4-4-2019 11:15 AM

Extracting Scales of Measurement Automatically from Biomedical Text with Special Emphasis on Comparative and Superlative Scales

Sara Baker
The University of Western Ontario

Supervisor
Mercer, Robert E.
The University of Western Ontario

Graduate Program in Computer Science
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

Acknowledgements

I would like to express my special thanks of gratitude to my supervisor Dr. Robert Mercer who gave me the golden opportunity to do this work. I am very grateful to his patience and support in overcoming numerous obstacles I have been facing through my research. He provided me with all the support and encouragement that I need. I appreciate every moment I spent working with him, and I feel honored that I got the opportunity to work with one of the most influential professors in my scientific life.

I would also like to thank the University of Western Ontario, the Faculty of Science and the Computer Science Department for their cooperation and support during my studies.

I would like to thank the Saudi Arabian government for supporting me financially.

Finally, but foremost, I would like to thank my family: my parents, my brothers, and my sister for supporting me spiritually throughout writing this thesis and in my life in general.

Abstract

In this thesis, the focus is on the topic of "Extracting Scales of Measurement Automatically from Biomedical Text with Special Emphasis on Comparative and Superlative Scales" Comparison sentences, when considered as a critical part of scales of measurement, play a highly significant role in the process of gathering information from a large number of biomedical research papers. A comparison sentence is defined as any sentence that contains two or more entities that are being compared. This thesis discusses several different types of comparison sentences such as gradable comparisons and non-gradable comparisons. The main goal is extracting comparison sentences automatically from the full text of biomedical articles. Therefore, the thesis presents a Java program that could be used to analyze biomedical text to identify comparison sentences by matching the sentences in the text to 37 syntactic and semantic features. These features or qualities would be helpful to extract comparative sentences from any biomedical text. Two machine learning techniques are used with the 37 roles to assess the curated dataset. The results of this study are compared with earlier studies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The amount of biomedical research papers is increasing at a fast rate, which contributes to creating many challenges for computer systems and humans. One of the most significant problems appears from the frequent use of degrees or scales of measurements and comparison sentences in these texts requiring continuous updating of biomedical databases such as PubMed and Medline databases. These databases include a sizable amount of information that could be helpful for scientists to discover more and more about several kinds of medicines and diseases. Moreover, comparative and superlative sentences play a very vital role in any human discourse which generates any text such as surveys that work on comparing different kinds of products and give information about them.

To collect this information, the process of extracting some particular types of degrees and comparison sentences would be useful for experts to learn more about many topics in the biomedical domain. But, it is important to know that identifying comparison scales and degrees automatically from the text is a complicated process from a computational point of view because there are several types of degrees and comparison sentences which could have different forms and structures. For instance, comparison sentences can be divided into gradable sentences which include words such as ‘greater’, ‘shorter’, and ‘older’, and non-gradable sentences which include phrases such as ‘same as’, ‘similarly’, or ‘as well as’. In contrast, the difference comparisons when considered as non-gradable comparisons appear in sentences that contain words like ‘between’, or ‘different from’. Some of these sentences may contain specific words that refer to the comparison, for instance, ‘more’, ‘most’, and ‘than’ but they are considered as non-comparative sentences. And, some of these sentences may not include any words that relate to the comparison, but they are regarded as comparison sentences.

Therefore, in this thesis, I discuss the problem of identifying comparison sentences and degrees by using a simple JAVA program, which works based on 32 syntactic and semantic features which have been presented by other research efforts (Park and Blake, 2012). In an attempt to improve the efficiency of these features, I created five new rules and techniques to extract comparisons and degrees automatically from the biomedical texts. These 37 syntactic and semantic features can be divided into two types. The first type concentrates on lexical rules that rely on extracting certain terms which could appear in the sentences and they refer to comparisons and scales of measurements. The second type focuses particular forms of dependency trees produced using the Stanford dependency parser (Software Stanford Parser, 2018) that could capture comparisons in the text.

Firstly, I curated a dataset that contains 1000 sentences. The dataset has been annotated manually and parsed by the Stanford parser. A rule-based system, implemented as a JAVA program, is used to extract comparison sentences based on the features. These sentences are compared to the manual annotation to find out the final result that includes accuracy, precision, recall, and F1 score. Then, I created some new rules or features to increase the proficiency of the program. The primary target for this thesis is extracting scales of measurement automatically from biomedical text with special emphasis on comparative and superlative scales. Finally, two machine learning techniques (Support Vector Machine and Naïve Bayes) were used on different datasets to make a comparison of the results. The evaluation that relies on using the JAVA program and machine learning models has indicated very promising results. More specifically, my contributions in this thesis are the following:

- Based on previous work, I identify the problem of extracting comparisons from the full-text articles by looking at the different types of degrees and comparative and superlative scales. In this step, I discover more about the classification of these degrees, the importance of defining this kind of text, and I find out the features and several special terms that could be used to determine them.
- I developed two systems for extracting comparison with particular emphasis on comparative and superlative relations in biomedical text. One is depending on using my rule-based JAVA program, and the other is depending on applying machine learning (Support Vector Machine and Naïve Bayes). I accomplished promising results using these systems based on precision, recall, and F1 score.
- To enhance the precision, I filtered data by relying on the classification for each sentence into "comparative" and "non-comparative" and adding new features to my system to automatically extract the target data.

This thesis is organized as follows: the third chapter talks about some related work. Chapter 4 discusses the problem statement and the classification for several different types of comparison sentences using what is provided by linguistics studies. Chapter 5 describes the suggested rule-based system. Chapter 6 provides a detailed assessment of the features or rules that form that system. And Chapter 7 concludes the thesis Idea and display various points for future work.

Chapter 2

Literature Review

This chapter in my thesis focuses on the topic of “Different Opinions and Perspectives to Clarify the semantics of degrees and the semantics of the scales of measurement”. The literature shows several articles that discuss five different themes which are: the meaning of degrees and gradable adjectives, the motivation behind classifying the measurements of scales in different classes, the type of terms that the researchers are looking for, several methods and techniques that could be helpful to extract the scales of measurements and degrees, and challenges and problems that the researchers encountered with degrees and measurements of scales. The literature review contains differences and similarities between opinions for each author based on some experiments in order to cover the problems, and to answer the question of how could the understanding of the degrees and the scales of measurement affect the clarity of the scientific texts positively? While some of the articles are experimental, citing either positive or negative arguments or maybe the both sides, others discuss several types of expressions that capturing gradability in order to understand different kinds of degrees.

Firstly, this literature review discusses Steven’s articles about “The theory of scales of measurement” (Stevens, 1946), and it describes that Stevens’ idea of scale types by classifying the scales of measurement to four types of classes which are nominal, interval, ordinal, and ratio.

Secondly, I present similar or different perspectives based on the work of Stephanie Solt in her article “Measurement scales in natural language” (Solt, 2014). The main idea of her article is to present a brief for several recent research papers that describing the meaning of scales from a linguistic perspective. She provides several different properties of scales with some examples to show their structures and how they look like in ontological position.

In the article “The Semantics of Degree” (Cresswell,1976), by Cresswell, he mentions that in order to understand different kinds of degrees such as grading, amount, or comparison, and be aware of the semantic relation between them, people need to recognize different types of expressions. Therefore, this article focuses on viewing several types of expressions that capture gradability. Moreover, he discussed several issues that would show up if scientists begin their investigation about the semantics of gradable adjectives. These issues are vagueness and relativity, the relative and absolute distinction, inference patterns, the distribution and interpretation of degree modifiers, comparison, polar opposition, and measure phrases.

The article, by Hospice Houngbo and Robert E. Mercer “Method Mention Extraction from Scientific Research Papers” (Houngbo and Mercer, 2012) presents new ways to automatically find method terminologies in the scientific research papers by using machine learning and rule-based approaches. I found this article to be a good reference for my thesis because I built a rule-based system, and I worked with two different types of machine learning techniques which are Naïve Bayes and Support Vector Machine.

In the article, “A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text” (Schwartz and Hearst, 2002), Ariel S. Schwartz and Marti A. Hearst. present an effective and very fast algorithm which helps in extracting acronyms or abbreviations from the biomedical text by finding pairs of short forms and long forms and matching them in two steps. Some of those abbreviations show different forms and structure for degrees such as (Kg) for Kilogram. In this article, the authors describe the process of this algorithm and how it has achieved a high result in evaluation of Precision and Recall. Also, this article shows that the evaluation or testing of the algorithm is very important to make sure that the matching between the short form and the long form is true and accurate. It is important to know that this article presents a JAVA program that has been created to extract specific types of terms from the biomedical text. The ideas in the article were helpful for me to create my own syntactic and semantic features using JAVA.

The article "Semantics in natural language" (Modifiers, 2012) talks about how to obtain the semantics from adjectives and adverbs in English. In this article, the authors have divided adjectives into three types or categories: intersective adjectives, non-predicative adjectives, and subsective adjectives.

The article "Identifying Comparative Sentences in Text Documents" (Jindal and Liu, 2006) discusses how comparative sentences could be found in text, and how the compared entities could be extracted. They also discuss the difference between objective comparative sentences and subjective comparative sentences, and they could be determined. So, the first step the authors do in their article is determining the types of sentences and classifying them. Then, they offered a modern approach or technique to identify comparison sentences in several types of texts such as articles, reviews, and Internet forums.

Lastly, in the article "Identifying Comparative Claim Sentences in Full-Text Scientific Articles" (Park and Blake, 2012) by Dae Hoon Park and Catherine Blake describe their work on detecting comparison claims automatically from the whole text in scientific articles. Additionally, they introduce a list of syntactic and semantic features or qualities which identify a sentence and then display how those features could be used through several techniques or classifiers such as Naïve Bayes, a Bayesian network, and a Support Vector Machine. Overall,

all articles agree that there are many different types of degrees and scales of measurement, but natural language is sensitive to a large number of scalar features.

2.1 Background Information

In people's lives, each new discovery could change humanity's future. But in order to discover anything people need to be aware of that collecting information or data is an essential step in any process of discovery. Scientific research papers are not only great reference to enhance people's knowledge, but also, they are very important to discover more and more about a specific topic. These days, the field of science has a sizable amount of biomedical research papers which means that there are many biomedical terminologies, that include synonyms, abbreviations, and complicated names, have been discovered. All of these kinds of terminologies could create ambiguity in scientific research papers if they do not interpret clearly and semantically. However, not all of these terminologies could be defined semantically easily. Some of them have been a debatable topic in the scientific domain for many years. One kind of these topics that have caused widespread controversies among scientists in the field of linguistics was called the scales of measurement or degrees. In 1946, the idea of the scales of measurement showed up by the director of Harvard University in that time S. S. Stevens when he tried to define the meaning of measurement of scales by classifying the whole numerical values and measurement of scales under only four types of classes. Those classes or categories are nominal, interval, ordinal, and ratio. He wrote an article about this classification process which called "The theory of scales of measurement" (Stevens, 1946). This contributed to a debate among several experts and scientists arguing that Steven's view was too simple or insufficient to describe all numerical values clearly. However, that also gripped the attention of most of the scientists to how much is important to understand scales of measurement, degrees, and gradable adjectives semantically in order to classify them clearly. Actually, they all agree on that finding the semantic of the degrees or clarifying the scales of measurement is a very useful step to avoid vagueness or ambiguity which could be found in the scientific research papers.

2.2 Gradable Adjectives

The first theme in my literature review is the meaning of degrees and gradable adjectives. According to Stephanie Solt in her article "Measurement scales in natural language" (Solt, 2014) the natural language expressions have much different meaning that could be considered as degrees on scales. Solt demonstrates in her article some similar and dissimilar definitions for scales or degrees by using many expert's points of views which make the topic of degrees as a debatable topic. For example, Solt stated that Stechow says about scale, "whatever they are, they are highly abstract objects" (as cited in Solt, 2014). And this point of view refers to that degrees associated with numerical values or numbers. Various authors such as Kennedy agree with Stechow that degrees are abstracts objects (as cited in Solt, 2014), but some of the others believe that scales considered as a process of comparing entities. Solt says that one of those authors who believe in this concept is Bierwisch. So according to Bierwisch "there is no degree without comparison and no comparison without degree" (as cited in Solt, 2014). Moreover, some experts such as Cresswell think that scales or degrees could be equivalence classes. For example, 'bigger than' or 'more expensive than'. In this relation, the equivalence classes become the degrees of the scale (as cited in Solt, 2014). Also, Cresswell in his article "The Semantics of Degree" think that the meaning of scales of measurement or degrees depends on studying different types of expressions. That would help scientists to understand the difference between degrees such as grading, amount, or comparison. In general, most of the articles in this literature review agree with Creswell's opinion which says that degrees could appear in different forms depending on the text or on the expressions. However, in this thesis, I focused on equivalence classes and comparisons or comparing entities as one of the most important types of scales of measurements.

