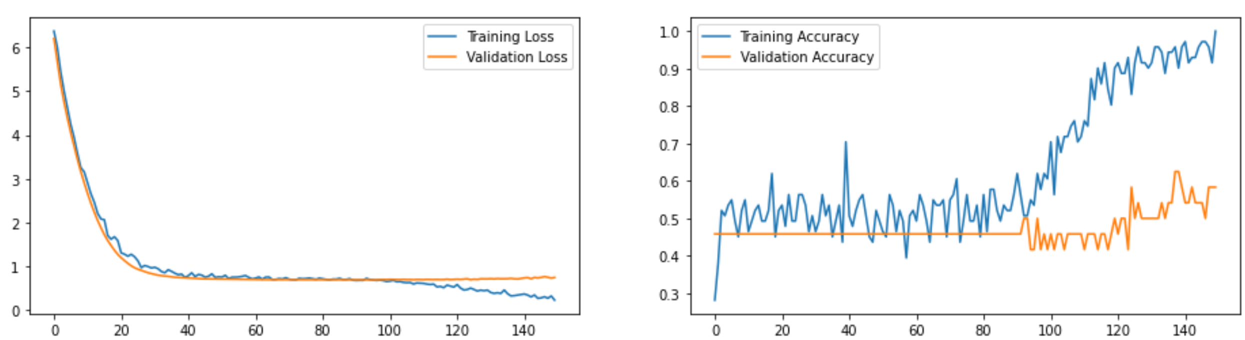


Figure 21. Build chatbot with keras Experiment 2 confusion matrix.