



 happiest minds  
The Mindful IT Company  
Born **Digital** . Born **Agile**

# NLP-Driven Qualitative Assessment

For creating content that  
clicks and converts

# Table of Contents

- ▶ Introduction
- ▶ Understand the Target Audience – Develop Customer-centric Marketing Content
  - Characteristics of Buyer Focus
  - Overview of the Algorithm
  - Sample Cue Extraction for Customer-centric Categorization
  - Present-day Limitations of the Algorithm
- ▶ What's in it for Me? Focus on Benefits
  - Characteristics of Benefits in Marketing Content
  - Overview of the Algorithm
  - Sample Cue Extraction for Benefits-related Categorization
  - Present-day Limitations of the Algorithm
- ▶ Back Your Words - Identifying Claims and Assertions
  - Characteristics of Assertions in Customer-facing Content
  - Overview of the Algorithm
  - Sample Cue Extraction for Identifying Claims and Assertions
  - Present-day Limitations of the Algorithm
- ▶ Use Words that Stimulate the Senses
  - Characteristics of Words that Tickle Human Senses
  - Repository Creation
  - Overview of the Algorithm
  - Sample Contexts
  - Sample Cue Extraction
  - Present-day Limitations of the Algorithm
- ▶ Experimental Results
  - Dataset
  - Output
- ▶ Conclusion
- ▶ References
- ▶ About the Author
- ▶ Happiest Minds Service Offerings in NLP (Natural Language Processing)
  - Content Tagging
  - Short Answer Grading
  - Sentiment Analysis
  - Content Mapping

A silhouette of a person's head and shoulders in profile, looking out of a window. The background outside the window is a bright, hazy sky. A network of thin lines and dots is overlaid on the left side of the image, extending from the person's head down towards the bottom left corner.

# Abstract

Powerful user-facing messaging can turn heads, spark conversations or lead to purchases or conversions. On the other hand, when content is not capable of engaging or appealing customers enough, it is a lost opportunity with the target audience leaving mid-way without making a purchase. User-focused content—which at times isn't given due importance—is key to generating business. Product descriptions, social media campaigns, marketing collaterals and any other form of customer-facing content are equally important.

This whitepaper recommends an automated mechanism to assess the quality of user-facing content—such as product descriptions—by quantifying the level of customer-centricity and benefits, rather than focusing on features. It also shows how Natural-Language-Processing-driven (NLP-driven) qualitative assessment may be used to verify if there are any assertions made in marketing content along with the use of keywords that entice human senses.

The approach explained in this whitepaper uses morphological, syntactic and semantic analysis techniques. Degree of customer-centricity is measured by using constituency parser and synsets from WordNet. Words that please the senses are traced using pre-trained BERT. Benefits are evaluated using Open Information Extraction (OpenIE) and coreference resolution. Verification of assertions in text is done using POS tagging and pertainyms in WordNet. The performance of the framework is evaluated for user-facing marketing content available online. The approach is shown to achieve acceptable levels of accuracy.



# Introduction

Customer-facing marketing material usually apprises a prospect about a company's products or services and their benefits. Persuasive content can lead to conversion while feeble content can steer a prospect toward competitors. This makes marketing collaterals important elements of an organization's success. Key parameters for evaluating the maturity of customer-facing content are:



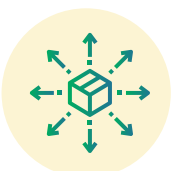
**Understand the target audience** - Specificity in language is achieved by keeping the ideal buyer in mind. For example, *"From Brand A's Honeydew collection, this shopping cart helps children advance their social skills through inventive play."* is a description of a toy shopping cart and is focused on children.



**What's in it for me?** It is important to focus on the performance aspect of products or services, highlighting benefits. For example, *"This Thermal Razor by Brand B provides heat instantly at the push of a button, giving you a comfortable shave."* is a description of a razor. It clearly explains how the razor enables a comfortable shave.



**Back your words** - Assertions should always be backed by evidence. This will also help stay away from legal hassles. For example, a mobile device provider's website reads: *"This device looks spectacular and comes in two of the most loved color variants: gold and black. Brand C always focuses on product quality and delivers the best after-sales service."* Though the website assures the "best" after-sales service, there's no concrete proof supporting the claim, and such statements should be avoided.



**Use words that stimulate the senses** – Enticing the senses by using the right adverbs and adjectives in marketing content often results in purchases. For example, the description *"Choose from our frozen desserts featuring rich ganache, velvety caramel, luscious praline, refreshing fruit, and crunchy nuts."* makes use of multiple words to elevate the content. However, that doesn't mean content should be saturated with adjectives and adverbs just to paint a colorful picture of the company's products.

NLP-driven qualitative assessment can help organizations across verticals and horizontals comprehend the quality of their sales and marketing content published across various platforms. The output of the analysis can be used to offer valuable feedback to content writers and marketing professionals, who in turn can help enhance content quality and overall sales of the organization.

# Understand the Target Audience – Develop Customer-centric Marketing Content

## Characteristics of Buyer Focus

User-facing marketing content can be made customer-centric by targeting a set of people who are potential buyers of the product. For example, the sentence *“Since ages, Indian mothers have been realizing that Brand C Drink helps choosy toddlers get the required nutrition to grow taller and gain weight.”* has been used to attract parents and children. Another example would be: *“Our fully breathable mattress with feather-soft material keeps parents happy and children comfortable.”* to attract parents.

The words “mothers”, “toddlers”, “parents” and “children” are **common nouns** and represent **groups of people**. “Mothers” is used as the **subject** of the sentence. The algorithm to identify customer-centricity should be aimed at determining these unique features.

## Overview of the Algorithm

The recommended algorithm for identifying customer-centric cues in a sentence is shown in Figure 1 and important steps of the algorithm are explained below.