2.3 Classifying the Measurements of Scale

The second theme in my literature review is the motivation behind classifying the measurements of scale in different classes. First of all, it is important to know more about the first classification that has been done by Stevens in his article "The theory of scales of measurement". As it has been described before, Stevens classified the measurement scales in four classes which are nominal, interval, ordinal, and ratio.

Firstly, nominal variables or categorical variables contain names or labels for definite entities that are mutual, equal, but not ordered. It measures identity and difference, but it does not measure the quantity. Moreover, nominal variables could be applied on statistics such as mode, frequency Distribution, and chi-square. According to Stevens "...the use of numerals as names for classes is an example of the assignment of numerals according to a rule. The rule is: Do not assign the same numeral to different classes or different numerals to the same class. Beyond that, anything goes with the nominal scale" (Stevens, 1971).

Secondly, the ordinal scale which ordered the individual characteristics of two objects or more in the same category to (1st, 2nd, 3rd, etc.) without depending on the element of measurement. It does not focus on equality between elements, but it orders them by focusing on greater than or less than. The race of horses could be a great example for the ordinal scale. The ordinal scale could be applied to statistics such as frequency distribution, median, and percentiles. Stevens observation recorded that psychological measurement mostly works on ordinal scales.

Thirdly, the interval scale which depends on quantitative attributes divides the information into categories or groups. In interval scale the zero-point considered as an arbitrary zero. The Likert scale is a good example for the interval scale (Likert, 1932). Also, Celsius temperature is an interval variable, and IQ tests have adopted the use of interval metric. Interval scale could be applied to statistics such as frequency distribution, median, and percentiles, add or subtract, mean, standard deviation, correlation, regression, and analysis of variance.

Finally, the ratio scale is considered as the main scale for the majority of the physical sciences and engineering. For instance, time, plane angle, energy and electric charge. In addition, ratio scale has the same properties of the interval scale except for the zero point because the ratio scale uses the true or the origin zero point, not an arbitrary one. Kelvin temperature scale is the perfect example for that. The last but not the least, ratio scale could be applied to statistics such as frequency distribution, median, and percentiles, add or subtract, mean, standard deviation, correlation, regression, analysis of variance, and ratio, or coefficient of variation.

This kind of classification that created by Stevens was a very abstract classification from many author's point of view. However, it highlights the importance of the classification of the scales of measurement in order to help scientists to know more about the types of scales, so they could extract these measurements of scales from different types of text such as the biomedical text easily and know more information about it. For example, Solt thinks that classifying the measurements of scales correctly is very important in order to understand them. Also, Nitin Jindal and Bing Liu in their article "Identifying Comparative Sentences in Text Documents" (Jindal and Liu, 2006) believe that the classification of the scales of measurement could bring many different types of degrees such as the comparative sentences that contained four types of comparative degrees which are equative as (ex: as good as), non-equal gradable as (ex: greater or less than, equal to), superlative as (ex: rank one object over all others such as Lee is the tallest), and non-gradable as (ex: similar to or different from). Also, it might have contained a variety of adjectives which could be helpful to collect information from particular types of text such as Internet forums, articles, consumer's reviews that could contain customers opinions. Dae Hoon Park and Catherine Blake (Park and Blake, 2012) agree on Jindal and Liu’s idea that depends on the classification for scales of measurement is very significant to find information about any kind of material (Jindal and Liu, 2006).

In contrast, Cresswell has tried to prove the same idea by reclassified the scale of measurement and degrees depending of the semantic of degrees. Therefore, in his article of “The Semantic of Degrees,” he provides several facts with examples in order to clarify his point of view. According to Cresswell, the first type of semantic expressions refers to the expressions that present common properties of vagueness. This kind of expressions could be true or false depending on contexts. For example, the Mars Pathfinder mission was expensive. This sentence could be false in a context in which the meaning refers to missions to outside of space. However, the same sentence could be true in a context in which the meaning refers to objects with the name of “Pathfinder” such as books, sport-utility vehicles, and mountain bikes. The second type of semantic expressions is expressions which could be altered by degree terms such as how, much, so, very, and too. For instance, Felix bought too many onions. The third type of semantic expressions is expressions which could show up in comparative constructions such as Kim is taller than Lee. The fourth type of semantic expressions is expressions which could be related to measure phrases. For example, ‘I live two blocks from Danny’. In this example, the meaning could be paraphrased in terms of the degree that an object holds some property.

Also, Cresswell sorts out several questions in his article in order to clarify the phenomena of the semantics of degree, for instance, how the language could encode amount, degree, or gradability, and what concepts could be learned about the grammar such as quantification, or lexical representation. Cresswell believes that to answer these questions, there are several facts that people need to be aware of them.

First of all, the fact of degree modification which indicates to that all gradable adjectives accept modification by degree terms such as ‘very’, ‘how’, ‘too’, ‘much’, etc., and modified by comparative constructions. However, there are some adjectives which considered as non-gradable adjectives. This kind of adjective does not accept modification. For example, ‘extinct’ is non-gradable adjective, so saying ‘Dinosaurs are very extinct’ is unacceptable but saying ‘My uncle Javier is very Spanish’ is ok. For this reason, the interaction of gradable adjectives and degree modifiers is not totally constant. In addition, there are several adjectives are modified by ‘pretty’. ‘Pretty’ could change the meaning of the sentence by two ways: the first one could be positive meaning, the second one could be negative meaning. For example, ‘The road is pretty long’ this sentence means that the road is long, but in the sentence of ‘The road is pretty straight,’ that’s mean the road is not straight. Moreover, in the situation of the distribution of “proportional” modifiers such as using words like ‘completely’, ‘partially’, and ‘half’. For example, people could say ‘completely empty’, but they could not say ‘completely long’.

The second fact that has been discussed by Cresswell is relative versus absolute gradable adjectives. The difference between relative gradable adjectives and absolute gradable adjectives is that people could recognize the vagueness in the sentences by looking to adjectives such as small and expensive. In contrast, the arguments in absolute gradable adjectives have the minimal or maximal amount of the property in the sentences.

The third fact in Creswell's article is dimensional versus evaluative gradable adjectives. Dimensional gradable adjectives are those that measure some tangible property of the object such as 'age', 'volume', and 'brightness', and evaluative gradable adjectives are those which measure several subjective, judgment based properties of the object. For example, adjectives like 'interest', 'beauty', and 'quality'.

The fourth fact in the article is called polar opposition which refers to the gradable adjectives that could come in pairs (positive form, and negative form). For instance, (tall, short), (likely, unlikely), (easy, difficult). Actually, it is significant to know that measure phrases would be acceptable just with positive gradable adjectives, but not the whole of positive adjectives accept them. For example, people could say 'Julian is two feet tall', but they could not say 'Julian is two feet short'. Also, people could not say 'Hillary was driving 160 mph fast' or 'Hillary was driving 20 mph slow'. There is also another point that are called factor phrases. Factor phrases are suitable just with positive dimensional adjectives. For instance, people could say 'Julian will soon be twice as tall as he is now', but they could not say 'Julian will soon be twice as short as me'. Finally, Cresswell has sorted some examples in order to explain that positive adjectives with degree morphology do not indicate the unmarked form. Positive evaluative gradable adjectives are more or less similar, nevertheless there may be a little slope to picture the inferences to the unmarked form in some cases. Furthermore, a negative adjective with degree morpheme signifies the unmarked form in the majority of the cases.

In my thesis, I worked on combining the most significant perspectives that have been presented by the authors, by reclassifying scales of measurements depending on using several lexical and syntactical features, and I give a detailed description of my classification in Chapter 4.

In short, classifying scales of measurement and extracting them from texts is very necessary and vital to find information and gathering many different types of data.

2.4 The Types of Terms

The third theme in my literature review is the type of terms that the scientists are looking for. Because the scale of measurements has many different types of degrees, the authors in their articles have discussed different types of terms. The first article by Houngbo and Mercer called "Method Mention Extraction from Scientific Research Papers" (Houngbo and Mercer, 2012) focuses on extracting biological terms from research papers in order to create lexical resources that could be useful for scientists. They concentrate on texts that include an explicit mention of method keywords, for instance, 'algorithm', 'technique', 'analysis', 'approach', and 'method' and other less explicit method terms. I did the same thing in my thesis; I created several lexicons that contain many different types of gradable and non-gradable terms that refer to comparison scales in the biomedical text. I explained these terms and lexicons in Chapters 4 and 5 of this thesis in more details. The second article by Schwartz and Hearst is called "A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text" (Schwartz and Hearst, 2002). The authors of this article are interested in extracting biomedical abbreviations or having more information about it by matching the short forms or the acronyms with the correct definitions or long forms to get more information about the acronyms.

It is essential to be aware that regular adjectives, comparative and superlative adjectives, and adverbs that modify adjectives are a vital part of my classification for degrees in this thesis.

For this reason, I now summarize several articles that discuss many types of adjectives in particular. In the article "Measurement scales in natural language" (Solt, 2014) Solt starts her research by looking for adjectives, specifically, the gradable adjectives, for instance, 'tall' from the adjectival field. Then, Solt's discovery drives her to the class of quantity and amount with some examples such as 'more beautiful than', or 'more residents than'. Then she discusses the scales in the domain of verbal semantics. For instance, a 'lot of books', or 'I slept a lot'. Finally, she states the nominal gradability (nouns) and the modal expressions with a detailed explanation about them. In addition, she mentions that not all of the languages around the world have the same types of scales or degrees. For example, some languages do not have words like 'too' and 'enough'. In addition, Stevens in his two articles about "The Theory of Scales of Measurement" was interested to search more on names, numbers, adjectives that refer to equality and quantity between elements, and the degrees which used in different specialties such as, time, plane angle, energy and electric charge. Cresswell also has focused his investigation on the semantics of gradable adjectives and the semantic analysis of expressions in the English language. Also, he showed several terms that related to comparison such as 'expensive', 'greater' and many other examples of terms which have discussed above in the second theme in my literature review. Likewise, the article of "Modifiers" from the course of "Semantics in natural language" focused on how to get the semantics from adjectives and adverbs in the English language. The authors in this article have divided adjectives in three types or categories, and they have given some examples for each one.

The first category is called intersective adjectives:

1a. Ralph's car is a yellow bus.

1b. Ralph's car is a Volkswagen.

1c. Ralph's car is a yellow Volkswagen.

1d. Ralph's car is yellow.

The intersective adjective in this category is 'Yellow'. That's mean (1c) would be true if and only if Ralph's car is both yellow and Volkswagen. However, in order to preserve this intuition in people's semantics, it is significant to combine the value of the Adj with the value of the N, such as in "yellow bus" which create the semantic value of the noun phrase.

The second category in adjectives is called non-predictive adjectives. And the sentences below are an example of this category:

2a. Ralph is a former basketball player.

2b. Ralph is a teacher.

2c. Ralph is a former teacher.

2d. Ralph is former.

Former is the non-predicative adjective in the sentences. In this type of category, the authors agree on that it is possible to treat 'former' as the way as they have treated 'yellow' before, but it would be given a different type of intension.

The third category in adjectives called subsective adjectives. The four sentences below are an example for this category:

3a. Bob is a tall midget.

3b. Bob is a basketball player.

3c. Bob is a tall basketball player.

3d. Bob is tall.

The main idea to understand this category semantically clearly is relying on treating the subsective adjectives as context-dependent intersective adjectives. So, the meaning of ‘tall’ or ‘large’ would be depending on the text or the context. For example, it could be the set of things that are tall for a snowman or the set of things that are tall for a snowman built by a kid with a 4 years old. Also, the article of “Modifiers” discusses the semantics of adverbs by show several sentences as an example:

4a. Kim kissed Lee passionately on the mouth.

4b. Kim kissed Lee passionately and Kim kissed Lee on the mouth.

4c. Kim kissed Lee passionately.

4d. Kim kissed Lee on the mouth.

4e. Kim kissed Lee.

Furthermore, the authors in the article “Identifying Comparative Sentences in Text Documents” (Jindal and Liu, 2006) and the authors in the article “Identifying Comparative Claim Sentences in Full-Text Scientific Articles” (Park and Blake, 2012) have concentrated their research on extracting comparative sentences from many different types of text. These two articles are vital to my thesis because they discuss different types of comparison sentences, and they present significant rules and features to identify them. Jindal and Liu did their experiments to classifying different types of comparative sentences that might show up in consumer reviews of products, news articles, and Internet forum postings.

