

# **A Study on the Content Analysis Capabilities of Voyant Tools -Targeting the Letters of the Southern Presbyterian Missionary W. Junkin**

Goo, Jiyeon\*

## **I. Introduction**

In the age of Artificial Intelligence (AI), what kinds of relations do the creative intelligence of humanities researchers tend to establish with high-tech software and electronic database? Could the relationship be compared to that of a driver and his car, the former being gradually replaceable by the latter as it equips itself with human attributes? It seems that some humanities researchers still have a strong belief in the superiority of human intelligence and creativity over computer technology for certain reasons. They argue that the use of printed materials compiled by humans is better than machine searches since the

---

\* Yonsei University, Department of English Language and Literature, Doctoral student.

perspectives on a work of literature contained in published bibliographies are unlikely to be duplicated by electronic technologies.<sup>1)</sup> They regard the published lists as the best channel to research sources because of the technical limitation in database coverage.

These days, however, this kind of position is revealing its limitations. This is mainly because of the Digital Revolution, which enabled some high-tech software to equip itself with Artificial Intelligence (AI) that provides researchers with various tools that reach beyond human abilities. Tools such as text visualization can spot patterns, reveal relationships, and shed light on some significant topics that would be inconceivable if it were not for the text mining function of AI. One important example of a web-based tool that exemplifies some of the major changes that have taken place since the start of the Digital Revolution is Voyant Tools. The aim of this paper is to explore the content analysis capabilities of Voyant Tools with the target texts of fifteen letters written by William Junkin, a Southern Presbyterian missionary to Korea during the early periods of the 20<sup>th</sup> century. To address this goal, the following chapters will provide qualitative descriptions of the process of interaction between a human scholar and various functions of Voyant Tools. The study will attempt to propose the ways of utilizing Voyant Tools in analyzing the missionary's letters and suggest implications for future research that deals with building digital archives in humanities projects.

---

1) For further discussions, see Altick, R. & Fenstermaker, J. (1993). *The Art of Literary Research* 4th ed. New York and London: W.W. Norton & Company, 180-81.

## II. Literature Review

Recent publications inform the reader that letters of a missionary are regarded as valuable primary sources in studying the sociohistorical backgrounds of a certain region. Pang and Song (2017) investigate the social impact of Christianity in Gunsan, Korea, by the means of analyzing the missionary correspondence of the Presbyterian Church in the U.S. This study provides insights in understanding the effect of missional works in the church, schools, and hospitals in the region during the early 20s. One of the primary sources in this study includes W. Junkin's several letters contained in *The Missionary*, which shows that the missionary's letters are considered as informative and reliable sources in exploring the region's sociohistorical contexts. However, this research calls for the need to investigate the emotional and psychological aspects of missionaries and believers by conducting in-depth analysis on more historical records (Pang & Song, 2017, 80).

It should be noted that these documents are often extensive both in their amount and in contents, making it difficult for a researcher to carry out a study within the limited resource of time and energy. Chung (2017) deals with early protestant missionary records in Korea, centering on the three types of materials retrieved from the missionary archives of Methodist Episcopal Church, South. Chung discusses the value and signification of these records in the fields of sociopolitical, economical, and cultural studies in Korea. However, regarding the third type of materials, which is *Warren Akin Candler papers, 1846-1977*, he informs that the amount of record is so vast that it is necessary to perform additional research in the future. In fact, *Warren Akin Candler papers* include 600 articles and 800 bundles of letters (Chung, 2017, 7). This may imply that the traditional humanist methods of analyzing text need be complemented

through the aid of computer software tools.

Next, Han and Chang (2017) provide a key guidance in locating the whereabouts of the collections of letters produced by the North American educational missionaries in Korea between 1885 and 1942. This study contains the classification and description of four types of letter collection according to the manner with which the archives retrieved the documents. This study clarifies the extensive number of letter collections which amounts to 61 collections and 34 holding institutions (Han & Chang, 2017, 38), which makes it seem almost impossible to identify their total amount and contents with humanistic approaches. For instance, one of the collections, owned by Presbyterian Historical Society, titled, “United Presbyterian Church in the USA. Commission on Ecumenical Mission and Relations Secretaries’ Files: Korean Mission, 1903-1972”, contains thousands of letters written by missionaries (Ibid., 42).

To address these issues, several attempts are made in the field of humanities to incorporate the use of digital software to undertake a content analysis on vast amount of textual data. This is considered as a critical part in building digital archives. For instance, Miller (2018) assesses content analysis capabilities of Voyant Tools in text mining for digital humanities project, which deals with Tennessee’s history between 1865 and 1965. This article also introduces a few studies that utilized Voyant Tools to extract specific information from massive amounts of field data (see Welsh, 2014; Maramba et al., 2015; Lynch, 2015). These articles report the usefulness and practicality of Voyant Tools for the average humanities researchers, admitting that the tool could be employed as an alternative and complementary means to traditional humanist methods of analyzing text (Miller, 2018, 172).

From this literature review, it is observed that there is a necessity of incorporating the use of the digital device such as Voyant Tools in developing

methodologies for analyzing missionaries' letters, considering their value as a primary source of study in the various fields of humanities.