**Tokenizing Content** – Sentence tokenizer available in the Natural Language Toolkit (NLTK)<sup>11</sup> is used to extract sentences from the overall content.

**Identifying Subject Phrase Contenders** – For each of the tokenized sentences, subject phrase contenders are identified by making use of constituency parser<sup>3,4</sup> available in Stanford CoreNLP. Subject phrase extraction is strictly based on the definition of a subject from the NLTK<sup>11</sup> website – One common way of defining the subject of a sentence *S* in English is as the noun phrase that is the child of *S* and the sibling of verb phrase. An example output of the constituency parser<sup>3,4</sup> is shown in Figure 2. In Figure 2, “Indian mothers” is: a) the child of sentence identified by “*S*”, b) a noun phrase identified by “*NP*”, and, c) the sibling of a verb phrase identified by “*VP*”. Thus, “Indian Mothers” is a contender subject phrase of this sentence. Similarly, “Brand C Drink” is another subject phrase contender.



**Identifying the Subject Phrase** – Out of all the identified subject phrase contenders, the right subject phrase is identified using the depth component of phrases in the sentence. The phrase with the lowest depth, that is, the phrase that is shown topmost in the constituency parser output is considered as the subject of the sentence. For example, “Indian Mothers” is the subject of sentence since it is the topmost contender shown in Figure 2.

**Identifying Subject Contenders** – For each of the tokenized sentences, all subjects that fall under categories mentioned in the Stanford typed dependencies manual are noted. Some example categories are ‘subj’, ‘nsubj’, ‘csubj’. This is done by making use of the dependency parser <sup>3,12</sup> available in Stanford CoreNLP <sup>3</sup>. A sample output from the dependency parser is shown in Figure 3.