These types of comparative sentences are subjective comparative sentences which refer to personal opinion such as ‘Car X is much better than Car Y’, and objective comparative sentence which refers to general fact and a real comparison such as ‘Car X is 2 feet longer than Car Y’. Additionally, they claim that some of the sentences include comparative words, but they are considered as non-comparative sentences. For instance, this kind of sentence ‘I cannot agree with you more’. And there are some sentences that do not include any comparative words, but they considered as comparative sentences, for instance, ‘Cellphone X has Bluetooth, but cell phone Y does not.’ In this article, the authors used sequential role method and machine learning technique in order to determine the difference between all of these types of comparative sentences. Similarly, Park and Blake did the same thing, but they concentrated on identifying comparison claims from the scientific text in articles automatically. In their experiments, they used the Naive Bayes and Support Vector Machine techniques to extract the comparative claim sentences. Also, they mention that comparative sentences have two types of relations which are gradable and non-gradable. Gradable comparisons contain words like ‘greater than’ or ‘shorter length than’. However, non-gradable comparisons could be described similarly or could be described differently. For example, the similarity comparisons appear in sentences that contain words like ‘the same as’, ‘similarly’ or ‘as- as’. In contrast, the difference comparisons appear in sentences that contain words like ‘between’ or ‘different from’. Then the authors have listed some features in their article to determine these types of comparative claim sentences and extract them from the scientific text. Generally, those are the types of terms which the authors, who involved in my literature review believed that they related to different kinds of degrees or scales of measurement.

2.5 Methods and Techniques

The fourth theme in my literature review is some methods and techniques that could be helpful to extract the degrees and the scales of measurement. My literature review contains two articles that focused on the techniques that scientists could use to extract several types of terms. For instance, the technique of machine learning, support vector machine, conditional random field, and role-based approach. Specialist and people who are interested in finding scales of measurement need to understand those important methods in order to extract any types of degrees. However, this literature review would put the emphasis on the types of terms that could be extractable in the text, and it would be giving a brief idea about the most important extraction techniques. In the article for Houngbo and Mercer which called “Method Mention Extraction from Scientific Research Papers” they have focused on extracting biological terms from the research papers in order to create lexical resources that could be useful for scientists. So, the first technique that have been used in order to extract these types of terms is the linguistics-based approach that depends on using some grammatical features to filtering terms. For example, POS (part of speech) in order to extract the biological terms. The second technique is Statistical approach and machine learning approach which depends on using some statistical information such as frequency to extract the terms. And the third technique is the hybrid approach which works by using the both previous approaches to extract the biomedical terms. However, not all of those techniques function properly with all types of terms. Scientists have found that the Linguistics-based approach has performed better than Machine learning approach in extracting explicit mention of method keywords and other less explicit method terms. In my thesis, to extract degrees and comparative and superlative scales, I used a rule-based approach and two machine learning techniques which are Naïve Bayes and Support Vector Machine. I discuss these approaches in detail in Chapters 5 and 6.

In another article “A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text” (Schwartz and Hearst, 2002) Schwartz and Hearst were interested in extracting biomedical abbreviations or having more information about it by matching the short forms or the acronyms with the correct definitions or long forms. They have presented a fast algorithm that works in extracting short forms from the biological text and matching them with the right long forms by using the adjacency to parentheses between the long and the short forms and vice versa. In my thesis, I also used the adjacency technique between specific types of terms to create several syntactical rules that work on extracting comparative and superlative scales. Moving to the article “Identifying Comparative Sentences in Text Documents” (Jindal and Liu, 2006) the authors focused on extracting comparative sentences from specific types of data such as reviews, articles, and forums. They have used a machine learning approach with sequential rules and keywords technique. In the machine learning approach with sequential rules, this technique depends on classifying the sentences automatically and determining which the sentence is a comparative sentence or non-comparative sentence by relying on the group of features that the scientists use with the machine learning. In keywords technique, this technique depends on Determining some terms or Keywords to be extracted to achieve a high recall and to help the machine learning to identify comparative sentences even if they do not include comparative words. In short, these are the most popular techniques which the authors discussed in most of the articles in my literature review.

2.6 Challenges and Obstacles

The last theme in my literature review has presented the challenges and the obstacles that the scientists encountered with degrees and scales of measurement. As Solt mentions in her article that natural language is sensitive to a large number of scalar features. Besides, the research about scales proves that there is a sizable amount of information that goes into different aspects in semantic domain.

Moreover, there is no static rule to classifying degrees or a well-known theory to determine them, and that is what makes the topic of the scale of measurement a debatable topic. Also, is what causes a lot of experts to have similar or different perspectives about the theory of scales of measurement. For example, in the article of "Modifiers", while authors are discussing adverbs, they debated on if the majority of adverbs work like intersective adjectives. But in the end of the article, the most of them decided that "Even if most adverbs work like intersective adjectives, there appear to be some which work like non-predicative adjectives." Moving to Cresswell's article, he discussed a list of issues which related to classifying many different types of degrees such as the relative and absolute distinction, vagueness and relativity, the distribution and interpretation of degree modifiers, inference patterns, comparison, polar opposition, and measure phrases. All of these types of issues extended the domain of degrees and scale of measurement which make the process of confining them and determining them not easy.

Also, there are some challenges that relate to identifying comparative sentences, and Jindal and Liu have sorted several of these challenges in their article. The first challenge is that not all of the sentences that have part of speech tags of (JJS, JJR, RBS, and RBR) are considered as comparative sentences. The second challenge is that there are several sentences that considered as comparative sentences even if they do not include any indicator word. The last challenge is having some badly formed sentences. For instance, sentences that include violation in grammar rules, sentences that are short and incomplete, and sentences that lack punctuation. The authors in the article tend to use the machine learning technique in order to overcome these challenges. Actually, Hospice Houngbo and Robert E. Mercer in their article have referred to the same problem which states that it is not possible to use the general rules in order to extract all types of terms from the text, especially biological text. In their paper, they provided several techniques in order to overcome this problem. For example, the statistical approach and machine learning approach depend on using some statistical information, such as the frequency of terms that appear in the corpus. Another technique that called the hybrid approach, depends on using the both of the previous techniques. Because a linguistics-based approach and a statistical approach and machine learning approach have their own advantages and disadvantages, so the scientists using the hybrid approach to get the best result from the text. The last problem that the scientists could face during their research in the scales of measurement field is related to extracting the abbreviations that refer to degrees or scales of measurement. Schwartz and Hearst in their article "A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text" have developed a fast algorithm to match short forms to the right long forms. They did that because the text contains the both short forms and long forms. However, in the case of degrees or scale of measurement, the text would have contained the short forms only. For example, 'Kg' as a short form in a sentence like 'Lisa's weight is 60 Kg', the text would not include a definition for 'Kg'. To conclude, the domain of scales of measurement and degrees is incredibly huge, so it is very important to know that identifying the whole types of degrees is a real challenge.

Overall, the majority of the authors have talked about many different types of scales of measurement in their articles, and some of them have reclassified the degrees by relying on their own perspectives. From my personal observation after reading, I found that most of the authors disagree on finding a specific rule to classify the scales of measurement correctly in a form that contains all kinds of the degrees and the relations between them. Therefore, relying on all of the opinions that I have mentioned in my literature review, I chose comparison scales to be the main topic of my thesis. I will try to find out a way to classify the scales of measurement in a form which combines several views that have been mentioned by the authors in their articles above. And in the same time, this classification should be suitable with my thesis topic "Extracting Scales of Measurement Automatically from Biomedical Text with Special Emphasis on Comparative and Superlative Scales." Also, I will discuss several types of degrees that relate to comparative and superlative sentences in a biomedical text and biological research papers. Furthermore, I agree with most of the authors that extracting all the kinds of terms which associated to the scales of measurement from any type of text automatically is a real challenge. However, I think with using the help of human's precise observation besides to the extracting techniques, scientists could figure out more and more features about the degrees and the scale of measurement. And that would be helpful to enhance the results and achieve a high performance in the final outcome.

Chapter 3

Related Work

Based on some early work on the identification of comparative sentences and degrees from a text, and as cited in (Saritha and Pateriya, 2014) researchers focus primarily on extracting comparisons from special types of text such as comparative claim sentences in biomedical text (Park and Blake, 2012) and comparative sentences in general business intelligence documents which compare several products to know customers' opinions (Jindal and Liu, 2006). Due to the growing importance of social media, product reviews have attracted much attention recently because they contain users' opinions that describe many different products and services. An example of resources which contain a lot of information in textual form are blogs, websites where people can express their perspectives and post comments on many different things to many other people online. In this type of text, opinion or sentiment most of the time determines the interest to some critical product. Comparing two or more products is a common way to get information about some product. For instance, "Camera X has a better lens than Y" describes a positive review towards Camera X and a negative or less favorable review towards Camera Y (Kessler, 2014). However, these previous studies concentrate on some limited linguistic terms and are not able to achieve my goal of extracting comparisons from the full-text of biomedical articles. This research deals with the identification of comparison sentences, and presents some linguistic techniques to extract comparative sentences in various text genres.

3.1 Comparative Sentences

A significant amount of research has defined a comparative as a sentence form that is used to compare two or more entities. Experts have divided comparative sentences into subjective and objective sentences.

1) Subjective sentences are sentences that refer to a personal opinion

   Ex: I like this phone more than the other one.

2) Objective sentences are sentences that refer to some general facts

   EX: IPhone is much more expensive than Huawei.

Most of the comparative sentences that indicate personal opinions contain explicit terms that are called opinionated comparative words such as ‘better’, ‘worse’, and ‘best’. However, many of the comparative sentences that indicate general facts depend on the meaning of the context. For instance, the term ‘longer’ is not opinionated as it is ordinarily used to say that the length of one entity is greater than the length for the second entity (Ganapathibhotla and Liu, 2008).

In general, comparative sentences tend to include adjectives that end with –er or –est, or include adverbs like ‘more’, ‘most’, ‘too’, or ‘as’ to describe a relation between two or more persons or objects. Comparison sentences can be nominal, verbal, adjectival, or adverbial (Bresnan, 1973) (Carol Friedman, 1989).

Another research work (Stechow, 1984) illustrates several examples to identify the difference between comparing entities and comparing degrees in comparison sentences. For example:

1) John is taller than Mary.
2) John’s height exceeds Mary’s height.

The first sentence states the relation between two entities ‘John’ and ‘Mary’, while the second sentence states the relation between two degrees ‘John’s height’ and ‘Mary’s height’ (Stechow, 1984).

Other researchers, such as Scheible, have focused on superlative scales as one of the essential types of comparison sentences. She wrote two articles (Scheible, 2007) (Scheible, 2008) about extracting superlative scales from text. She described how superlative scales can be attached to the useful information that could appear in text, and to what degree that would be helpful to meet the needs of experts in the natural language field. In her articles (Scheible, 2007) (Scheible, 2008) she used a variety of syntactic and semantic features captured with the part of speech tagging (POS) technique to obtain her goal.

Also, in the past, several studies have used various systems for extracting the comparison sentences and degrees from several types of documents such as the ones that have been mentioned above.

3.2 Approaches and Methods for Identifying Comparisons

A number of studies (Kennedy, 2004; Saritha and Pateriya, 2014; Jindal and Liu, 2006; Park and Blake, 2012; Bakhshandeh and Allen, 2015; Yang and Ko, 2009; Yang and Ko, 2011) have done research on several supervised methods and unsupervised methods to extract comparisons from the text. These methods and techniques have been sorted below.

3.2.1 Linguistic Approach

In this approach, one of the experts tried to classify comparative sentences based on using syntax and semantics (Kennedy, 2004). Syntax and semantics focuses on the terms and linguistic features that could describe the relation in the sentences. This method concentrates on the language form such as the grammatical structure of sentences without looking to the meaning of the sentences (Saritha and Pateriya, 2014).

Two other experts, Bakhshandeh and Allen in 2015 presented a semantic framework for comparison. This framework worked by extracting some indicators for comparison (explicit and implicit comparison terms) and the context related to the indicators by using several semantic features (Bakhshandeh and Allen, 2015).

The majority of research mentioned that many comparison sentences come with some indicators such as ‘more’, ‘less’, ‘most’, ‘–er’ and ‘as’ to indicate the comparative relation in the sentences. In addition, this earlier research illustrates the categories of comparison sentences which split into equative, non-equal gradable, superlative, and non-gradable. Chapter 4 discusses these kinds of comparisons in more detail.

3.2.2 Sequential Pattern Mining Approach

Another technique to extract comparisons from the text is called the sequential pattern mining (SPM) approach. It relies on discovering some statistically relevant forms in the datasets which involve values that delivered in a sequence. The values would be predicted as separate from each other (Jindal and Liu, 2006). SPM works to solve many computational problems, for instance, creating indexes and useful databases for sequence data, identifying repeated patterns or forms, comparing the similarity in a sequence, and returning the missing members of a sequence. Sequential role mining has two types, which are class sequential rules mining and label sequential rule mining.

1) Class Sequential Rules Mining (CSR) is used to classify sentences into different classes and to determine the types of these sentences (Jindal and Liu, 2006).

2) Label Sequential Rules (LSR) are used for extracting the comparison sentences and identifying the relations in these sentences (Jindal and Liu, 2006).

3.2.3 Machine Learning Approach

Supervised machine learning is an approach that is considered as one of the most effective learning techniques in identifying comparison sentences automatically from text. It depends on introducing a group of syntactic and semantic features to distinguish the comparative sentences.