### III. Methodology

In analyzing the target texts, this research utilizes Voyant Tools, a web-based text analysis device that adds functionality to a document so that a researcher can perceive and understand it in different ways. The Voyant Tools website provides tools such as Cirrus (a word cloud that visualizes the most frequent lexical words of a document), Collocates Graph (a network in which keywords are linked to their collocates, showing each node's frequency and proximity), Topics (a table that displays what topics exist in a text and how each of them is distributed), and RezoViz (a network that shows the connections between people, places and organizations that concur in multiple texts).<sup>2)</sup>

Next, brief research on the author of the target text is undertaken to better comprehend the content of the letters. William McCleery Junkin is a missionary from the Southern Presbyterian Church in the United States, the organization that dispatched the first seven missionaries to Korea, including Mr. Junkin. The missionary was born in 1865 in Virginia, studied English literature at Washington and Lee University and then went to Union Theological Seminary, where he met Horace G. Underwood. At the seminary, Mr. Underwood made a speech to the students about the urgent need of dispatching missionaries to

---

2) For more explanations of tools, see Han, M. (2020). Building a Digital Archive of Letters from Protestant Missionaries in Korea(1884-1942). Paju-si: Bogosa. [In Korean], 160-64; Sinclair, S., & Rockwell, G. (2016c). List of tools. Voyant Tools. Retrieved from <http://www.voyant-tools.org/docs/#!/guide/tools>

Korea. Later, Mr. Junkin married Mary Leyburn and went to Korea with her in 1892. He established the first modern Christian school in Kunsan<sup>3)</sup>, served as a pastor at Cheonju<sup>4)</sup> Church and ran an orphanage in the town. Mr. and Mrs. Junkin lost three sons in 1899 in Korea on account of endemic disease. Mr. Junkin himself died in 1908 of acute pneumonia and typhoid fever.

The target texts analyzed in this research are fifteen letters written by W. Junkin between 1892 and 1905. The places that Mr. Junkin sent his letters from are Seoul, Kunsan, Mokpo, and Cheonju, among which Kunsan occupies the largest number. The recipient of the first four letters is a magazine named *The Missionary*, and that of the other eleven is A.J.A. Alexander, who financially sponsored several major missionaries in Korea including Mr. Junkin. The place where Mr. Alexander received the letters is mainly Spring Station, Kentucky. The list of Mr. Junkin's letters analyzed in this research is presented in Table 1.

Additionally, in performing a content analysis on the fifteen letters through Voyant Tools, four goals are set forth to guide research, which aims to explore: 1) What terminologies appear the most frequently throughout the missionary's first seven letters; 2) What factors influence in interpreting the result of visualization tools; 3) How the Distinctive Words list relates to the actual topics of each of the next eight letters, and 4) Any underlying implications when trying to extract significant data through Voyant Tools.

However, it should be noted that the present research was carried out in two stages, which will be titled as the first (1<sup>st</sup>) and the second (2<sup>nd</sup>) study hereafter. This is because the goals of each study are different, making the approach and methodologies vary consequently. The goals 1) and 2) are addressed in the first

---

3) Spelled as "Gunsan" nowadays.

4) Spelled as "Jeonju" nowadays.

study, whereas the goal 3) is dealt with in the second study, and the goal 4) is explored in both the first and the second study. The target letters for each study are chosen randomly, without any differentiation in their length and contents. This is to limit the variations that could affect the results.

<Table 1> The List of W. Junkin's Letters Analyzed with Voyant Tools

Sender	No.	Date	City	Receiver	Destination	1st Study	2nd Study		
Junkin, William M.	1	1892-11-23	Seoul	The Missionary		O			
	2	1897-05-00	Kunsan			O			
	3	1898-06-15	Kunsan			O			
	4	1902-01-00	Kunsan			O			
	5	1903-00-00	Mokpo	Alexander, A.J.A.	Spring Station, KY		O		
	6	1903-03-26	Kunsan			O			
	7	1903-04-05	Kunsan				O		
	8	1903-04-24	Kunsan			O			
	9	1903-10-28	Kunsan			O			
	10	1904-01-11	Kunsan				O		
	11	1904-04-20	Kunsan				O		
	12	1904-06-03	Kunsan				O		
	13	1904-09-08	Kunsan				O		
	14	1905-02-02	Kunsan				O		
	15	1905-12-15	Cheonju				O		
Total						7	8		

Here is a brief explanation of the major differences between the first and the second studies' methodologies. With the first study, the researcher constructs a single corpus by collating the seven letters and making them into one MS Word file. This is because the first study expects to figure out the major concern

and interests of Mr. Junkin by examining his natural language data throughout the whole text (Goal 1). Also, before processing the corpus via Voyant Tools, the researcher skims through the target letters without necessarily paying close attention to details. This is because the first study attempts to investigate the humanistic elements that may affect the interpretation of visualization tools' results (Goal 2), to see if there is any coincidence or contradiction between the scholar's expectation and the AI tools' outcomes. Concerning the second study, on the other hand, the researcher builds a corpus that consists of eight letters contained in each MS Word files, resulting in eight sub-corpora. The reason is that the second study aims to compare and contrast the outcome of the Distinctive Words list (see Chapter V) and that of the color-coded topics analyzed by the human scholar (Goal 3). To do this, the researcher conducts a line-by-line reading of the eight letters and classifies the contents of texts according to the set of principles (see Chapter V). The more detailed processes of the first and second study and the consequent results are illustrated in the following chapters.

#### IV. Contents and Results of the First Study

In the first study, the researcher skimmed through<sup>5)</sup> seven letters written by W. Junkin and constructed a corpus in the form of a single MS Word file. Next, the corpus was uploaded to Voyant Tools and analyzed accordingly, which took less than 20 seconds. Then the researcher went over the results of the content analysis that appeared on the display panel and reconstructed the contents of the

---

5) According to Oxford Learner's Dictionaries, to "skim" means to "read something quickly in order to find a particular point or the main points".



seven letters. As a result, it was found that Voyant Tools enabled the researcher to ascertain the various types of objective indicators such as the list of vocabulary with top frequencies and the contextual information of certain terms. However, it was discovered that there was room for the scholar's background knowledge, selective perception, and prejudice to be involved in the process of interpreting these indicators. The following case revealed such findings.