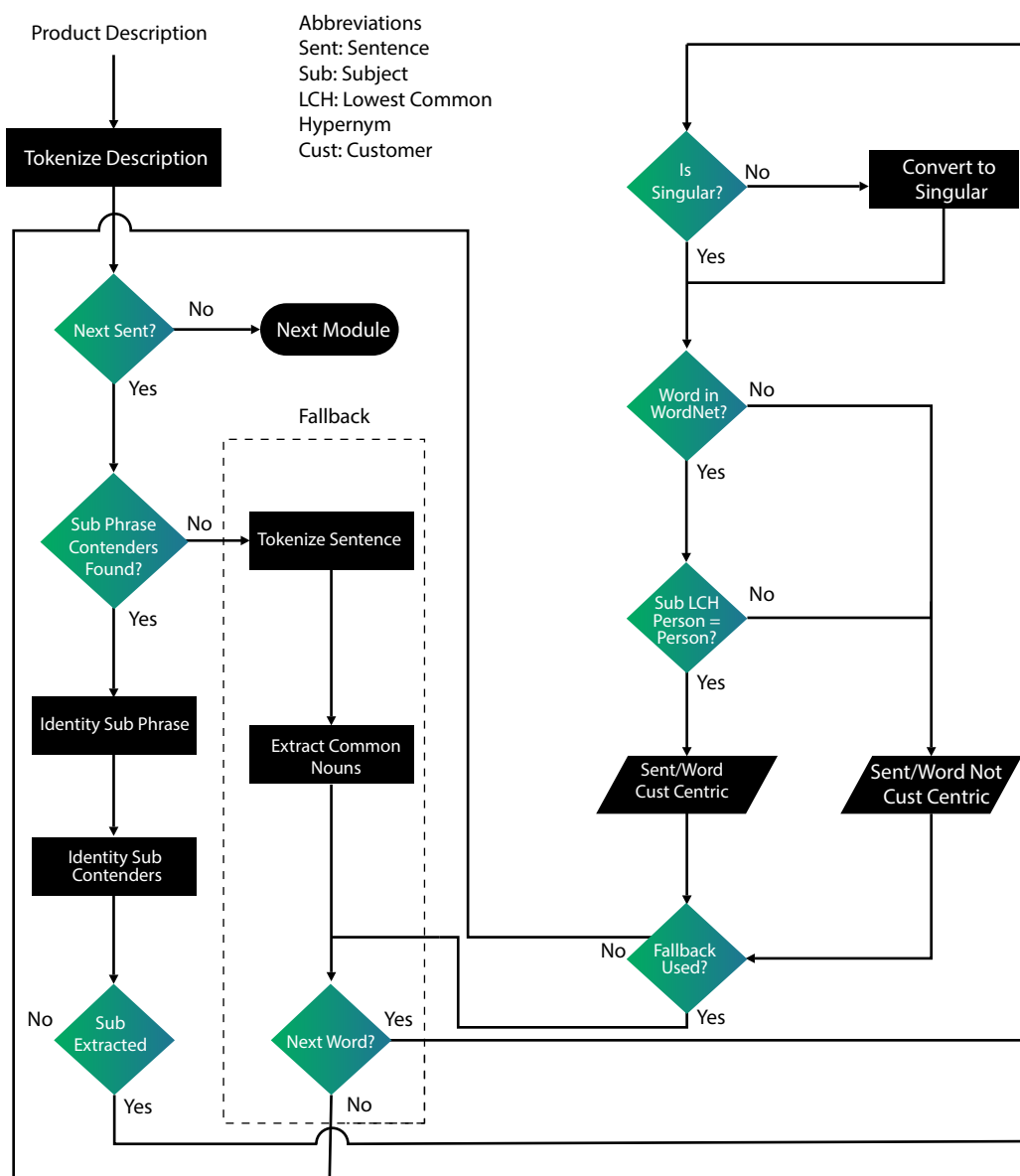


Figure 1. Proposed algorithm for identifying buyer centric cues in sentences

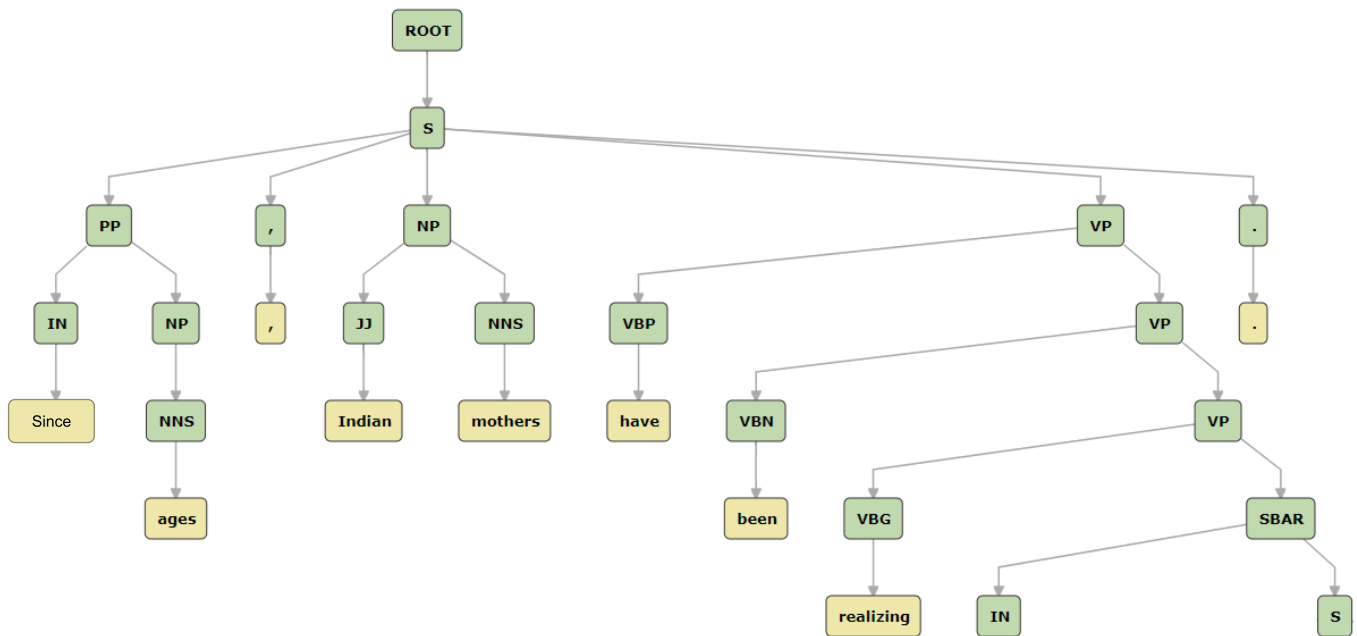


Figure 2. A sample output from Constituency Parser available at <https://corenlp.run>

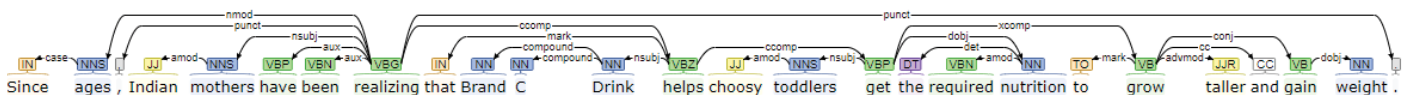


Figure 3. Sample output from Dependency Parser available at <https://corenlp.run>

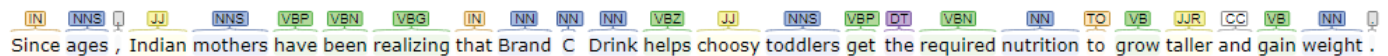


Figure 4. Sample output from Part of Speech Tagger available at <https://corenlp.run>

**Subject Extraction** –Next, the subject of the sentence is calculated through an intersection of the identified subject phrase and subject contenders.

Subject Phrase n Subject Contenders

**Convert Subject to Singular** – In case the subject is in plural form, it will have to be converted to singular form, and this can be achieved using the Inflect<sup>13</sup> library. This step is essential as the Subject Extraction phase in this process is dependent on the WordNet<sup>1</sup> corpus that contains the singular form for entity definitions.

**Lowest Common Hypernym** – Using the WordNet<sup>1</sup> corpus available in the NLTK library<sup>11</sup>, find out the lowest common hypernym between the subject and “person”. Hypernyms provide a way to categorize words based on their similarity to each other. If the lowest common hypernym in the subject and “person” is “person”, it means the subject represents “person”, and hence, the word is classified as a customer-centric keyword in the product description.

**Fallback Branch** – In case there isn’t a valid subject phrase contender identified through the constituency parser, a fallback branch is identified by tokenization<sup>3,10</sup> and common noun extraction. The sentence is considered buyer-focused if one of the common nouns identified in the fallback branch shares the lowest common hypernym with “person”.

## Sample Cue Extraction for Customer-centric Categorization

An example to explain customer-centric categorization for a sample product description is given in Table 1.

**Sample Content** - Since ages, Indian mothers have been realizing that Brand C Drink helps choosy toddlers get the required nutrition to grow taller and gain weight.

Step	Description	Output
1.1	Tokenization Product Description	Since ages, Indian mothers have been realizing that Brand C Drink helps choosy toddlers get the required nutrition to grow taller and gain weight.
1.2	Identify Subject Phrase Contenders	Indian Mothers, Brand C Drink
1.3	Identify Subject	Indian Mothers
1.4	Identify Subject Contenders	Mothers, Drink
1.5	Subject Extraction	Mothers
1.6	Convert Subject to Singular	Mother
1.7	Lowest Common Hypernym	Mother

Table 1: Sample of Customer-centric Categorization

## Present-day Limitations of the Algorithm

**Missing Words in WordNet<sup>1</sup>**– Customer-centric words can be identified only when they are present in the WordNet<sup>1</sup> corpus.