After that, Park and Blake have taken these features and assessed them on biomedical text by using different classifiers such as Naïve Bayes (NB), Bayesian network (BN), and Support Vector Machine (SVM). Their study has obtained 0.71, 0.69, and 0.74 on F1 score using NB, SVM, and BN for the development set, and respectively, the NB, SVM and BN have achieved 0.76, 0.65, and 0.74 on a validation set (Park and Blake, 2012).

Some other experts have used machine learning to remove non-comparative sentences from the text, and then they extract only the comparison sentences. To do this, they classified comparison sentences into six groups similarity, difference, equality, superlative, greater or less than, and predicative (Yang and Ko, 2009). This study obtained 68.39% on precision, 95.96% on recall, and 79.87 on F1 score.

Yang and Ko improved their previous result by classifying comparison sentences in Korean text into seven comparative classes and one non-comparative class and presenting several characteristics for each comparative type, and by collecting comparative keywords and finding relevant features which indicate comparison in Korean text. In this research, they also removed non-comparative sentences from appearing in the corpus. The results using machine learning techniques and 5-fold cross-validation show high performance: 88.59% on accuracy and 90.23% on F1 score (Yang and Ko, 2011).

By looking at these results, all studies have accomplished good results by using a machine learning approach. For this reason, in my thesis, I am interested in using machine learning techniques to assess my dataset. So, I considered the first study (Park and Blake, 2012) as a useful reference that could direct me to obtain my goal of extracting comparative and superlative scales automatically in all sentences in biomedical text not only the claim sentences.

Chapter 4

The Problem Statement and The Classification of Comparison Sentences

In this chapter, I present the problem that I want to solve. Firstly, I describe the comparison from a linguistic perspective and list several limitations. I will further discuss about some implicit and explicit comparisons in order to define the problem that I worked toward solving in this thesis in Chapter 5.

4.1 The Problem from a Linguistic View

Linguists define comparisons as the processes that are used to illustrate an ordering between entities by describing the relation or the degree that has some gradable attribute. The sentence can contain one or more entities under a relation or topic that indicates a comparison. These entities could be names of people, objects, or products. The comparison relation has four different types. The first three types are considered as gradable comparisons, while the last one is considered as a non-gradable comparison:

1) Equative comparison: This type of relation captures the equality between two objects on a specific topic. The sentence in this case would include terms or keywords such as 'same as', 'equal to', 'similar to', etc.

   EX: "Camera A is similar to Camera B".

2) Non-Equal Gradable relation: This type of relation captures the ordering between the objects. The sentence in this case would include terms or keywords such as 'taller', 'better', 'smaller', etc.

   EX: "Lina is taller than her sister".

3) Superlative relation: This type of relation states that one object is greater or less than all the other objects on some gradable scale. The sentence in this case would include terms or keywords such as ‘most’, ‘best’, ‘highest’, etc.

   EX: “Camera A is the best Camera in the store”.

4) Non-Gradable relation: This relation compares different features between the objects without grading them explicitly.

   EX: “Camera A has a good design and Camera B has a good memory card”.

4.2 The Classification for Comparative Sentences

In my thesis, I have classified comparative sentences by adding new types to the previous classification depending on some specific kinds of terms:

1) Finding the difference comparison between two entities in the sentences. In this case the sentence includes words such as ‘the difference’, ‘different to’, ‘differently’, ‘differ’, etc.

   EX: “Mobile A is different than Mobile B”.

2) Finding the contrast between two entities in the sentences. In this case, the sentence includes words such as ‘in contrast’, ‘on the contrary’, ‘however’, ‘on the other hand’, ‘but’, etc.

   EX: “I like apple, but not orange”.

3) Words that capture directions by referring to the degree of some scale in the sentence. In this case, the sentence includes words such as ‘increase’, ‘decrease’, ‘below’, etc.

   EX: “Sales **decreased** by five percent this year”.

4) Finding some adverbs that modify adjectives in the sentence. In this case, the sentence includes words such as ‘significantly’, relatively’ ‘dramatically’, ‘very’, ‘much’, ‘too’, etc.

   EX: “The results between the two groups were not significantly high”.

5) Finding some terms that refer directly to the comparison in sentences. In this case, the sentence includes words such as ‘compared’, ‘comparing’, ‘comparison’, ‘contrast’, ‘compare’, ‘relative’, etc.

   EX: “She is tall, compared to you”.

4.3 Syntactic and Semantic Features

The syntactic and semantic features or rules play a very significant role in the process of distinguishing comparative sentences and various kinds of scales from biomedical text. These scales could be adverbial, adjectival, superlatives, and many other types. One research article that has developed some of these features in order to identifying comparative claim sentences automatically has been written by Park and Blake (2012). They have introduced a set of rules and characteristics that could be used with three different classifiers which are Naïve Bayes, Bayesian Network, and Support Vector Machine. In this thesis, I will describe these features in detail and I will add my new features to them in order to increase their efficacy. The first type of feature depends on some lexicons and terms which make them lexical features. The second type of features depends on using universal dependency in the Stanford dependency parser.

4.4 Features with specific terms and lexicons

The selection of features and rules could affect the classification process considerably. Based on (Blake, 2010) there are 32 features that refer to the lexical and syntactic forms of the sentence. The six lexical features are described blow.

The 26 syntactic features are described in the next section.

L1: The first rule uses words or terms from a lexicon called the SPECIALIST lexicon. This rule would set to true if it has been determined that the sentence is a comparison sentence, when the sentence includes words that refer to comparison. Park and Blake added some terms to the lexicon such as ‘more’, ‘better’, ‘worse’, ‘less’, ‘lesser’, ‘fewer’, and they removed terms such as ‘good’, ‘later’, ‘few’, ‘ill’, ‘low-dose’, ‘well’, ‘long-term’, ‘number’, ‘well-defined’. The SPECIALIST lexicon contains in total 968 terms.

L2: The second feature relies on a direction verb lexicon that contains a list of 104 words that have been created to capture the direction verbs. There are 22 terms of this lexicon have been selected manually by the Park and Blake. This rule would set to true when the sentence includes any terms in the lexicon.

It is important to know that I tried to obtain the lexicon of direction verbs which used by Park and Blake, but I couldn't have it. So, I created my own list for the direction verbs.

L3: This rule would set to true when the sentence contains ‘from’, ‘above’, or ‘over’.

L4: This rule would set to true when the sentence contains ‘versus’ or ‘VS’.

L5: This rule would set to true when the sentence contains ‘twice the’.

L6: This rule would set to true when the sentence contains 'times that of', 'half that of', 'third that of', 'fourth that of'.

4.5 Rules that Work with the Stanford Parser

Before I present the rules that are related to this part, I will mention the Stanford parser and the SimDif lexicon that some of the rules depend on.

Stanford parser: is the parser that analyzes the sentences syntactically and grammatically by producing the Stanford dependency representation and phrase structure trees. For example, in the sentence "My sister also likes eating an apple.", the Stanford parser refers to the subject of 'likes' at 'sister' (Software Stanford Parser, 2018).

SimDif lexicon: is the lexicon that includes words that indicate differences and similarities between the entities in the sentences. It is important to mention that I did not find the exact SimDif lexicon that has been used by the previous experts, so I needed to create my own SimDif lexicon that contains more than 25 words. This lexicon can be found in Appendix A.2. Next, I will list the rules given in Park and Blake (2012) and a short comment about the propose of the rules.

Important explanation for the symbols:

W1_than	'W1' is the word identifier, and 'than' is the constraint that applied to a word.
W4_SIMDIF	'W4' is dragged from terms in the SIMDIF lexicon.
'|'	Depict disjunctions 'OR'.
'¬'	Negations
'?'	Optional.
'*'	Wildcard operators.

Table 1: Explanation of symbols

4.2.1 SimDif lexicon

The SimDif lexicon has been used with the first four syntactic rules below to extract similarities and differences in comparison sentences:

S1: [*root* W_{1_SIMDIF} [*nsubj|cop* W_2, (*prep* W_3)?]]

S2: [¬ *root* W_{1_SIMDIF} [*nsubj|cop* W_2, (*prep* W_3)?]]

The sentence below is a good example for **S2**:

"However, tumor cells were negative for desmin, myocin and myoglobin, **while** being strongly positive for vimentin and actin and slightly positive for HAM56".

Syntactic rules 3 and 4 shows other structures of non-gradable comparisons that combined with prepositions.

S3: [(prep W_1)?, (* W_2)? [(*prep* W_3)?, (*acomp|nsubjpass|nsubj|dobj|conj*) W_{4_SIMDIF} [(*prep* W_5)?]]]

S4: [(prep W_1)?, (* W_2)? [(*prep* W_3)?, ¬ (*acomp|nsubjpass|nsubj|dobj|conj*) W_{4_SIMDIF} [(*prep* W_5)?]]]

The sentence below is a good example for **S3**:

"There was a significant difference in somatostatin-immunoreactive cells between the four groups (PANOVA=0.027)".

4.2.2 Syntactic Rules with 'than'

The following syntactic rules 5, 6, 7, 8, 9, and 10 used to extract gradable and non-gradable comparisons that combined with 'than':

S5: [*prep* W_{1_than}]

S6: [*advmod* W_1_than]

S7: [*quantmod|mwe* W_1_than]

S8: [*mark* W_1_than]

S9: [*dep* W_1_than]

S10: [¬ (*prep|advmod|quantmod|mwe|mark |dep)* W_1_than]

The sentence below is a good example for **S5**:

"In the diabetic rats, stomach and duodenum LPO levels were significantly higher **than** those of the other group".

4.2.3 Syntactic Rules with 'compared', 'comparing', 'comparison'

The following syntactic rules 11, 12, 13, 14, 15, 16, 17, 18 and 19 used to extract gradable and non-gradable comparisons that combined with these terms 'compared', 'comparing', 'comparison':

S11: [*advcl|prep* W_1_compared]

S12: [*dep* W_1_compared]

S13: [¬ *(advcl|prep|dep)* W_1_compared]

The sentence below is a good example for **S11**:

"In diabetic rats, a significant decrease in stomach GSH levels was observed when **compared** with control group (bp < 0.0001)".

S14: [*advcl* W_1_comparing]

S15: [*partmod|xcomp* W_1_comparing]

S16: [*pcomp* W_1_comparing]

The sentence below is a good example for **S16**:

“Relative quantitation of the PCR products was accomplished by **comparing** the signals densitometrically”.

S17: [*nsubj* W1_comparison]

S18: [*pobj* W1_comparison]

S19: [¬ *(nsubj|pobj)* W1_comparison]

The sentence below is a good example for **S18**:

“The relative mRNA level in each band was calculated by **comparison** with the expression level of the endogenous control β-actin mRNA, which was used as an endogenous control”.

4.2.5 Syntactic Rules with ‘contrast’, ‘relative’

The following syntactic rules 20, 21, 23, and 24 used to extract gradable and non-gradable comparisons that combined with these terms ‘contrast’, ‘relative’:

S20: [*dep* W1_contrast]

S21: [*pobj* W1_contrast]

The sentence below is a good example for **S21**:

“In **contrast** to the hamster visual cortex, more than half of the GluR1-IR neurons are located in layer VI in rat visual cortex”.

S22: [*advmod* W1_relative]

S23: [*amod* W1_relative]

S24: [¬ *(advmod|amod)* W1_relative]

The sentence below is a good example for **S22**:

"This established the linear phase of amplification for each PCR product **relative** to internal control β-actin”.

4.2.6 Syntactic Rules combined two terms or character together

The following syntactic rules 25, and 26 used to extract gradable and non-gradable comparisons that combined two terms or special character together in one sentence such as ‘compare’ with ‘well’ or ‘favorably’, and ‘%’ with ‘of’:

S25: W1_compare [*advmod* W2_(well|favorably)]

S26: W1_% [*nsubj* W2 [*prep* W3_of]]

Chapter 5

The Rule-based System and the Annotated Corpus

In this chapter, I illustrate the JAVA program that I implemented as a rule-based solution using the lexical and syntactic rules to extract comparisons from biomedical text. Firstly, I describe the corpus used in all of the experiment. Then, I explain how the program functions, and I talk about the challenges that I faced to accomplish my goal. At the end of this chapter, I list some additional rules and several modifications to my Java program to improve accuracy and reduce challenges so that I can get better results from my system.

First of all, I generated a corpus containing 1000 sentences which have been chosen randomly. To prepare these sentences for use by the rule-based system, I converted these sentences into dependency trees by using the Stanford dependency parser (Software Stanford Parser, 2018). Also, since the rule-based system is based on the lexical and syntactic rules, I used version 1.6.9 to obtain the same labels in the dependency trees that Blake and Park used in their article "Identifying Comparative Claim Sentences in Full-Text Scientific Articles" (Park and Blake, 2012) and because newer version of the Stanford parser give different labels than older versions. Secondly, I created a JAVA program that implement the 32 rules that have been described in chapter 4.