To apprehend the major concerns the missionary had in his mind, the researcher read one of his letters contained in the U.S. magazine *The Missionary*, which was issued in April 1902. As reading the document, the researcher noticed the following contents:

During my absence eleven Korean brethren had been preaching in rotation at these places. Three classes for catechumens, or new professors, had been regularly taught, and I found seventy-one people waiting to be examined for baptism. Many of them had husbands, parents, or other kindred and friends in the church, and were the fruits of their work, as well as the proof of their faith.

A poorly clad little boy, who walked in ten miles last Sunday to see a few people baptized, came up not long ago for examination. "Do you know any of the Bible?" was incredulously asked him. But after he had recited the 1<sup>st</sup> Psalm, the Commandments, and would go on to give the 23<sup>rd</sup> Psalm, Matthew 5:3-12, etc., we stopped him, for fear he might begin on something we couldn't recite ourselves.

While eating dinner one day on this baptizing trip a little fellow about four or five-years old toddled in. After we were through, he was told to eat a little rice left over in one of the bowls. It was only a spoonful or so, but, instead of swallowing it down at once, the little fellow bowed his head and said his little "blessing." This is a result of family instruction and example. For whereas family worship seems to be universally neglected at home, it is universally respected in

the Korean church with reading, prayer, and song morning and evening, even in families no member of which can carry a tune. Pastors at home should adopt our rule of not receiving a head of a family into the church until he begins to hold worship in the home. (Junkin, 1902, 194-195)

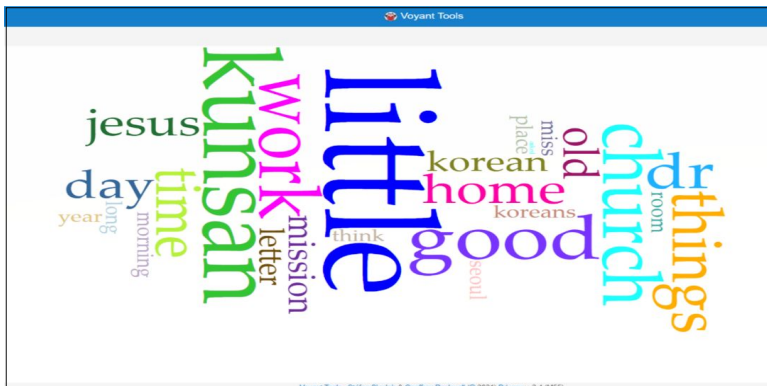
What followed this part of the letter was the news that a site, along with a large pile of bricks, was prepared for the proposed Drew Hospital at Kunsan. The bad news was that the doctor in charge had been forced to leave the field due to his health problems. By the end of the letter Mr. Junkin mentioned that no medical work was possible, and the hospital could not be built unless the doctor came back or someone else was sent.

Although the initial purpose for reading the document was to grasp the main issues in the missionary's life, the research's focus shifted from this goal to the descriptions of anonymous young Koreans who are poor, weary, and humble. The two sentences in the letter were especially noticeable: *A poorly clad little boy, who walked in ten miles last Sunday to see a few people baptized, came up not long ago for examination and while eating dinner one day on this baptizing trip a little fellow about four or five-years old toddled in.* It was regarded noteworthy to see the fact that Mr. Junkin paid special attention to these minor characters when considering sending an overseas letter from Korea to the U.S. in 1902 would not have been an easy task. This unexpected discovery led to academic curiosity in the import of a baptizing trip for Mr. Junkin and the Korean kids when Korea was in a state of destitution. Here, the scholar's compassion and inquisitive mind, which are the elements of subjectivity, were involved in the process of research. Before observing the results of Summary and Cirrus tools, an assumption was made that the missionary had special concerns for the little people and things in Korea.

## 1. Summary and Cirrus

First, the Summary device revealed that the corpus has 5,819 total words and 1,637 unique word forms, and that the average words per sentence are 19. The tool also reported that the most frequent content words in the document are: “little” (32), “Kunsan” (16), “church” (13), “good” (13), and “work” (13).<sup>6)</sup> Then the result of the text visualization device called Cirrus (a word cloud) was consulted under the assumption that the analysis of lexical words that Mr. Junkin frequently used can be an objective indicator of his major concerns. Figure 1 indicates the result of the function visualizing the most frequent lexical words of the corpus (addressing Goal 1).

As regards Cirrus, it locates the words in such a way that the terms that appear the most frequently are positioned near the center with the largest size. As Cirrus’s algorithm processes, the list of vocabulary continues to draw the terms as close as possible to the center so that it could fit the less frequent words within the spaces around the larger words. This visualization enables a researcher



<Figure 1> The Result of Cirrus Tool

6) The number in the parenthesis accounts for the number of occurrences.

to see through the text with intuition and capture the gist and keywords of the whole story. When putting a cursor over a certain word, the frequency count for that term appears in a small box.

As soon as the Cirrus appeared on the display terminal, a verse from the Bible came up in the researcher's mind: "And the King will answer and say to them, 'Assuredly, I say to you, inasmuch as you did it to one of the least of these My brethren, you did it to Me.'" (Matt. 25. 40) It was conjectured that the missionary had a deep concern for "little" things in doing God's work in Korea.