Consider the example, *“Your pulmonologist prescribed Brand D, but your pharmacy may suggest switching to a generic inhaler. Are you sure it contains the same medicine as in the Brand D inhaler?”* is the description of an inhaler. However, the word “pulmonologist” is missing in WordNet<sup>1</sup>.

**Missing Context in WordNet<sup>1</sup>**– Not just words, but certain contexts of some words too could be missing in WordNet<sup>1</sup>, increasing false negatives.

Consider the example, *“The Brand E patch repairer will ensure your lawn looks its best.”* Though this description refers to the patch repairer product, in WordNet<sup>1</sup>, the word “repairer” refers to a person and not a product.



# What's in it for Me?

## Focus on Benefits

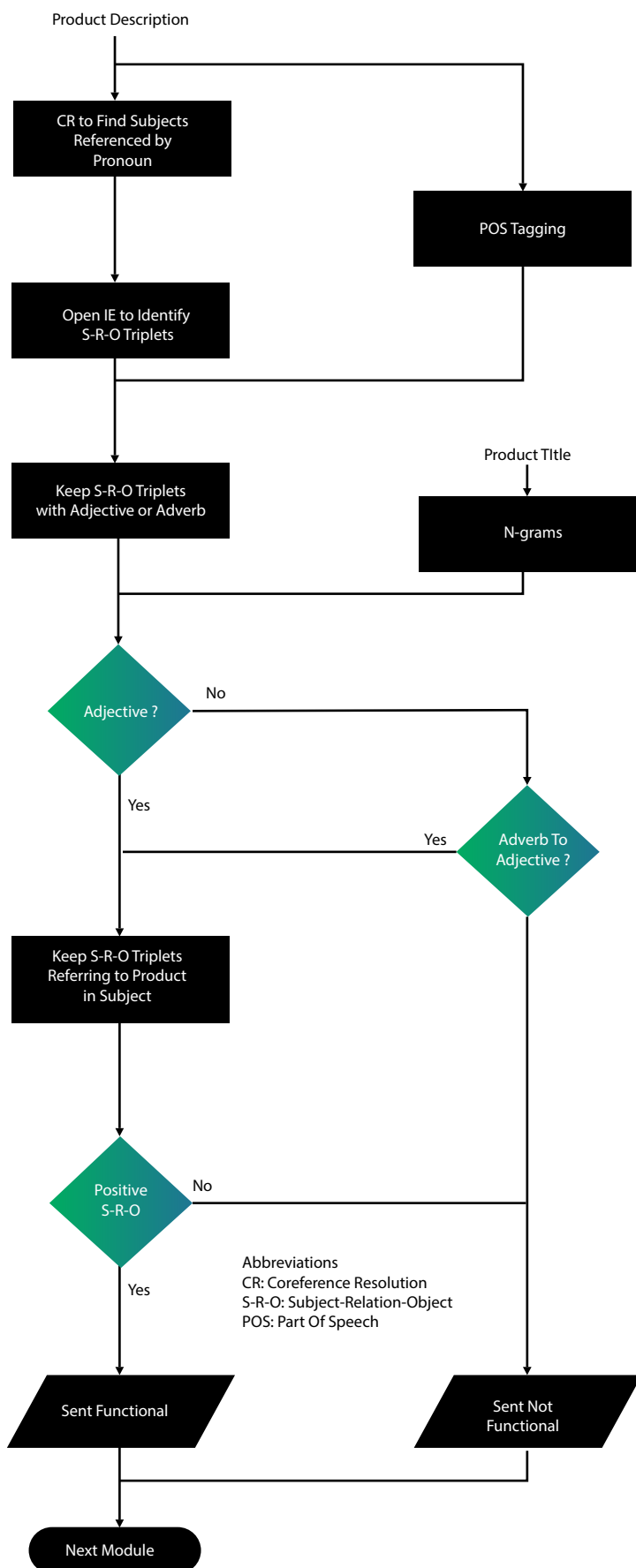


Figure 5. Proposed algorithm for identifying benefits-related cues in sentences

## Characteristics of Benefits in Marketing Content

Focus on the performance of the product or service by clearly mentioning benefits rather than focusing on features alone.

For example, “*Brand E whole wheat flour is hygienically ground using a modern mill-grinding method. Brand E whole wheat flour contains 100% pure wheat flour and 0% maida.*” describes the brand’s whole wheat flour that explains the hygienic and modern grinding process.

The words “modern” and “hygienically” are added to highlight the benefits of the product.

## Overview of the Algorithm

The proposed algorithm for identifying benefit-related cues in the sentence is shown in Figure 5. Important steps of the algorithm are:

**Coreference Resolution**<sup>3, 6-9</sup> – In marketing material that comprises over a single sentence, a pronoun refers to a noun used earlier in the same document. Stanford CoreNLP<sup>3</sup> is used to find such pronouns and replace them with corresponding nouns. An example is shown in Figure 6.

**Tokenize Description** – Sentence tokenizer available in the NLTK<sup>11</sup> library is used to extract sentences out of the user-facing content.

**Open Information Extraction** – Next, Subject–Relation–Object (S-R-O) triplets are extracted from each of the sentences tokenized in the earlier step. An example output of OpenIE<sup>3,5</sup> from Stanford CoreNLP<sup>3</sup> is shown in Figure 7.

**POS Tagging** – Part-of-speech tagging<sup>3,10</sup> is done for all words in the input text using Stanford CoreNLP Parser<sup>3</sup>.

**Adjectival/Adverbial Triplets** – Only those S-R-O triplets that have an adjective or adverb are retained.

**N-grams** – N-grams are extracted from the title using the NLTK<sup>11</sup> library. This is done to ensure only the S-R-O relations that refer to the input product in their subject are retained.