5.1 The Corpus

The corpus contains 1000 sentence which I have chosen randomly from different biomedical articles.[1] I annotated this corpus manually to classify the sentences as comparison and non-comparison sentences.

[1] FTP Service. Current Neurology and Neuroscience Reports. U.S. National Library of Medicine. Retrieved on 25 July 2009. www.ncbi.nlm.nih.gov/pmc/tools/ftp/.

During my annotation, I removed all the tables, references, citations and figures from the corpus. It is significant to be aware that I built this corpus to use it in my rule-based system to assess the lexical and syntactic features. After the annotation, I got 277 comparison sentences and 723 non-comparison sentences.

5.2 Rule-Based System (JAVA Program)

I uploaded the file of 1000 sentences on the program. Then the program shows a table of 32 columns and 1000 rows. The columns contain the names of the rules as labels, and each row contains the number of the sentence and the result of every rule for that sentence. The result appears as 0s and 1s. So, when the result for any rule is 1 that means the sentence is a comparison sentence, and when the result for each rule is 0 that means the sentence is not a comparison sentence. The interface of the program and how it functions is shown in Figure 5.2.

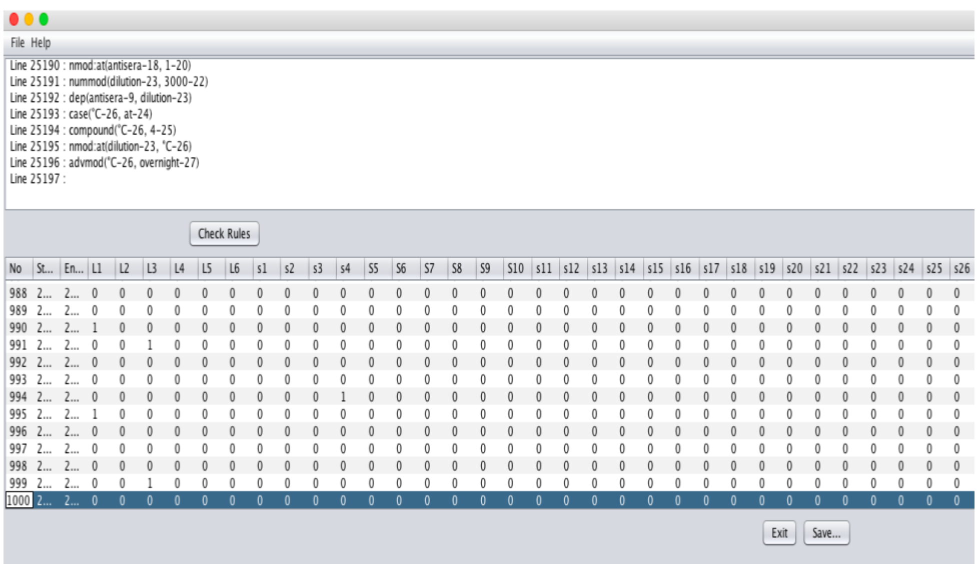

Figure 5.2: Main interface of JAVA program

Figure 5.2 shows the GUI that is part of my rule-based JAVA program. I uploaded the file of 1000 sentences in dependency tree format that have been produced by the Stanford parser by clicking on file. Then, the sentences appear in the text area. I then got the result by clicking on the button ‘Check Rules.’ Then the output of the rules shows in the table that appears below the text area. Also, I added a save button to save the result and an exit button for closing the program.

Both lexical and syntactic rules have been coded by using JAVA. For the lexical features, I uploaded each lexicon as an array list to my rule-based system. Each rule is implemented individually as a regular expression.

5.3 Problems and Challenges

While working on the program and analyzing the data, I encountered many challenges:

1) Technical problems: After converting the dataset to a dependency tree, I found that using the latest version of the dependency tree could change the result because there are several new tags have been added to the new version and they do not match with that tags that Park and Blake used to create their 32 syntactic rules (Park and Blake, 2012). Because I want to obtain the correct results for the 32 features in my dataset, I used the same version of 1.6.9 that Blake and Park used in their article.

For example, the scientists in the article used this sentence “DBP is several orders of magnitude more mutagenic/ carcinogenic than BP”. By using version of 1.6.9 of dependency tree, the result of the tags was as appears below:

nsubj(orders-4, DBP-1)

cop(orders-4, is-2)

amod(orders-4, several-3)

root(ROOT-0, orders-4)

prep_of(orders-4, magnitude-6)

advmod(mutagenic\/carcinogenic-8, more-7)

amod(magnitude-6, mutagenic\/carcinogenic-8)

prep_than(magnitude-6, BP-10)

However, by using the latest version of dependency tree, the result of the tags changed as appears below:

nsubj(orders-4, DBP-1)

cop(orders-4, is-2)

amod(orders-4, several-3)

root(ROOT-0, orders-4)

case(magnitude-6, of-5)

nmod:of(orders-4, magnitude-6)

advmod(mutagenic/carcinogenic-8, more-7)

amod(magnitude-6, mutagenic/carcinogenic-8)

case(BP-10, than-9)

nmod:than(magnitude-6, BP-10)

2) The problem of finding the lexicons: finding the lists of terms that have been used by the scientists in their article was a real challenge for me, so I built my own lists or lexicons.

The Lexicon	Number of terms
SPECIALIST Lexicon	987 terms
Direction Verbs Lexicon	212 terms
SimDif Lexicon	33 terms

Table 2: The list of terms used throughout this thesis is provided in Appendix

3) A compound term that consists of two words or more: In the SimDif lexicon, there are compound terms that refer to similarity or difference in the sentence such as ('as well as', 'as far as', 'one way', 'on the other hand'…etc.)

   The main challenge with this type of term is that the dependency tree does not work with phrases or compound terms, so it gives different tags for each word in the compound term. For instance, the dependency tree will divide the phrase of 'as well as' in three lines and it will give different tags for every word.

   EX: "Terminal Schwann cells (TSCs) that cover motor neuron terminals, are known to play an important role in maintaining neuromuscular junctions, **as well as** in the repair process after nerve injury".

   The dependency tree for 'as well as' in the sentence above is:

   advmod(well-26, as-25)

   cc(in-20, well-26)

   mwe(well-26, as-27)

4) Regular adjectives in the SPECIALIST Lexicon: I found that most of the words in the SPECIALIST Lexicon are regular adjectives, and there are very few words that can be considered comparative adjectives.

5.4 Modifications and New Rules for Improvement

To deal with some challenges that I described above and to enhance the result of the rule-based system, I added several new rules and modifications to the JAVA program.

1) I improved the SPECIALIST Lexicon by adding two more lists. The first list includes comparative adjectives that refer to comparison in the sentences, and I add this list of

words to the same SPECIALIST Lexicon that related to L1 to enhance its efficiency. Then, I connected this list with the syntactic rules which focus on 'than' and use rules from S5 to S10. And the other list is a separated list which contains superlative adjectives that could appear in the text. I created a new rule for the superlative list:

MR5: [superlative (W1)]

This new lexicon is shown in Appendix A.5.

2) I built a new lexicon that includes adverbs that modify adjectives. In fact, during my analysis of the biomedical text, I found some sentences that contain expressions such as 'significantly high,' or 'relatively low.' So, I considered these types of expressions as another form or structure for superlative scales:

MR2: [acomp|amod Adverbs (W1)]

This new lexicon is shown in Appendix A.7.

3) I divided the SimDif lexicon into two lists which are SimDif and SimDif 2Word. The SimDif list includes words that consist of one part such as 'similar', 'difference', 'likewise', 'while', 'unlike', etc. The SimDif 2Word includes words that consist of two parts such as 'much as', 'in common', 'on the other hand', 'much like', etc. After that, I created a new rule that works with SimDif 2Word to extract words that consist of two parts:

MR4: [SimDif2Word (W1)]

This new lexicon is shown in Appendix A.6.

4) I added a new rule that uses a construct of 'as – as' to extract sentences with expressions like 'as well as', 'as far as'etc:

MR3: [W1_ as [*W2 [W3_as]]]

5) I added a new rule to extract any sentence that includes the expression of 'as compared

to' which appears as an advmod in the sentence. I found that this type of sentences refers to a comparison for sure:

MR1: [advmod W1_as [W2_compared [W3_to]]]

A complete list of all of words in all of the lexicons is found in the Appendices. A summary of the lexicons and the number of words in each together with the source of these words is given in Table 3.

The Lexicon	**Number of terms**
SPECIALIST Lexicon (Browne, McCray, & Srinivasan, 2000) (*LexAccess,* 2016)	987 terms
SPECIALIST Lexicon Comparative Adjectives (Easy Pace Learning, 2011)	202 terms
Superlative Lexicon (Easy Pace Learning, 2011)	203 terms
Direction Verbs Lexicon (*VerbNet*, 2013)	212 terms
SimDif Lexicon 1 (Blauman, 2017)	26 terms
SimDif Lexicon 2 (SimDif2Word) (Blauman, 2017)	7 terms
Adverbs Lexicon (*100 Adverbs,* 2015) (Comprehensive site for English learning, 2018)	30 terms

Table 3: The list of lexicons used throughout this thesis and the number of words contained in each.

5.5 The New Syntactic features

I have assembled the new rules below to have them in one place.

MR1: [advmod W1_as [W2_compared [W3_to]]]

MR2: [acomp|amod Adverbs (W1)]

MR3: [W1_ as [*W2 [W3_as]]]

MR4: [SimDif2Word (W1)]

MR5: [superlative (W1)]

Chapter 6

Testing and Evaluation Systems

In this chapter, I demonstrate three evaluation systems that I used to assess my work with the 32 syntactic features (Park and Blake, 2012), the new lexicons, and the five new rules that I added to enhance the process of extracting comparison and non-comparison sentences from biomedical text. In my evaluation, I used a Java program which is a rule-based system that I created to obtain my goal. Also, I used two types of machine learning which are Support Vector Machine and Naïve Bayes. With these machine learning techniques, I classified my dataset by using tenfold cross-validation to check the accuracy, precision, recall, and F score for all the syntactic rules.

6.1 Classifiers

In my thesis, I used two classifiers which are the Support Vector Machine and Naïve Bayes. I have chosen these two machine learning techniques because they work effectively with different kinds of text and Park and Blake (2012) used them, so well.

6.1.1 Naïve Bayes (NB)

Naïve Bayes is one of the machine learning classifiers. It is simple, works effectively with text, and is frequently used to classify different datasets. It is one of the very common classifiers that work in a classification process for text, and it is one of a conventional way to solve many problems such as spam detection (Zhang, 2004).

The Naïve Bayes' proposition depends on the independence hypotheses between predictors. The model is easy to build, with a very simple iterative parameter valuation that helps to make the use of the classifier suitable for massive datasets. Bayes' algorithm works to calculate the posterior probability, P(c|x), from P (c), P (x), and P(x|c).

The Naïve Bayes classifier supposes that the impact of the amount of the predictor (x) on an offered class (c) is separated from the amounts of other predictors. This kind of assumption is known as class conditional independence (Zhang, 2004).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Symbol	Description
P(c|x)	The conditional probability of a class (goal) given the predictor (property).
P(c)	The prior probability of class.
P(x|c)	The conditional probability of the predictor given the class.
P(x)	The prior probability of the predictor.

Table 4: Description of symbols

6.1.2 Support Vector Machine (SVM)

Support Vector Machine is one of the supervised machine learning techniques that works to solve regression problems or classification challenges. It is commonly used in classification problems because it works well with text (Lin and Wang, 2002).

The support vector machine training works to decrease the error function:

$$\frac{1}{2}\, W^T W + C \sum_{i=1}^{N} \zeta i$$

With the following constraint:

$$y_i\,(w^T \varnothing(x_i) + b) \geq 1 - \zeta_i \text{ } and \text{ } \zeta_i \geq 0, i = 1, \dots, N$$

Symbol	Description
C	The capacity constant.
W	The vector of coefficients. (w^T is the transpose)
b	The bias constant.
ζ_i	Represents parameters for handling non-separable data (inputs), and the i labels the N training cases.
$y \in \pm 1$	Demonstrates the class labels.
xi	Demonstrates the independent variables.
$\varnothing$	Kernel function used to transform the input to the feature space.

Table 5: Description of symbols

It is essential to be aware that when the C be larger, the error would be more detectable. For this reason, the C must be chosen with carefulness to prevent overfitting (Wang, 2005).

6.2 Results and Discussion

In this thesis, my study was conducted using a data set of 1000 sentences that have been chosen randomly from several different full-text biomedical articles to assess the initial set of 32 syntactic features (Park and Blake, 2012). Then, I added my development set of features which includes my five rules that I created to improve the result.