## 2. Contexts

Once the result of the Cirrus tool was apprehended, the outcome of Contexts was examined. It allowed the researcher to explore the surrounding texts of the target word "little", which exhibited the highest occurrences. An attempt was made to discover what kind of little things Mr. Junkin had in mind along with the circumstances in which he used the word. Figure 2 is the result of Contexts that displays each occurrence of the target word with the surrounding texts.

Document	Left	Term	Right
1) Seoul...	high form, a low form and a middle. But cha, cha, cha (little by	little	), as the Koreans say, we hope to get it. Kumsan, May 1897 Just two years
1) Seoul...	bursts at last with a drenching of tears. As the hill crest is reached, our	little	boy looks back and sees things for which our eyes are too dim. But let
1) Seoul...	let the merciful slackness of the cloud hide hearts that are breaking. Thank God, the	little	child looks not ahead at the many separating years. The long voyage ended on a
1) Seoul...	the many separating years. The long voyage ended on a crisp November day. As our	little	coasting steamer cast anchor in the river harbor of Kumsan, Korea, we noticed a number
1) Seoul...	thinking of that suffering that helped me get through the cold today. A poorly clad	little	boy, who walked in ten miles last Sunday to see a few people baptized, came
1) Seoul...	something we couldn't see the outside. While eating dinner one day on this baptizing trip a	little	fellow about four or five years old toddled in. After we were through, he was
1) Seoul...	five years old toddled in. After we were through, he was told to eat a	little	rice left over in one of the bowls. It was only a spoonful or so
1) Seoul...	was only a spoonful or so, but, instead of swallowing it down at once, the	little	fellow bowed his head and said his little "blessings." This is a result of family
1) Seoul...	of swallowing it down at once, the little fellow bowed his head and said his	little	"blessings." This is a result of family instruction and example. For whereas family worship seems
1) Seoul...	one of our members who moved to that field during my furlough in America. The	little	seven years old daughter of the leader at this place is a wonder. When I
1) Seoul...	of the Word and in ability to express themselves. Miss Doty is down for a	little	visit, and with Miss Straeffer and the Bults we had a regular congregation at church
1) Seoul...	spend this first day trying to recall the sad hours of the past week. Our	little	daring did beautifully for the first two weeks and then caught cold, probably from sleeping
1) Seoul...	and we kept him well covered but the air was perhaps too cold for the	little	one to breathe. We became uneasy when he began to sleep constantly and refused to
1) Seoul...	But alas pneumonia set in when the bronchitis left off and eleven o'clock April 22nd (	little	George's birthday) his little heart could stand it no longer he would breathe a little
1) Seoul...	in when the bronchitis left off and eleven o'clock April 22nd (little George's birthday) his	little	heart could stand it no longer he would breathe a little rest, and try again
1) Seoul...	little George's birthday) his little heart could stand it no longer he would breathe a	little	rest, and try again until at one A.M. He gave a little sigh and his
1) Seoul...	would breathe a little rest, and try again until at one A.M. he gave a	little	sigh and his poor baby spirit nestled in Jesus' bosom at his little sorrow over
1) Seoul...	gave a little sigh and his pure baby spirit nestled in Jesus' bosom at his	little	sorrow over but our deepened and made to bleed awhet. Mary and I set by
1) Seoul...	sorrow over but our deepened and made to bleed awhet. Mary and I set by	little	bed till morning. Then we had to tell Edward and Willie and the knife cut
1) Seoul...	morning. Then we had to tell Edward and Willie and the knife cut deeper. Edward's	little	heart ran over but (sprang) Willie only went and put away all his playthings and
1) Seoul...	then went off to himself in the fields. After a while he came with a	little	boy and said "Mama I made this all myself must I put it in little
1) Seoul...	little boy and said "Mama I made this all myself must I put it in	little	brother's coffee? We dug the grave in the corner of our yard above the house
1) Seoul...	corner overlooking the river until it merges with the sea. The dear friends covered the	little	casket with white silk, a calla lily and violets. Mr. Bull read "Suffer the children

<Figure 2> The Result of Contexts Tool

As regards the tool's function, it is possible to decrease or increase the number of surrounding words from 5 to 30. This function was useful in that it allowed for a more precise examination on how the keyword was used in different contexts. As observing the language data, there occurred a necessity to analyze the contexts of the target word "little" to ascertain if they reveal certain dominant patterns. Therefore, the total of 32 concordance lines were analyzed more in detail, with attention to the target word's lexical use and the part of speech. Interestingly, the most frequent occurrences were observed in relation to a tragic event of Mr. Junkin's own son. Table 2 shows the result of this analysis.

According to the result in Table 2, it turned out that the keyword "little" occurred most frequently (14 instances) in the descriptions of a person, an object, or a situation associated with the death and burial of Mr. Junkin's child. The next frequent occurrences were of young Koreans (6 instances), followed by Mr. Junkin's living child (3 instances), his company, and work (3 instances respectively). After consulting the result, the researcher's previous assumption on the missionary's major concern was amended, which was supposed to be about doing little things in Korea for God's sake. Furthermore, it led to a discovery that the human scholar had a tendency to draw a hasty conclusion from his or her previous knowledge, worldview, and the object of compassion when interpreting the visualization results (addressing Goal 2). It was also observed that there was a discrepancy between what the human scholar presupposed from the visualization results of the target texts and what the texts were actually about.

<Table 2> The Result of Contexts Analysis

Case	Contexts
Describing a person, an object, or a situation	- Our <i>little</i> darling did beautifully for the first two weeks and then caught cold, probably from sleeping in a room in which...