**Adverb to Adjective** – Only those adverbs that have a corresponding adjective are retained using pertainyms in WordNet<sup>1</sup>. These adverbs are known as “adverbs of manner”.

**Product-only Triplets** – Next, using the extracted N-grams, triplets that have fragments of the title in their subject part of the relation are identified, and only such S-R-O relations are considered for the next steps.

**Sentiment Analysis** – Concatenate the S-R-O triplets to form phrases and analyze sentiments in these phrases using Text Blob<sup>14</sup>. Phrases with positive sentiments are considered to be describing product or service performance.

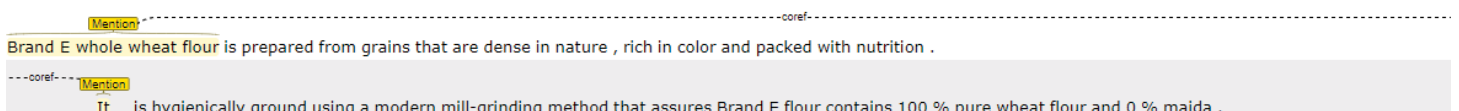


Figure 6. A sample output from Coreference Resolution available at <https://corenlp.run>

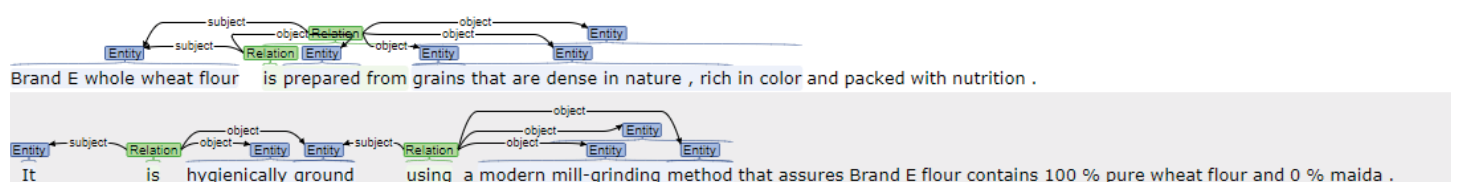


Figure 7. A sample output from OPEN IE available at <https://corenlp.run>

# Sample Cue Extraction for Benefits-related Categorization

An example to explain benefits-related categorization for a sample product description is given in Table 2.

**Sample Content** - Brand E whole wheat flour is prepared from grains that are dense in nature, rich in color and packed with nutrition. It is hygienically ground using a modern mill-grinding method that assures Brand E flour contains 100% pure wheat flour and 0% maida.

Step	Description	Output
1.1	Coreference Resolution	Brand E whole wheat flour is prepared from grains that are dense in nature, rich in color and packed with nutrition. Brand E whole wheat flour is hygienically ground using a modern mill-grinding method that assures Brand E flour contains 100% pure wheat flour and 0% maida.
1.2	Tokenize Description	Brand E whole wheat flour is hygienically ground using a modern mill-grinding method that assures Brand E flour contains 100% pure wheat flour and 0% maida.
1.3	POS Tagging	Flour (NN), hygienically (RB), Modern (JJ), etc.
1.4	Open IE	(ground, using, modern mill-grinding method), (Brand E whole wheat flour, is, hygienically ground)
1.5	Adjective Triplets	(ground, using, modern mill-grinding method)
1.6	Adverbial Triplets	(Brand E whole wheat flour, is, hygienically ground)
1.7	Adverbs of Manner	(Brand E whole wheat flour, is, hygienically ground)
1.8	Product-only Triplets	(Brand E whole wheat flour, is, hygienically ground)
1.9	Sentiment Analysis	Slightly positive

Table 2: Sample of benefits-related categorization

## Present-day Limitations of the Algorithm

**Indirect Benefits Cannot Be Identified** – Sentences which do not have the product in the Subject part of S-R-O triplets are avoided. Hence, indirect explanations such as this cannot be identified with the algorithm: “Dough prepared from Brand E flour soaks up more water, and hence, flatbreads stay fresh for longer.”

**Syntactically Incorrect and Incomplete Sentences Remain Unaccounted** – Only complete sentences are considered. Text which does not qualify to be a complete sentence does not work well with OpenIE<sup>3,5</sup>, and is avoided. As an example, bullet points that do not contain complete sentences are not taken into account. Phrases in advertisements such as “Organically grown without pesticides and fertilizers” are not considered.

## Back Your Words - Identifying Claims and Assertions

### Characteristics of Assertions in Customer-facing Content

Customer-facing content can contain positive assertions made about products, offerings or services. For example, the content “Brand C always focuses on product quality and delivers the best after-sales service.” asserts that the product delivers the best post-sale service.

### Overview of the Algorithm

The algorithm for identifying assertion cues in the sentence is shown in Figure 8, and the important steps of the algorithm include:

**Tokenizing Content** – The sentence tokenizer available in the NLTK<sup>11</sup> library can be used to extract sentences out of marketing content.

**POS Tagging** – Part-of-speech tagging is done for all words in the input text using Stanford CoreNLP Parser<sup>3</sup>.

**Identifying Superlative Adjectives and Adverbs** – All text, except words identified as superlatives while performing POS tagging<sup>11</sup>, is removed. Superlative adjectives are identified by ‘JJS’ and superlative adverbs are identified by ‘RBS’.

**Identify Superlative Adjective** – Superlative adjectives represent an assertion made in the creative customer-facing content—in product descriptions, ad copies, etc. The word “best” is a popular example.

**Separate Adverbs from Adjectives** – Only “adverbs of manner” that have a corresponding adjective and are generally representative of a quality (bad or good) are retained. This step is executed using pertainyms in WordNet<sup>1</sup>.