Sentence Type	Development	True Positive and True Negative
Comparative Sentences	402 (40.2%)	275 (31.57%) (True Positive)
Non-comparative sentences	598 (59.8%)	596 (68.43%) (True Negative)
Total	1000 (100%)	871 (100%)

Table 6:Distribution of comparative and non- comparative sentences

It is important to know that when sentences are randomly chosen from the corpus of biomedical articles, tables, figures, references, and citations have been eliminated from that corpus. Also, I annotated my corpus manually to identify comparison sentences, and after that, I compared my manual result to the result that has been generated by the rule-based system which is my JAVA program. Based on the comparison of the manual annotation and the rule-based system, 129 sentences have been distributed between false positive and false negative sentences. I will talk about this in detail in my analysis for the dataset below.

6.2.1 Rule-Based Result (JAVA program)

My evaluation shows that the F1 score on the rule-based is 0.63 for comparison sentences, and 0.88 for non-comparison sentences before adding the improvement rules. However, the result has been improved after adding the improvement rules to be 0.81 for comparison sentences on an F1 score and 0.90 for non-comparison sentences on an F1 score. The result has been described clearly in the table below to present accuracy, precision, recall, and F1 score for both comparison and non-comparison sentences:

	Rule-Based 32 Rules	**Rule-Based 37 Rules**
Accuracy	0.816	0.871
Comp. Precision	0.704845815	0.68408
Comp. Recall	0.577617329	0.99278
Comp. F1 score	0.634920635	0.810015
Non-comp. Precision	0.848641656	0.996656
Non-comp. Recall	0.907330567	0.824343
Non-comp. F1 score	0.877005348	0.902347

Table 7: Java programm Results before and after adding the new rules

Note:
Accuracy = TP+TN/TP+FP+FN+TN
Precision = TP/TP+FP
Recall = TP/TP+FN
F1 Score = 2*(Recall * Precision) / (Recall + Precision),
where TP are the number of true positives, TN are the number of true negatives, FP are the number of false positives, and FN are the number of false negatives (Joshi, 2016).

6.2.1.1 Analyzing the Dataset

I found that the rule-based system has recorded 127 sentences as a comparison which based on my annotation are considered as false positive sentences. I carefully analyzed these sentences to provide reasons for these errors. This analysis appears below:

1) There are 74 sentences that include words such as 'from,' 'over,' and 'above'. These sentences follow rule L3, and I found that no sentence which contains these three words is considered as a comparison sentence for two reasons. Firstly, during my analysis, I detect 68 sentences that involve 'from,' 'over,' and 'above', but they are considered as non-comparison sentences because they include two entities or more with no comparison word or any other indicator of comparison. For example, the sentence

"Gauze pads were placed **over** the bowel, frequently moistened with warm PBS, and covered with aluminum foil to prevent photobleaching". Secondly, I detected six sentences that come with 'from,' 'over,' and 'above', and they contain comparison words. But, they have only one entity. For instance, "Zinc deficiency in diabetics could result **from** the hyperglycemia or the impaired intestinal zinc absorption or increased oxidative stress". In addition, some of the sentences include the word 'above,' but they are considered as non-comparison sentences because it is being used to refer to something in the previous text. An example of this kind of sentence "After washing **above**, a purple color was developed with 0.02 % 3, 3'-diaminobenzidine and 0.3 % nickel ammonium sulfate in 50 mM Tris-HCl buffer (pH 7.6)".

2) There are three sentences which follow the syntactic rule S3 which extract sentences that include SimDif words with specific tags in the dependency tree. After the annotation process, I find that these sentences are non-comparison sentences because there are no entities for comparison in these sentences, even if they include SimDif terms. For instance, this sentence "Significant **differences** were defined as $p < 0.05$".
3) There are 49 sentences which follow the syntactic rule S4 that extract sentences that include words from the SimDif lexicon. After analyzing these sentences, I found that they are non-comparison sentences because even if they contain terms from the SimDif lexicon, all of these sentences talk about only one entity. For example, the sentence "The primers for β-actin PCR were designed to encompass **different** exons, and were expected to yield a 266 bp PCR fragment".
4) There is one sentence that follows the syntactic rule MR5 that refers to words from the superlative list. This sentence is considered as non-comparison because it contains the phrase 'at least,' but from the context, I realized that not every use of 'at least' could be

an indicator for comparison. For instance, the sentence “Before staining, these sections were kept for **at least** 4 days at 4°C in 0.1 M PBS, pH 7.4, containing 0.3 % Triton X-100 (PBST)”.

In addition to the sentences that were misclassified as comparisons, I found that the rule-based system has recorded two sentences as a non-comparison. Those two sentences are related to each other, and they contain two entities with an adverb that modifies an adjective which makes them comparison sentences, but the problem is that the two entities have been divided between the two separate sentences. So, at this point, I have to agree with the rule-based system that those two sentences are non-comparisons since the rule-based system looks at each sentence separately and does not look at the text as a whole. That’s why the rule-based system did not detect the comparison at this point. The two separate sentences below are a good example for this case:

Sentence 1: “Treatment with zinc sulfate for 60 days was found to increase in duodenum GSH levels in diabetic rats (dp < 0.01)”.

Sentence 2: “The NEG levels in duodenum tissue were **very low**.”

- Also, I found a sentence that is considered as a false negative sentence, but based on my annotation, I classified this sentence as a comparison for this reason:

The sentence contains two entities, but there is no any comparison word that refers to a comparison. However, this sentence is considered as a comparison sentence because of the meaning of the bold-faced words in the sentence.

“In the DMNV, **a small number** of ChAT neurons in the lateral part were positive for ChAT and **a large number** of ChAT-**positive** neurons in the medial part were **negative** for FGF1”.

6.2.2 Machine Learning Result

I used Support Vector Machine (SVM) and Naïve Bayes (NV) to determine the accuracy, precision, recall, and F1 score for all of the syntactic features. I classified my dataset by using tenfold cross-validation on both approaches to make sure that I achieved my target. I did that by creating a small python program using Anaconda Navigator software (Anaconda, 2018). For Naïve Bayes, I got 0.94 on F1 score for comparison sentences, and 0.79 of F1 score for non-comparison sentences. For Support Vector Machine, I got 0.91 on F1 score for comparison sentences, and 0.77 on F1 score for non-comparison sentences. The result has been described clearly in the table below to present accuracy, precision, recall, and F1 score for both comparison and non-comparison sentences:

	NB	**SVM**
Accuracy	0.902024302	0.877093009
Comp. Precision	0.888420969	0.904761086
Comp. Recall	0.988907914	0.928120244
Comp. F1 score	0.935921149	0.916087236
Non-comp. Precision	0.959092584	0.800828814
Non-comp. Recall	0.675264549	0.743650792
Non-comp. F1 score	0.791752002	0.769788508

Table 8: Support vector machine and Naïve bayes

6.2.2.1 Analyzing the Dataset (SVM)

I found that the Support Vector Machine has incorrectly recorded three sentences as a comparison. After comparing SVM result to my result, I found that those sentences are considered as false positive sentences. I carefully analyzed these sentences to provide the reasons which appear below:

1) There are two sentences that follow rule S4 which extracts any sentence that includes a SimDif word. For example, this sentence contains ‘but’. However, it talks only about one entity.

   “Claudin-4 expression was observed not only at the apical **but** also at the lateral surfaces of the cell”.

2) The third sentence also follows rule S4 which extracts any sentence that includes a SimDif word. The sentence contains ‘either’, but it talks only about one entity that involves two different composites:

   “The primary antiserum was diluted at **either** 1: 250 (GluR1 and calbindin D28K), or 1: 200 (parvalbumin and GABA)”.

I also compared the result of SVM and the Rule-Based result, and I found that the Support Vector Machine has misclassified eight sentences out of 11. The main reason for this misclassification is that some rules have been activated in the dataset less than others. So, the SVM did not train enough on these rules to extract those eight sentences correctly. That’s why the SVM considered these sentences as non-comparison. The table below shows how many times the several features have been activated, specifically for the misclassified sentences.

Sentences No	Rules	Activated in the Dataset	My annotation	Rule-Based Result	SVM Result
41	S22	One time	True	True	False
346	MR2	17 times	True	True	False
471	S3	19 times	True	True	False
542	MR4	14 times	True	True	False
614	S4	171times	False	True	True
687	S4	171times	False	True	True
734	S3	19 times	True	True	False
781	MR3	7 times	True	True	False
856	S3	19 times	True	True	False
887	S4	171times	False	True	True
949	S7	16 times	True	True	False

Table 9: Explanation for Sentencs that SVM missclassified them

6.2.2.2 Analyzing the Dataset (NB)

I recognized that the Naïve Bayes had recorded six sentences as a non-comparison out of six. But comparing to my annotation result and to the rule-based result, I found that those sentences are considered as false negative sentences for the same reason which emphasis on that some rules have been activated in the dataset less than others. The table below shows how many times the several features have been activated, specifically for the misclassified sentences.

Sentences No	Rules	Activated in the Dataset	My annotation	Rule-Based Result	NB Result
41	S22	One time	True	True	False
221	S4	171times	True	True	False
471	S3	19 times	True	True	False
627	S13	6 times	True	True	False
901	S4	171 times	True	True	False
946	S4	171 times	True	True	False

Table 10: Explanation for Sentencs that NB missclassified

It is important to realize that when I say that S4 has been activated 171 times in the dataset, this means S4 is only active in 171 sentences out of 1000. In addition, the 10-fold cross-validation further reduces this number (171* 9/10 = 153.9). So, the activation number for rule S4 compared to the number of sentences in the dataset is low. That maybe the reason why the NB system incorrectly classifies these sentences and considers them as not comparable.

Another reason, I found that when there is more than one rule has been activated besides rule S4 in the sentence, the Naïve Bayes method will classify this sentence correctly. However, when the S4 has activated alone in some sentences, the Naïve Bayes classified these sentences incorrectly.

Chapter 7

Conclusion and Future Work

7.1 Contributions

In this thesis, I emphasized the importance of comparison sentences and how they play a very significant role in biomedical text. Extracting comparison sentences automatically from scientific text would be useful to help scientists and experts in the biomedical domain to find the information contained in these sentences easily. This information could present a comparison between findings, laboratory data, and earlier research hypotheses and new discoveries. Therefore, I have introduced different types of degrees and scales of measurements that are related to comparison scales in this thesis. My goal focused on extracting scales of measurement automatically from biomedical text with particular emphasis on comparative and superlative scales. To obtain this target, I built a rule-based system (JAVA Program) that works based on several syntactic and semantic features., I first used 32 syntactic and semantic characteristics that have been presented by other research efforts (Park and Blake, 2012). These had to be reimplemented because no code was publically available. Then I concentrated on improving these characteristics by adding my new features and enhancing the related lexicons.

In addition to the rule-based system, I had to create a dataset of biomedical sentences since no dataset targeted to the task of extracting comparison sentences was publically available. Experiments considered 1000 sentences which have been chosen randomly from many different full-text biomedical papers. From these sentences, I extracted at least 275 comparison sentences.

To summarize, my contributions in this thesis are the following:

- I developed a rule-based system for extracting comparative sentences from the biomedical text. The extracting process involves several steps: I generated a dataset that includes 1000 sentences. These 1000 sentences were preprocessed using the Stanford parser in preparation for the evaluation of the semantic and syntactic rules. I implemented these rules in my JAVA program to extract comparisons from the dataset. I annotated the dataset manually and compared this manual annotation to the program result. The results show that the accuracy and F1 scores of the 32 features on the rule-based system were reasonably low. The rule-based system got 82% accuracy, 63% F1 score, 70% precision, and 58% recall for comparison sentences. Also, it obtained 88% F1 score, 85% precision, and 91% recall for non-comparison sentences.
- I implemented my new rules in the rule-based system: I enhanced the SPECIALIST Lexicon by adding two more lists which are the comparative adjectives list and the superlative adjectives list. And I connected the comparative list to the syntactic rules that involve the word 'than.' Also, I built a special rule which works to extract adverbs that modify adjectives, and I created a new lexicon for this rule. Moreover, I divided the SimDif lexicon into two lists which are SimDif and SimDif 2Word. These lists are useful to extract all words and phrases that consist of one or two parts and refer to similarities and differences in the text. Additionally, I added a new rule to extract the expressions of 'as-as,' and I added another rule to extract any sentence that contains the expression 'as compared to.' Based on my analysis, I found that these types of expression indicate comparison in the text. The result shows that the accuracy and the F1 scores have improved after enhancing the lexicons and adding my five-new syntactic and semantic features to the rule-based system. The improved results were

87% accuracy, 81% F1 score, 68% precision, and 99% recall for comparison sentences, and 90% F1 score, 99% precision, 82% recall for non-comparison sentences.

- I used two machine learning techniques to find out accuracy, F1 score, precision, and recall for the final set of 37 rules. I classified my dataset by using tenfold cross-validation with both Support Vector Machine and Naïve Bayes to check the results. Results show that the accuracy and F1 scores of the NB were statistically higher than the SVM. Support Vector Machine has achieved 88% accuracy and 92% F1 score. However, Naïve Bayes has reached 90% accuracy and 93% of the F1 score. At the end of my thesis, I can say that I obtained a promising result using my rule-based system to achieve my target that focusing on extracting scales of measurement automatically from biomedical text with a special emphasis on comparative and superlative scales.