<p>related to the death and burial of Mr. Junkin's child (14 instances)</p>	<p>- but the air was perhaps too cold for the <i>little</i> one to breathe. We became uneasy when he began to sleep constantly and refused to eat...</p> <p>- But alas pneumonia set in when the bronchitis left off and eleven o'clock April 22nd (<i>little</i> George's birthday)...</p> <p>- Mary and I sat by <i>little</i> bed till morning. Then we had to tell Edward and Willie and the knife cut deeper.</p> <p>- his <i>little</i> heart could stand it no longer...</p> <p>- he would breathe a <i>little</i> rest, and try again until at one A.M....</p> <p>- he gave a <i>little</i> sigh and his pure baby spirit nestled in Jesus's bosom...</p> <p>- all his <i>little</i> sorrow over but our[s] deepened and made to bleed afresh</p> <p>- The dear friends covered the <i>little</i> casket with white silk a calla lily and violets.</p> <p>- After a short Korean service at the grave we covered the <i>little</i> mound with a white cross of Flowering Almond, Lilac, daisies and verbenas.</p> <p>- After a while he [Willie] came with a <i>little</i> toy and said "Mama I made this all myself must I put it in [little brother's coffin?]"</p> <p>- Mr. Bull read "Suffer the children." It is not the will of God that one of these <i>little</i> ones perish," and "In heaven their angels do always behold the Father's face."</p> <p>- The school-boys and Edward and Willie at their own request were the <i>little</i> pall-bearers. After a short Korean service at the grave...</p> <p>- must I put it in <i>little</i> brother's coffin? We dug the grave in the corner of our yard above the house in a corner...</p>
<p>Describing a person, an object, or a situation related to young Koreans (6 instances)</p>	<p>- A poorly clad <i>little</i> boy, who walked in ten miles last Sunday to see a few people baptized, came up not long ago for...</p> <p>- While eating dinner one day on this baptizing trip a <i>little</i> fellow about four or five years toddled in...</p> <p>- instead of swallowing it down at once, the <i>little</i> fellow bowed his head and said...</p> <p>- The <i>little</i> seven years old daughter of the leader at this place is a wonder. When I asked her if she could...</p> <p>-After we were through, he was told to eat a <i>little</i> rice left over in</p>

	<p>one of the bowls. It was only a spoonful or so, but...          -[the <i>little</i> fellow] bowed his head and said his <i>little</i> "blessings." This is a result of family instruction and example.</p>
<p>Describing Mr. Junkin's children who are alive          (3 instances)</p>	<p>-As the hill crest is reached, our <i>little</i> boy looks back and sees things for which our eyes are too dim          -But let the merciful blackness of the cloud hid hearts that are breaking. Thank God, the <i>little</i> child looks not ahead at the many separating years.          -Edward's <i>little</i> heart ran over but [sprang].</p>
<p>Describing Mr. Junkin's company          (3 instances)</p>	<p>-Mary Virginia is some consolation, tho., for bless her <i>little</i> heart...          -she is just a <i>little</i> bundle of smiles. She is a tiny lady but lie the other precious things...          -One <i>little</i> discouraging feature, she seems always to think that I am on fire, and need putting out!</p>
<p>Describing the events and circumstances related to work (3 instances)</p>	<p>-(about learning Korean language) But cha, cha, cha, cha (<i>little</i> by <i>little</i>), as the Koreans say, we hope to get it.          -while we did not exactly expect the new people to do that we did hope for something to make the silences a <i>little</i> less felt          -Miss Doty is down for a <i>little</i> visit, and with Miss Straeffler and the Bulls we had a regular congregation at church last Sabbath.</p>
<p>Describing objects          (2 instances)</p>	<p>-The long voyage ended on a crisp November day. As our <i>little</i> coasting steamer cast anchor in the river harbor of Kunsan...          -This is the time the Koreans are busy cutting the grass off the hills for fuel thus making the <i>little</i> uncut patches regular traps for the bird s...</p>

## V. Contents and Results of the Second Study

The second study employed a different methodology in processing the data and building a corpus to address Goal 3 (see Chapter III for reference). Here, the researcher read through the eight letters, without speeding up by skimming

them, and carried out color coding according to the following six categories: *time*, *place*, *person*, *object*, *topic*, and *faith*. This categorization, adapted from the proposal by Chang (2021), is one way of identifying key elements that capture the gist of a target text. Next, a corpus was constructed per letter, not the entire target letters as done in the first study. A corpus consisting of eight MS Word files was then uploaded to Voyant Tools and analyzed accordingly, with special attention to the Distinctive Words list provided in Summary. The results of color coding done by the researcher were compared and contrasted with the outcomes of Voyant Tools' analysis, to verify the capabilities of the content analysis and to find out any implications for future research.

### 1. Description of Color Coding Process

First, the left picture of Figure 3 shows one of Mr. Junkin's letters in the form of MS Word file, color coded by the researcher according to the following principle: the "Time" class is colored as yellow, the "Place" class as red, the "Person" class as blue, the "Object" class as green, the "Topic" class as gray, and the "Faith" class as pink. After the color coding process, the researcher managed to classify each color-coded vocabulary into an Excel table as shown in the right picture of Figure 3. For the reference, the category of "Faith" includes expressions regarding the writer's comments on God, prayers, and sermons which often appeared in Mr. Junkin's letters during his missionary work.