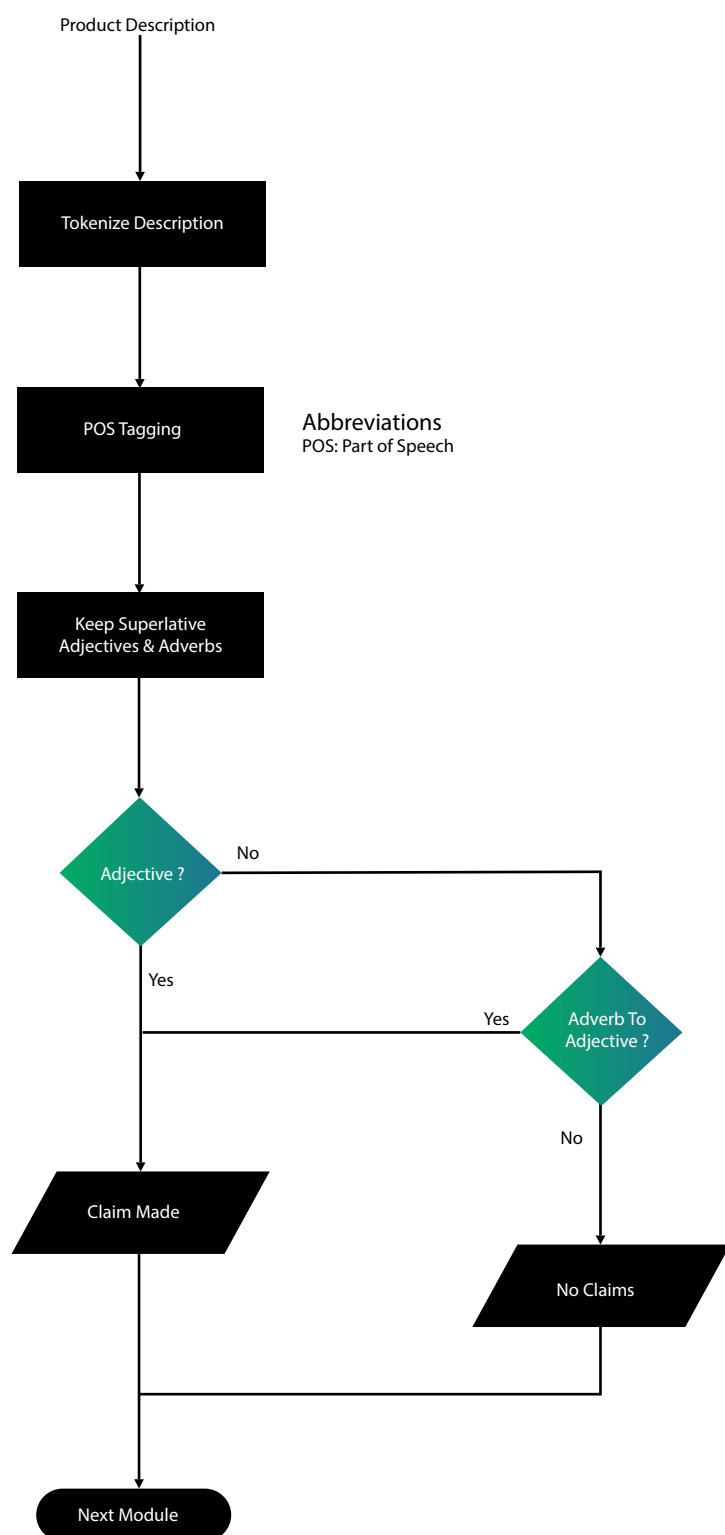


Figure 8: Proposed algorithm for identifying assertion related cues in sentences.

## Sample Cue Extraction for Identifying Claims and Assertions

An example to explain assertions in marketing material is given in Table 3.

**Sample Content** – This device looks spectacular and comes in two of the most loved color variants: gold and black. Brand C always focuses on product quality and delivers the best after-sales service.

## Present-day Limitations of the Algorithm

**Identifying Assertions that are Not Superlatives** – The algorithm does not flag claims and assertions that do not directly indicate the highest degree of quality or attributes as superlatives. Work is in progress to expand the algorithm to include such assertions. As an example, *“Brand E whole wheat flour was launched on January 16, 2000, and has now become the number one in branded packaged wheat flour across the continent.”* Here, the algorithm does not identify the superlative “number one”.

**Verifying the Presence of Supporting Elements** – There could be assertions made in customer-facing material that are not supported by evidence. Work is in progress to expand the current approach to identify proofs of submitted assertions and flag the rest.

Step	Description	Output
1.1	Tokenize Description	Brand C always focuses on product quality and delivers the best after-sales service.
1.2	POS Tagging	Always (RB), Best (IJS), Deliver (VB), etc.
1.3	Superlative Adjective and Adverb	Best
1.4	Adverb to Adjective	NA

Table 3: Sample of assertions-related categorization

## Use Words that Stimulate the Senses

**Characteristics of Words that Tickle Human Senses** – User-facing content works great when it contains words that entice human senses. For example, the ad copy “Our curry paste contains a hand-picked combination of aromatic chili peppers, natural sweeteners, black truffle, and spicy flavors.” contains words such as “spicy” and “chili” to stimulate the human senses.

**Repository Creation** – A repository of such words and their corresponding senses and sentences has been developed for the purpose, and this repository acts as the basis of the algorithm. For example, the word ‘bitter’ can refer to the sense of taste, and it can also describe relationships. These are termed as two senses of the word “bitter”. The relevant sense is identified by understanding the context in which the word is used in the sentence.