7.2 Future Work

There are some problems and challenges that I encountered while working on this thesis but only partly solved them because of time constraints. If I get enough time in the future, I will create some new rules and features that focus on extracting the type of sentences that are considered as a comparison, but they do not include any straight forward indicator for comparison. I could do that if I study several different types of text semantically and syntactically in depth, or trying to discover a new approach to extract these sentences by looking to the context and without relying on explicit terms. I found that following sentence is a good example of this type of comparison "In the DMNV, **a small number** of ChAT neurons in the lateral part were positive for ChAT and **a large number** of ChAT-**positive** neurons in the medial part were **negative** for FGF1".

Also, I could consider different types of scales of measurements or degrees without concentrating on comparisons as the main topic, so I could find more characteristics to identify them. That would broaden my scope of research, and increase its comprehensiveness. For example, I could look at numerical values or numbers, grading, and amount.

During my analyzing process, I faced some problems in determining the comparison sentences. Next time, it would be helpful for me to obtain a dataset that has been annotated by some linguistic experts. That would enhance my results, and make it more accurate.

One of the challenges was finding the lexicons and lists which have been used by the previous researchers. I created my own lexicons in this thesis because some of them were not available. So, it might be I have missed some important terms. Having the previous lexicons together with my new ones would be possibly improve this work.

Another improvement would be to find comparisons that include an entity which are compared to some unknown entities in the previous text. For instance, in the following example the two sentences talk about diabetic rats and give some specific information about them. This information could be comparing to information about normal rats such as in this sentence "Treatment with zinc sulfate for 60 days was found to increase in duodenum GSH levels in **diabetic rats** (dp < 0.01). The NEG levels in duodenum tissue were **very low** (data not shown)."

Additionally, to help ordinary users who are not familiar with using the Stanford parser, I will improve my JAVA program by implementing some JAVA code that works to receive any regular text and analyzing it directly. It would then not be necessary for the user to do this step first before using the rule-based system.

Lastly, I could use other types of machine learning such as Neural Network to process the datasets.

Bibliography

Anaconda. (2018) *Anaconda 2018.02 for MacOS Installer*. Retrieved October 21, 2018, https://www.anaconda.com/download/#macos.

Bakhshandeh, O., & Allen, J. (2015, September). *Semantic Framework for Comparison Structures in Natural Language*. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 993-1002).

Blake, C. (2010). Beyond genes, proteins, and abstracts: *Identifying Scientific Claims from Full-Text Biomedical Articles*. Journal of Biomedical Informatics, 43(2), 173-189.

Blauman, L. (2017). *Key Words for Compare and Contrast.* Retrieved from the companion website for Teaching Evidence-Based Writing.

Bresnan, J. W. (1973). *Syntax of The Comparative Clause Construction in English*. Linguistic Inquiry Volume IV. Linguistic inquiry, 4(3), 275-343.

Browne, A. C., McCray, A. T., & Srinivasan, S. (2000). *The Specialist Lexicon*. National Library of Medicine Technical Reports, 18-21.

Comprehensive site for English learning. *Modifying Adjectives*. Retrieved November 02, 2018,http://myenglishgrammar.com/exercise-4-adverbs/432-modifying-adjectives-part-1.html

Cresswell, M. J. (1976). *The Semantics of Degree*. In Montague grammar (pp. 261-292). Academic Press.

Easy Pace Learning. Retrieved November 02, 2018, https://www.easypacelearning.com/all-lessons/grammar/1436-comparative-superlative-adjectives-list-from-a-to-z.

Friedman, C. (1989, June). *A General Computational Treatment of The Comparative*. In Proceedings of the 27th Annual Meeting on Association for Computational Linguistics (pp. 161-168). Association for Computational Linguistics.

"FTP Service." *Current Neurology and Neuroscience Reports*., U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/pmc/tools/ftp/.

Ganapathibhotla, M., & Liu, B. (2008, August). *Mining Opinions in Comparative Sentences*. In Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1 (pp. 241-248). Association for Computational Linguistics.

Houngbo, H., & Mercer, R. E. (2012). *Method Mention Extraction from Scientific Research Papers*. Proceedings of COLING 2012, 1211-1222.

Jindal, N., & Liu, B. (2006, July). *Mining Comparative Sentences and Relations*. In AAAI (Vol. 22, No. 13311336, p. 9).

Jindal, N., & Liu, B. (2006, August). *Identifying Comparative Sentences in Text Documents*. In Proceedings of the 29th annual international ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 244-251). ACM.

Joshi, R. (2016). *Accuracy Precision Recall F1 Score*: Interpretation of Performance Measures. Retrieved April 01, 2018.

Kennedy, C. (2004). *Comparatives, Semantics Of*. Concise Encyclopedia of Philosophy of

Language and Linguistics, 68-71.

Kessler, W. (2014). *Improving the Detection of Comparison Arguments in Product Reviews*. In Proceedings of Informatik 2014, 2311-2316.

Likert, R. (1932). *A Technique for The Measurement of Attitudes*. Archives of Psychology, 140, 1-55.

LexAccess. (2016). Retrieved November 02, 2018, https://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/lexAccess/2013+/web/download.html.

Lin, C. F., & Wang, S. D. (2002). *Fuzzy Support Vector Machines*. IEEE Transactions on Neural Networks, 13(2), 464-471.

Park, D. H., & Blake, C. (2012, July). *Identifying Comparative Claim Sentences in Full-Text Scientific Articles*. In Proceedings of The Workshop on Detecting Structure in Scholarly Discourse (pp. 1-9). Association for Computational Linguistics.

PHIL. (2012). *Modifiers*. Retrieve 2016.

Saritha, S. K., & Pateriya, R. K. (2014). *Methods for Identifying Comparative Sentences*. International Journal of Computer Applications, 108(19), (pp.23-26).

Schwartz, A. S., & Hearst, M. A. (2002). *A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text*. In Proceedings of Biocomputing 2003 (pp. 451-462).

Scheible, S. (2008). *Annotating Superlatives*. In the International Conference on Language Resources and Evaluation, (pp. 28-30).

Scheible, S. (2007). *Towards A Computational Treatment of Superlatives*. In Proceedings of the ACL 2007 Student Research Workshop, (pp. 67-72).

Software Stanford Parser. Retrieved October 21, 2018, https://nlp.stanford.edu/software/lex-parser.shtml

Solt, S. (2015). *Measurement Scales in Natural Language*. Language and Linguistics Compass, 9(1), 14-32.

Stechow, A. V. (1984). *Comparing Semantic Theories of Comparison*. Journal of Semantics, 3(1-2), 1-77.

Stevens, S. S. (1971). *Issues in Psychophysical Measurement*. Psychological Review, 78(5), 426.

Stevens, S. S. (1946). *On the Theory of Scales of Measurement*. Bobbs-Merrill, College Division.

VerbNet v3.2. (2013). Retrieved November 02, 2018, http://verbs.colorado.edu/verb-index/vn/calibratable_cos-45.6.php#calibratable_cos-45.6-1

Wang, L. (Ed.). (2005). *Support Vector Machines: Theory and Applications* (Vol. 177). Springer Science & Business Media.

Yang, S., & Ko, Y. (2009, August). *Extracting Comparative Sentences from Korean Text Documents Using Comparative Lexical Patterns and Machine Learning Techniques*. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers (pp. 153-156). Association for Computational Linguistics.

Yang, S., & Ko, Y. (2011, June). *Extracting Comparative Entities and Predicates from Texts Using Comparative Type Classification*. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 1636-1644). Association for Computational Linguistics.

Yang, S., & Ko, Y. (2011). *Finding Relevant Features for Korean Comparative Sentence Extraction*. Pattern Recognition Letters, 32(2), 293-296.

Your dictionary. (2015). *100 Adverbs*. Retrieved October 15, 2018, http://www.yourdictionary.com/index.ph

Zhang, H. (2004). *The Optimality of Naive Bayes*. In: Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference, AAAI Press, (pp. 562-567).

Appendix A

Lexicons and Terms

A.1 Specialist lexicon terms

able	abrupt	acerb	achy	acute
airy	ample	angry	antiphase	apt
ashy	atonal	bad	baggy	bald
balky	balmy	bare	barky	base
batty	bawdy	beachy	beady	beaky
beefy	beery	bendy	better	big
Bismarck brown	bitchy	bitter	black	bland
bleak	bleary	blind	bloaty	blobby
blond	bloody	blotchy	blue	blunt
blurry	boggy	bold	bony	bossy
bouncy	brainy	branchy	branny	brash
brassy	bratty	brave	brawny	breathy
breezy	brief	bright	briny	brisk
bristly	brittle	broad	broody	brown
browny	brushy	brusk	bubbly	bucky
buggy	bulgy	bulky	bumpy	bunchy
burly	burpy	burry	bursty	bushy
busty	busy	buttery	buxom	calm
canny	catchy	chaffy	chalky	chancy
chatty	cheap	cheesy	cherty	chesty
chewy	childproof	chilly	choice	choppy
chubby	chunky	churchly	civil	cladogenetic
clammy	clarion	classy	clean	clear
clever	cliffy	clingy	cloddy	close
cloudy	clubby	clumsy	clunky	coarse
cobbly	cobwebby	cocky	cold	comfy
common	compleat	contrasty	cool	corky
corny	costly	cosy	countercultural	courtly
couth	coy	crabby	crackly	crafty
craggy	crampy	cranky	crappy	crass
crawly	crazy	creaky	creamy	creepy
crinkly	crisp	crispy	croaky	cross
crude	cruel	crumbly	crunchy	crusty
cuddly	curdy	curly	curt	curvy
cushy	cute	cyan	dainty	damned
damp	dandy	dank	dark	dark-eyed
dead	deadly	deaf	dear	deep
deft	dendriticlike	dense	dewy	dicey
dim	dingy	dire	dirty	disputant
distinct	dizzy	dodgy	doggy	dopy
dotty	doughty	doughy	dour	dowdy

downy	drab	drafty	dreamy	dreary
drippy	droopy	droughty	drowsy	druggy
drunk	dry	dull	dumb	dumpy
dusky	dusty	dwarf	early	earthy
Eastern Orthodox	easy	eery	electron poor	emissary
empty	exact	faddy	fain	faint
fair	fancy	far	fast	fast acting
fast growing	fast moving	fat	fatty	faulty
feeble	feisty	fewer	fiddly	fierce
fiery	filmy	filthy	fine	firm
fishy	fit	fizzy	flabby	flaky
flashy	flat	fleecy	fleshy	flicky
flighty	flimsy	flinty	flip	floaty
floppy	fluffy	fluky	foamy	foggy
folksy	fond	foul	foxy	frail
frank	freaky	free	fresh	friendly
frilly	frizzly	frizzy	frosty	frothy
frowsty	frowsy	fruity	frumpy	full
funky	funny	furry	fussy	fuzzy
game	gappy	gassy	gauche	gaudy
gaunt	gawky	gay	gentle	ghastly
giddy	girly	glad	glassy	glib
glitzy	gloomy	gluey	glum	golden
goodly	gooey	goosy	gory	gouty
grainy	grand	grassy	grave	gray
greasy	great	greedy	green	grim
grimy	gritty	groggy	groovy	gross
grouchy	grubby	gruff	grumpy	guilty
gummy	gusty	gutsy	hacky	hairy
hammy	handsome	handy	happy	hard
hardy	harsh	hasty	haughty	hazy
heady	healthy	healthy appearing	hearty	heavy
hefty	herby	high	high density	high-caliber
high-efficiency	high-mortality	high-priority	high-quality	high-resistance
highrisk	hilly	hip	hippy	hoar
hoarse	hoary	hollow	holy	homely
homy	hooplike	hoppy	horny	horsy
hot	huge	humble	humpy	hungry
husky	hypertriploid	icy	idle	iffy
imbecilic	impure	inappetent	inapt	infelicitous
instinctual	intense	jaunty	jerky	jiggly
jittery	joky	jolly	juicy	jumpy
keen	kind	kingly	kinky	knobbly
knobby	knotty	lacy	laky	lame
lanky	large	late	lax	lazy
leafy	leaky	lean	leary	lengthy
less	lesser	lewd	light	light skinned