1903

My dear doctor **Alexander**:

Since my last to you we have had more trouble and anxiety both here and at **Seoul**. First I had a long and very severe attack of **tonsillitis**. Dr. **Ingold** finally came over but **Myung** woke just as she arrived. **The girls** decided to return to **Seoul** with her and then the very day they were to start the other side of my throat went up and I had to go into a second **illness**. Fortunately this was a short one, and Dr. went over to see her guests in four days to find **Miss Bull** and **Miss Mamy** a little sick and then a **contended** (certified) by **Mrs. G** and stirred us up considerably. Before they left **Dr. Ingold** went to bed with **typhus fever** and is now very ill. How poor Dr. Ingold stands all of this I don't know. At any rate we'll breathe a deep sigh of relief when this summer is over and we again have a **rest** at **Seoul**. I need a **rest** from **Seoul** telling us that **Mrs. F. S. Miller** had just died of peritonitis. It is a sad straggles across to come to a fairlier with three little ones. Pretty **Mrs. McRay** of **Seoul** who came out a bride a year ago has **typhus fever** and of course you have heard of the death of **Mr. Johnson**. She died on her way here at **Seoul**, and he soon after his arrival of **Seoul** of small-pox.

Your short stay among us started the ball of patience to rolling, and we are called on to see and send away some very pitiable cases. Only **this morning** **two young men** fearfully afflicted came in from **Mr. O's** former home **Seoul**. They had both long and great hopes on the "great man" who had come to **Seoul** to cure the sick. It was a hard duty to tell them that all their trouble and suffering in coming here were for naught. I gave them breakfast and some books but to such **difficulties** it seemed to me to be giving a stone when they asked for bread.

I am sending measurements for the water works. I don't know how the plumbing will be done. But if we have only the **land** up on level with **Seoul** it will be a vast improvement.

**The Koreans** are always asking about **the boys** but judging by some stone work that I have had done recently it will cost some 600 **dollars** or so and the road half as much again. **Bill** and I did not like to start on it without letting you know this, and to tell the truth we did not expect you to carry out these plans so long as the way opened closed to your return to **Seoul**.

I feel sure you **will** for us and we sorely need your **papers**. I have no copying paper so fear you will not be able to read the copy enclosed.

**Mrs. Junkin** is very much run down but is still the cheerful busy little **character** you knew. She always wishes to be remembered to you as do the boys.

Write when you can.  
Cordially and truly yours in the work.  
**W.M. Junkin**

No.	Time	Place	Person	Object	Topic	Faith
1	1903	Chunju (3)	Alexander	tonsillitis	sick	pray
2	this morning	the station	Dr. Ingold	my throat	ill	prayers
3		Seoul (2)	The Bulls	sage	died	Christian
4		Wonsan	Mrs. Bull	typhus fever	death	
5		Kobe	Miss Mamy	doctor	cure the sick	
6		Kenju	Mrs. B	telegram	suffering	
7		Kuokmal	Mrs. Harrison	small-pox (2)	the water works	
8		the hospital	Dr. Ingolds	water	cost	
9		Korea	Mrs. F. S. Miller	the bridge		
10			Mrs. McRay	Yen		
11			the Johnsons			
12			two young men			
13			Mr. O			
14			the Koreans			
15			Bull			
16			I (W. Junkin)			
17			Mrs. Junkin			
18			W.M. Junkin			

<Figure 3> The Color-coded Letter(left) and Excel Table(right)

## 2. Summary and the Distinctive Words List

First, the Summary device reported that the corpus has 8,406 total words and 1,983 unique word forms. It also revealed the most frequent content words in the document, listed as: “Dr” (29), “good” (26), “just” (24), “sent” (21), “come” (20), “bull” (18), “little” (18), “boys” (17), “Korea” (17), and “think” (17).<sup>7)</sup> Additionally, the Summary device allowed the researcher to figure out the top five letters in the longest order: L19051215 (1,873), L19040111 (1,721), L19050202 (1,227), L19040603 (1,010), and L19040420 (974). Also, the top five letters in the shortest order are as follow: L19030405 (144), L19030000 (548), L19040908 (909), L19040420 (974), and L19040603 (1,010).<sup>8)</sup>

In the second study, it should be noted that the list of Distinctive Words in the Summary device is utilized substantially to address Goal 3. The

7) The number in the parenthesis accounts for the number of occurrences.

8) The name of each letter’s corpus is titled according to its production date, with the capital letter “L” in the front: LYYYYMMDD. The number in the parenthesis is the word count.

Distinctive Words list provides unique occurrences of certain vocabulary compared to the rest of the corpus, which explains why each letter, not the entire letters, was built into a corpus. Also, the list of Distinctive Words shows difference from the most frequent content words in the document, as the list provides the user with the expressions that stand out in a specific letter. For example, Mr. Junkin's letter written in 1903 reveals that the following words appear uniquely in the text in question: *throat, stone, pox, came, Mrs, trouble, short, return, died, and works*. These words are not likely to be found dominantly in the rest of the letters. Table 3 exhibits the list of prominent words in each letter compared to the rest of the corpus.

&lt;Table 3 &gt; The List of Distinctive Words

Letter	Distinctive Words Compared to the Rest of the Corpus
L19030000	throat (2), stone (2), pox (2), came (3), Mrs (6), trouble (2), short (2), return (2), died (2), works (1)
L19030405	avril (2), suffering (2), wrong (1), unexpectedly (1), U.S.A (1), top (1), tonight (1), rtard (1), presentation (1), postmark (1)
L19040111	Xmas (4), loss (4), game (4), points (3), geese (3), door (3), hedges (4), donkey (4), yesterday's (2), winter (2)
L19040420	reason (4), missionaries (4), feet (3), dogs (3), black (3), white (4), war (4), walks (2), tribe (2), threaded (2)
L19040603	inch (3), comfort (4), view (2), timber (2), inside (2), gone (2), angry (2), affliction (2), outfit (3), building (3)
L19040908	bamboo (4), cable (3), tank (3), sight (2), opportunity (2), necessary (2), ka (2), fence (2), elected (2), boat (2)
L19050202	city (7), Daniel (4), K.S. (3), class (3), can't (3), say (4), trip (3), preaching (3), preach (3), Miss (3)
L19051215	Forsythe (8), week (4), priest (4), came (8), Tate (5), Japs (5), taken (3), I'll (3), debt (3), check (3)

Note. The number in the parenthesis represents the number of occurrence. Also, the raw data in Voyant

Tools is case insensitive.