## Overview of the Algorithm

The proposed algorithm for identifying sense-related cues in a sentence is shown in Figure 9. Important steps of the algorithm include:

**Tokenizing Content** – Sentence tokenizer available in NLTK<sup>11</sup> is used to extract sentences out of content.

**POS Tagging** – Part-of-speech tagging<sup>3,10</sup> is done for all words in the input text using Stanford CoreNLP Parser<sup>3</sup>.

**Lemmatization** – The extracted verbs and adverbs are converted into the base form of the respective words using NLTK<sup>11</sup>.

**Extracting Common Words** – Next, common words are extracted by intersecting the words from the repository and the lemmatized output from the previous step.

$$\text{Repository} \cap \text{Lemmatized O/P}$$

**Closest Context** – Context of the input words are verified using the pre-trained BERT<sup>2</sup> model and closeness of inputted words with all contexts present in the repository is calculated. The context sense from the repository that is closest to the input user-facing content is considered as the context.

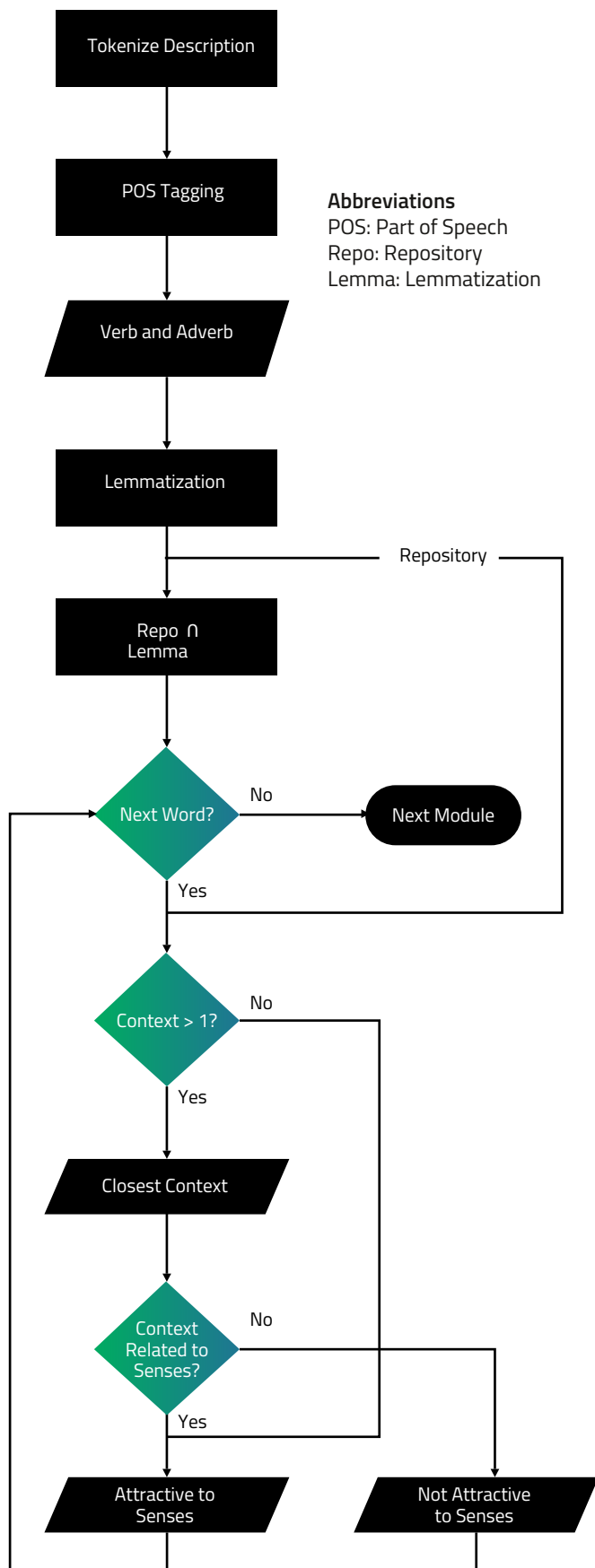


Figure 9. Proposed algorithm for identifying sense-related cues in sentences

## Sample Contexts

Sample contexts from the repository for the word “bitter” are shown in Table 4.

## Sample Cue Extraction

An example to explain this category is given in Table 5.

**Sample Product Description** – Fine dark chocolate with a pinch of sea salt! Splendidly silky, deliciously rich Brand G dark chocolate with a pinch of sea salt - Dark chocolate bites with a pinch of hand-harvested sea salt.

## Present-day Limitations of the Algorithm

**Limited Words/Contexts in the Repository** – The repository has been created manually from WordNet and is limited to my understanding. Consider, “Brand H introduces to you the velvety and intense experience of bittersweet dark chocolate.” In this example, though the word “intense” describes taste, it is not associated with taste anywhere in WordNet. Hence, the same result reflects in the repository developed for the algorithm.

**Sentence Formation in the Repository** – Sentences created in the repository must be generic and sufficient in length to create a generic context and this can be a laborious process. For example, “Wow! That’s tasty.” isn’t a good candidate to be entered in the current version of the repository for the word “tasty”.

Context	Sentence
Taste	It was far too sweet and had a bitter aftertaste.
Resentment	His ecclesiastical legislation, too, met with bitter opposition from the Church.
Harsh tone	Despite the bitter words, there was regret on Wynn's face.
Painful	Winters are quite bitter in the Canadian region.
Cynicism	She felt extremely bitter about the divorce.
Difficult to bear	She went into bitter sorrow after her husband's death.

Table 4: Sample contexts for the word “bitter”

Step	Description	Output
1.1	Tokenize Description	Fine dark chocolate with a pinch of sea salt! Splendidly silky, deliciously rich Brand G dark chocolate with a pinch of sea salt - Dark chocolate bites with a pinch of hand-harvested sea salt.
1.2	POS Tagging	Fine (JJ), Deliciously (RB), Pinch (NN), etc.
1.3	Verb and Adverb	Deliciously
1.4	Lemmatization	Delicious
1.5	Repository $\cap$ Lemmatized O/P	Delicious
1.6	Contexts	Entertaining, interested, taste
1.7	Closest Context	Taste

Table 5: Sample of sense-attractive categorization

# Experimental Results

## Dataset

Dataset 1

We extracted product titles and product descriptions of chocolates by top brands available on different e-commerce web sites.