likely	Lilliputian	limp	limy	little
lively	loamy	lofty	lonely	long
longlasting	long-lived	long-living	loopy	loose
loud	lousy	lovely	low	low valued
lowdensity	low-efficiency	lowfat	lowgrade	low-heat
lowly	lown	low-priced	low-profile	low-risk
loyal	lucky	lumpy	luny	lush
mad	malty	mangy	manly	marshy
massy	mature	mealy	mean	meaty
meek	mellifluous	mellow	mendicant	mere
merry	meshy	messy	mettlesome	mighty
mild	milky	minatory	minty	minute
miry	misty	modern	moist	moldy
moly	moody	more	mossy	most
mousy	muddy	mule foot	murky	mushy
mussy	musty	muzzy	nappy	narrow
nasty	natty	naughty	near	neat
needy	nerdy	new	newsy	nice
nifty	nitty	noble	noisy	nonclose
nonlow	nonwoody	nosy	nubby	nude
numb	oaky	obfuscatory	obtuse	odd
often	oily	old	oozy	over-clean
overhasty	overlarge	overripe	overthin	overwet
painty	palatial	pale	palmy	pappy
paraparetic	pasty	patchy	patternless	paunchy
peachy	peaky	pearly	peaty	pebbly
peppy	perichondral	perky	pesky	petty
phlegmy	picaresque	picky	piggy	pink
pipid	pitchy	pithy	placentate	plain
pleasant	plucky	plump	plush	podgy
pointy	poky	poor	portly	posh
potty	pouty	preachy	predeterminate	preppy
pretty	pricey	prickly	prim	princely
prissy	profound	prompt	prone	proteomic
proud	pseudogranulomatous	psycholinguistic	puckery	puddly
puffy	pulpy	punchy	pure	purple
purply	quaint	quarter-hourly	queasy	queer
quick	quick acting	quiet	quirky	quivery
racy	rainy	randy	rapid	rare
raspy	ready	real	red	remote
resupinate	rich	ridgy	right	ripe
ripply	risky	rocky	roomy	ropy
rosy	rough	round	rowdy	ruddy
rude	runny	runty	rusty	rutty
sabulous	sad	safe	saintly	salty
sandy	sane	sappy	sassy	saucy
savvy	savy	scabby	scaly	scant
scanty	scarce	scarry	scary	scrappy
scratchy	scrawny	scrubby	scruffy	scurfy
scurvy	seamy	secure	sedate	sedulous

seedy	sententious	severe	sexy	shabby
shady	shaggy	shaky	shallow	shaly
shapely	sharp	sheer	shelly	shifty
shiny	shoddy	short	shortlived	shortterm
showy	shrewd	shrill	shrubby	shy
sick	sickly	sightly	silky	silly
silty	simple	sincere	sketchy	skimpy
skinny	skunky	slack	slangy	sleazy
sleek	sleepy	slender	slick	slight
slim	slimy	slinky	sloppy	sloughy
slow	sludgy	slushy	sly	small
small bodied	smart	smeary	smelly	smoggy
smoky	smooth	smug	smutty	snaky
snappy	sneaky	sneezy	snooty	snotty
snowy	snuffly	snug	soapy	sober
soddy	soft	soggy	soily	sooty
sore	sorry	sound	soupy	sour
spacy	spare	sparkly	sparse	speedy
spicy	spiky	spindly	spiny	split foot
splotchy	spongy	spooky	sporty	spotty
sprightly	spruce	spry	squally	square
squashy	squatty	squeaky	squiggly	squinty
squishy	stable	staggy	stagy	stale
stanch	starchy	stark	starry	stately
steady	stealthy	steamy	steely	steep
stemmy	stern	sticky	stiff	still
stingy	stocky	stodgy	stoney	stony
stormy	stout	straggly	straight	strange
strawy	streaky	stretchy	strict	stringy
stripy	strong	stubbly	stubby	studly
stuffy	stumpy	stupid	sturdy	subtile
subtle	sugary	sulky	sunny	supernatant
sure	swampy	sweaty	sweet	sweet tasting
swift	swirly	tall	tame	tan
tangy	tardy	tart	tasty	taut
tawdry	tawny	teary	teeny	tender
tense	terse	testy	thermoplastic	thick
thin	thirsty	thorny	threadbare	thrifty
throaty	tickly	tidy	tight	tight fitting
timely	tinny	tiny	tippy	tipsy
toothy	topsy-turvy	touchy	tough	tranquil
trashy	trendy	tricky	trim	trusty
tubby	tubewell	turquoise blue	twangy	twiggy
twirly	twisty	ugly	uneasy	unfriendly
unhappy	unhealthy	unlikely	unlovely	unruly
unsightly	untidy	unwieldy	vague	vain
vast	veiny	vile	viny	void
wacky	wambly	wan	warm	wary
wavy	waxy	weak	wealthy	weary
webby	wee	weedy	weighty	weird

well accepted	well conserved	well directed	well kept	well managed
well matched	well positioned	well predicted	well protected	well proven
well received	well represented	well resolved	well trained	well understood
wet	wheezy	white	wicked	wide
wide-angle	wild	wily	windy	wintry
wiry	wise	wispy	witty	woodsy
woody	wooly	woozy	wordy	wormy
worse	worthy	wrinkly	wrong	yawny
yeasty	yellow	young	yucky	zany
zigzaggy				

A.2 SimDif lexicon terms

alike
also
both
but
comparable
contrary
differ
difference
differences
different
differently
either
equivalent
however
like
likewise
otherwise
same
similar
similarly
though
too
unlike
whereas
while

A.3 Direction Verbs lexicon

abate	abbreviate	accelerate	advance	aggrandize
ameliorate	amplify	appreciate	ascend	attenuate
augment	become stronger	bend	billow	bloat
boost	break down	bring down	broaden	burst
capsize	cave	cheapen	climb	collapse
come	compound	constrict	contract	crawl
crumble	cut	dangle	decelerate	decline
decrease	decreased	deepen	deflate	degrade
depreciate	depreciate	descend	deteriorate	devalue
develop	die down	dilate	dilate	dim
diminish	diminish	dip	distend	dive
double	down	download	drop	dwarf
dwindle	dwindle	enhance	enlarge	escalate
exceed	expand	extend	fade	fail
fall	famish	federate	flatten	flood
flop	fly	gain	get up	go up
grow	hang	harden	hasten	heighten
hike	hip	hoot	hush	ignite
immerse	improve	incline	increase	increased
inflate	inform	intensify	jump	kneel
lengthen	lessen	lift	linger	loll
lollop	lop	lose	lower	magnify
maximize	minimize	molt	mount	multiply
narrow	overflow	overlie	overthrow	overturn
pass	perch	pick up	plop	plummet
plump	plunge	pop up	project	proliferate
propagate	pump	push	push down	push up
quicken	quiet	range	recede	recline
redouble	reduce	regress	retract	retreat
rise	rocket	run down	sag	scramble
screw off	shake off	sharpen	shed	shoot up
short circuit	short-circuit	shorten	show up	shrink
sink	sit	sit down	skyrocket	slack
slacken	slant	slash	slide	slim
slip	slope	slow	slue	slump
smarten	soar	soften	speed up	spill
sprawl	sprout	stand up	steep	steepen
stoop	strengthen	stretch	submerge	supervene
surge	surge	surmount	swarm	swell
thicken	top	topple	tower	transcend
treble	tumble	tumble down	up	upload
volatilize	wane	wax	widen	win
withdraw	worsen			

A.4 Lexicon1 comparative terms

angrier	better	better	bigger	bitterer
blacker	blander	bloodier	bluer	bolder
bossier	braver	briefer	brighter	broader
busier	calmer	cheaper	chewier	chubbier
classier	cleaner	clearer	cleverer	closer
cloudier	clumsier	coarser	colder	cooler
crazier	creamier	creepier	crispier	crueller
crunchier	curly	curvier	cuter	damper
darker	deadlier	deeper	denser	dirtier
drier	duller	dumber	dustier	earlier
easier	fainter	fairer	fancier	faster
fatter	fewer	fiercer	filthier	finer
firmer	fitter	flakier	flatter	fresher
friendlier	fuller	funnier	further/farther	gentler
gloomier	grander	graver	greasier	greater
greedier	grosser	guilter	hairier	handier
happier	harder	harsher	healthier	heavier
higher	hipper	hotter	humbler	hungrier
icier	itchier	juicier	kinder	larger
later	lazier	less	lighter	likelier
littler	livelier	longer	lonlier	louder
lovelier	lower	madder	meaner	messier
milder	moister	more	narrower	nastier
naughtier	nearer	neater	needier	newer
nicer	noisier	odder	oilier	older/elder
plainer	politer	poorer	prettier	prouder
purer	quicker	quieter	rarer	rawer
richer	riper	riskier	roomier	rougher
ruder	rustier	sadder	safer	saltier
saner	scarier	shallower	sharper	shinier
shorter	shyer	sillier	simpler	sincerer
skinnier	sleepier	slimier	slimmer	slower
smaller	smarter	smellier	smokier	smoother
softer	sooner	sorer	sorrier	sourer
spicier	steeper	stingier	stranger	stricter
stronger	sunnier	sweatier	sweeter	taller
tanner	tastier	thicker	thinner	thirstier
tinier	tougher	truer	uglier	warmer
weaker	wealthier	weirder	wetter	wider
wilder	windier	wiser	worldlier	worse
worse	worthier	younger		

A.5 Superlative lexicon

angriest	best	best	biggest	bitterest
blackest	blandest	bloodiest	bluest	boldest
bossiest	bravest	briefest	brightest	broadest
busiest	calmest	cheapest	chewiest	chubbiest
classiest	cleanest	clearest	cleverest	closest
cloudiest	clumsiest	coarsest	coldest	coolest
craziest	creamiest	creepiest	crispiest	cruellest
crunchiest	curliest	curviest	cutest	dampest
darkest	deadliest	deepest	densest	dirtiest
driest	dullest	dumbest	dustiest	earliest
easiest	faintest	fairest	fanciest	fastest
fattest	fewest	fiercest	filthiest	finest
firmest	fittest	flakiest	flattest	freshest
friendliest	fullest	funniest	furthest/farthest	gentlest
gloomiest	grandest	gravest	greasiest	greatest
greediest	grossest	guiltiest	hairiest	handiest
happiest	hardest	harshest	healthiest	heaviest
highest	hippest	hottest	humblest	hungriest
iciest	itchiest	juiciest	kindest	largest
latest	laziest	least	lightest	likeliest
littlest	liveliest	loneliest	longest	loudest
loveliest	lowest	maddest	meanest	messiest
mildest	moistest	most	narrowest	nastiest
naughtiest	nearest	neatest	neediest	newest
nicest	noisiest	oddest	oiliest	oldest/eldest
plainest	politest	poorest	prettiest	proudest
purest	quickest	quietest	rarest	rawest
richest	ripest	riskiest	roomiest	roughest
rudest	rustiest	saddest	safest	saltiest
sanest	scariest	shallowest	sharpest	shiniest
shortest	shyest	silliest	simplest	sincerest
skinniest	sleepiest	slimiest	slimmest	slowest
smallest	smartest	smelliest	smokiest	smoothest
softest	soonest	sorest	sorriest	sourest
spiciest	steepest	stingiest	strangest	strictest
strongest	sunniest	sweatiest	sweetest	tallest
tannest	tastiest	thickest	thinnest	thirstiest
tiniest	toughest	truest	ugliest	warmest
weakest	wealthiest	weirdest	wettest	widest
wildest	windiest	wisest	worldliest	worst
worst	worthiest	youngest		

A.6 SimDif2Word lexicon

another way
in common
much as
much like
on the other hand
one way
same as

A.7 Adverbs lexicon terms

absolutely
badly
barely
completely
dramatically
entirely
exceptionally
extremely
frequently
gradually
highly
incredibly
little
much
notoriously
partially
quite
really
relatively
significantly
simply
slightly
so
somewhat
too
totally
truly
unusually
utterly
very

Curriculum Vitae

Name: Sara Talal Baker

Academic Background:

- **May 2016-April 2019** — Studying Master's degree of Computer Science (Artificial Intelligence) at Western University-Canada-London Ontario.

- **January 2016-April 2016** — Certificate of Achievement from London Language Institute- London Ontario.

- **February 2016** — IELTS Test Certificate.

- **December 2015** — Certificate of graduation from CultureWorks ELS-London Ontario.

- **July 2012** — Open University United Kingdom –the degree of

  Bachelor of Science with Second Class Honours (1st division) in Information Technology and Computing.

- **September 2006-September 2012** — Arab Open University- Jeddah- KSA Bachelor degree in Information Technology and Computing (ITC).

- **September 2006-January 2007** — Arab Open University- KSA Certificate in English language Intensive Course.

Professional Background:

- **May 2013-Jan 2014** Dallah Al-Baraka Holding Company-Experience Certificate as Events and Programs Officer in Dallah voluntary academy.

- **February 2008 -June 2009** Al -Rowad School- Riyadh-KSA-Experience Certificate for teaching English and Computer.

Languages:

- English Fluent.

- Arabic Fluent.

Extra-Curricular activities:

- Experienced in Object Oriented Programming (SQL, C++, JAVA, Web Application, and VB.net).
- Experienced in Object Oriented analysis and design (Information visualization, interface designing skills, human-computer interaction(HCI)).
- Working with different systems (Windows, and Mac OS X)
- Computer skills (Word, Power point, Photoshop, Moviemaker).
- Ability to work with different kinds of diagrams.
- Very good skills in management.
- Good skills in writing reports.
- Ability to work as a team leader.
- A social character who works great with others in life and on social media.