At first glance, it seemed that the advantage of utilizing the Distinctive Words list was to figure out the unique subjects or topics of each letter efficiently, without having to read it line by line. However, it was necessary to ascertain the accuracy of each entry of Distinctive Words to check if it reliably reflects the prominent topics of the letter in question. Therefore, the eight lists were compared and contrasted with the color-coded Excel tables produced by the researcher. Due to the limitation of space, this paper will provide three instances that demonstrate a closer look on the comparison process. The following three letters are chosen, which vary in their length: L19030405 (144 words), L19040420 (974 words), and L19051215 (1,873 words). This variable in length is included to gain additional information on the consistency level of each entry contained in the Distinctive Words list and the color-coded topics.

First, the Distinctive Words list of the letter written in April 5, 1903 was compared with the color-coded results indicated in Figure 4. The letter’s list of prominent words that are uniquely used in the text are: *avril*, *suffering*, *wrong*, *unexpectedly*, *U.S.A*, *top*, *tonight*, *rtard*, *presentation*, and *postmark* (underlined by the author). The word “rtard” in the list seems to be a spelling mistake by the writer.

No.	Time	Place	Person	Object	Topic	Faith
1	April 2nd	SPRING STATION, Kentucky, U.S.A.	A.J.A. Alexander	fever	suffering (2)	The good Lord
2	tonight	Kunsan, Chumulpo, Coree, Avril	Doctor (AJAA)			
3			<u>Young gentlemen</u>			
4			Dr.			
5			Dr. I.			
6			<u>all our servants</u>			
7			Junkin			

<Figure 4> The Result of Color Coding (L19030405)

As a result, it was discovered that the vocabulary “suffering” appeared both in the Distinct Words list of Voyant Tools as well as in the Excel table under

the Topic class (cf. suffering (2)). For one's reference, the two expressions 'Young gentlemen' and 'all our servants' under the Person class were underlined and boldfaced. This is because in Mr. Junkin's letters, Koreans often appeared in anonymous terms, not as proper nouns. Therefore, special marks were placed when these indicative expressions signify Koreans to compensate for the fact that they were referred to anonymously.

Second, the Distinctive Words list of the letter written in April 20, 1904 was compared with the respective color-coded results presented in Figure 5. Voyant Tools reported that the list of Distinctive Words that are prominently used in the text are: *reason*, *missionaries*, *feet*, *dogs*, *black*, *white*, *war*, *walks*, *tribe*, and *threaded* (underlined by the author).

No.	Time	Place	Person	Object	Topic	Faith
1	April 20th, 1904.	Kunsan, Korea	doctor (AJAA)	your things	a mistake	(N.P./these/the/0) missionaries (4)
2		Spring Station, Ky	1 (W. Junkin) (6)	the pump rod	apparatus	mission work
3		Mokpo	Bell	(200) yen	pipe	the mission
4		Ky	Chang	the pump (2)	the war (3)	
5		Kunsan (2)	the Jap	water	A transport	
6		Southern Korea	Bull (2)	your correspondence	send	
7		our port	we (Bull and Junkin)	the Committee	the Chino-Japanese war	
8		the Philippines	the Hedges	railroad material	(Korean precedent)	
9		Korea (5)	a Japanese guard	the Observer	dog	
10		Sunchun	Russians	the Cemetery	vaccinated	
11		Pveng Yang	us (on the spot)	(\$22/fifty) gold (2)		
12		the North	the Koreans	a badger		
13		the garden	<b>thirteen men</b>	some animal		
14		the village	the boys	Alexander / Alex (dog) (4)		
15			a Korean	Bess		
16			Edward	(Bull's two/the) dogs (2)		
17			Jack	the whole tribe		
18			Mr. Bull	his tribe		
19			papa (Junkin?)			
20			Mrs. Owen			
21			Preston			
22			W.M. Junkin			
23			Mr. O			
24			<b>the babies</b>			
25			<b>the kaksies (ladies)</b>			

Note. The number in the parenthesis represents the number of occurrence.

<Figure 5> The Result of Color Coding (L19040420)

It was observed from the study that the word "war" appeared both from the list of Distinctive Words and from the color-coded Excel table, under the Topic class (cf. the war (3)). This is a significant finding because this concurrence

informs that the content analysis capabilities of the Voyant Tools are partly reliable. Additionally, the letter in question conveyed the news that the Korean Peninsula was under the influence of the Russo-Japanese War. The letter included discussions on relocating the missionaries in Korea, including Mr. Junkin, to other countries such as the Philippines. It was notable that the issue of the missionaries' relocation was reflected consistently in the list of Distinctive Words and the color-coded topics under the Faith class (cf. missionaries (4), mission work, the mission).

It should be mentioned, however, that not all the entries in the list of Distinctive Words accurately reflected the subject matters of a letter in question. This was because the function of the Distinctive Words was not to exhibit the major topics that were frequently commented on but the expressions that were prominent in the target letter, compared to the rest of the letters. It is, therefore, recommended for the user to examine the result of Cirrus, a word cloud that reflects the most frequent lexical words of a document, to grasp the major subjects in the target letter and then to go over the list of Distinctive Words to identify topics that occur uniquely in the text.