Dataset 2

We extracted product titles and product descriptions of baby foods by various brands available on multiple e-commerce web sites.

## Output

The output is detailed in Table 6.

Category	Output
Buyer-centric	602 out of 1336 chocolate-related descriptions.
Benefits-related	723 out of 5676 baby food-related descriptions.
Assertions	200 out of 1336 chocolate-related descriptions.
Attractive to the senses	206 out of 1336 chocolate-related descriptions.

Table 6: Output for the algorithm



**Performance of the Algorithm** – The dataset at hand is unsupervised. 100 descriptions were manually labelled based on my understanding, and an average accuracy of 73% was noticed on the overall dataset.

## Conclusion

User-facing content such as marketing content in product descriptions, brochures, flyers and websites appraises prospects about the organization's products, services and their benefits. A compelling, customer-centric copy with an appropriate focus on: benefits, solid claims or assertions backed with proof, and words that stimulate the senses can result in increased conversions. At the same time, content that doesn't include any of these essential features can shift the prospects' attention toward competitors. Marketing professionals, creative copywriters and sales executives can leverage the automated mechanism detailed in this whitepaper to assess the quality of their marketing content by scoring the usage of personal, sensorial, functional and superlative words and phrases in their material. The method proves that lexical, syntactic and contextual NLP techniques can help improve sales by enhancing the quality of content.

## References

- 1 Fellbaum C, 1998. WordNet: An Electronic Lexical Database, Bradford Books.
- 2 Devlin, Jacob and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina, 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint arXiv:1810.04805.
- 3 Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.
- 4 Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430.
- 5 Gabor Angeli, Melvin Johnson Premkumar, and Christopher D. Manning. Leveraging Linguistic Structure for Open Domain Information Extraction. In Proceedings of the Association of Computational Linguistics (ACL), 2015.
- 6 Marta Recasens, Marie-Catherine de Marneffe, and Christopher Potts. The Life and Death of Discourse Entities: Identifying Singleton Mentions. In Proceedings of NAACL 2013.
- 7 Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu and Dan Jurafsky. Deterministic coreference resolution based on entity-centric, precision-ranked rules. Computational Linguistics 39(4), 2013.

- 8 Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky. Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In Proceedings of the CoNLL-2011 Shared Task, 2011.
- 9 Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky, Christopher Manning A Multi-Pass Sieve for Coreference Resolution EMNLP-2010, Boston, USA. 2010.
- 10 Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003.
- 11 Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.
- 12 Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O'Reilly Media Inc.
- 13 Danqi Chen and Christopher Manning. 2014. A Fast and Accurate Dependency Parser Using Neural Networks. In Proceedings of EMNLP 2014.
- 14 Van Rossum, G. & Drake, F.L., 2009. Python 3 Reference Manual, Scotts Valley, CA: CreateSpace.
- 15 TextBlob (2017, 30 June). TextBlob: Simplified Text Processing [Online]. Website, Access: [textblob.readthedocs.io/en/dev/index.html](http://textblob.readthedocs.io/en/dev/index.html).
- 16 Mohammad Hossein Amirhosseini, Neuro Linguistic Programming Automation for Improvement of Organizational Performance. Diss. The School of Computing and Digital Media, 2018. Web. 10 Nov, 2019.

## About the Author



*Siddharth is a Lead Data Scientist at Happiest Minds. His role primarily involves solving problems in the NLP, Optical Character Reader (OCR) and Chatbot areas. He is also responsible for helping customers address challenges in the Data Analytics area. His love for his job is driven by his interest in writing code and playing with data structures, and he likes to share his knowledge with the wider community. Siddharth has worked across multiple domains such as Digital Marketing, Edu-tech and Construction.*

# Happiest Minds Service Offerings in NLP (Natural Language Processing)

## 01

### Content Tagging

- ▶ Automated tags generation for audio, video and text content.
- ▶ Deep learning algorithms along with static features for tagging

## 02

### Short Answer Grading

- ▶ Grading natural language-based student's answers w.r.t reference answers.
- ▶ Amalgamation of contextual encoding algorithms & static techniques.

## 03

### Sentiment Analysis

- ▶ Custom approach using classification algorithms.
- ▶ Chatbot historical conversation analysis and scoring.

## 04

### Content Mapping

- ▶ Mapping articles to relevant learning objectives for a course curriculum
- ▶ Use of state-of-the-art algorithms for content search and mapping.

Business Contact [business@happiestminds.com](mailto:business@happiestminds.com)



[www.happiestminds.com](http://www.happiestminds.com)

#### About Happiest Minds Technologies

Happiest Minds, the Mindful IT Company, applies agile methodologies to enable [digital transformation](#) for enterprises and technology providers by delivering seamless customer experience, business efficiency and actionable insights. We leverage a spectrum of disruptive technologies such as: [Big Data Analytics](#), AI & Cognitive Computing, Internet of Things, Cloud, Security, SDN-NFV, [RPA](#), Blockchain, etc. Positioned as "Born Digital . Born Agile", our capabilities spans across product engineering, digital business solutions, infrastructure management and security services. We deliver these services across industry sectors such as retail, consumer packaged goods, edutech, e-commerce, banking, insurance, hi-tech, engineering R&D, manufacturing, automotive and travel/transportation/hospitality.

Headquartered in Bangalore, India; Happiest Minds has operations in the U.S., UK, The Netherlands, Australia and Middle East.