Third, the Distinctive Words list of the letter written in December 15, 1905 was compared with the respective color-coded results exhibited in Figure 6. The list of Distinctive Words that were particularly mentioned in the letter are: Forsythe, week, priest, came, Tate, Japs, taken, I'll, debt, and check (underlined by the author).

No.	Time	Place	Person	Object	Topic	Faith
1	12/15 1905	Chunju	Dr. (AJAA)	your letters	write	pastor
2	18-Dec	Kansas City	Mrs. Reynolds	your wedding presents	your (AJAA's) anniversary	a Church
3	last week & this week	the U.S.	(a returning) missionary	the old mission houses	the happenings here	God's leading
4	yesterday	Seoul (5)	Dr. Forsythe	(15,000) yen	sell (us) back	faith
5	tomorrow	Corea	he (Forsythe)	(the) money (3)	pay	the church
6	Sunday	the station (2)	Hounshell	houses	borrow	asked the Lord
7		Kunsan	McFarland	price	I got deeper & deeper in debt	the Lord sent
8		Chunju Church	Forsythe (6)	(660) gold	(the Japs) buy	God is good to us
9		New York & VA	the government	(the / our) church (3)	(the Japs) raised prices tenfold	(260) have been examined
10		the country (3)	a Brother	(300) yen	add to (our church) lot (경매)	preach
11		the city	Ad. Int Committee	his (Forsythe's) salary	the building	Romanism
12		the hospital	the Korean officials	a letter	lot	An English priest
13		the Chunju field	I	(our little) orphanage	cancel the church debt	the priest
14		Chemulpo	the Japs (5)	orphanage funds	his personal check	Blest
15		Kangwha	the natives	(unasked) a check	rounding up the check	the Man of God
16	the coast of Wonsan	(New York & VA) friends	an armed band	would not go into debt for myself	theological classes	
17	PyeongYang	Judge Phlegan of Va.	guns	filled the bill	the priests' compound	
18	the Tate's place	my home church	ammunition	the price	preaching	
19	here	he (Judge Phlegan) (3)	the station	shot	the mission	
20	this place	a friend on Wall St	pump	arrest & beat		
21	K.C.	F. (Forsythe) (5)	water (3)	Love		
22		the committee	dispensary	sickness--the result of work or worry		
23		the orphans	D.J.' house	a fire in Forsythe's kitchen		
24		a brand new baby	the well	his lady love		
25		a country man	hurting (hunting?)	needs a good wife		
26		a patient	yen (600)	a wife for Forsythe		
27		the man		a crisis		
28		Mr. Tate (2)		we shall be unable to work in...		
29		A Romish priest		their (Japanese)' misdeeds		
30		a Korean		an ideal school site		

&lt;Figure 6&gt; The Result of Color Coding (L19051215)

It was observed from Figure 6 that the name of Dr. Forsythe was frequently mentioned under the Person class. This was consistent with the analysis result of the Distinctive Words list of Voyant Tools which reported that the word ‘Forsythe’ appeared eight times (see Table 3). In fact, this letter included an unfortunate story that happened to Mr. Junkin’s church. According to the letter, Mr. Junkin and other missionaries worked hard to prepare a building to use as a church. But the Japanese bought the lands near the building, causing the cost of the surrounding area to rise higher. As a result, the missionary became increasingly in debt.<sup>9)</sup> To solve this problem, Mr. Junkin wrote a letter to his home church, where Judge Phlegan consequently called for help from his friend who was working on Wall Street.<sup>10)</sup> Mr. Junkin’s difficulty in finance was solved miraculously when the friend of Judge Phlegan wrote a check to the

9) Refer to the expressions, ‘I got deeper & deeper in debt,’ ‘(the Japs) buy,’ ‘(the Japs) raised prices tenfold,’ and ‘add to (our church) lot’ under the Topic class in Figure 6.

10) Refer to the expressions, ‘my home church,’ ‘Judge Phlegan’ and ‘a friend on Wall St’ under the Person class in Figure 6.

missionary.<sup>11)</sup> The letter stated that Mr. Junkin's church used this check to pay off its debts.<sup>12)</sup>

In this case, it was observed that the story's key expressions were included in the list of the Distinctive Words, such as *Forsythe*, *priest*, *Japs*, *taken*, *debt* and *check*. It revealed the fact that this tool has the function of extracting important information from the target text in an efficient manner (addressing Goal 3). This is significant because it enables a researcher to figure out major concepts or events that are particular in each letter without having to read it meticulously and comparing it to other letters. According to the research results, it can be assumed that the longer the target text, the more precise the function of the Distinctive Words' capturing the prominent contents of the text.

## VI. Conclusion

Thus far, this paper investigated the content analysis capabilities of Voyant Tools through fifteen letters written by the Southern Presbyterian missionary W. Junkin between 1892 and 1905. The results and findings of the study are as follows. First, the Summary tool reported that the most frequent content words in the first seven letters are *little*, *Kunsan*, *church*, *good*, and *work*. Second, the major element that affected the interpretation of the result of Cirrus turned out to be the human researcher's subjectivity that is derived from his or her worldview, prejudice, and background knowledge (Chapter IV). To be specific, the researcher's focus on the missionary's depiction of anonymous young

---

11) Refer to expressions, 'his personal check' and 'rounding up the check' under the Topic class in Figure 6.

12) Refer to the expressions, 'cancel the church debt' under the Topic class in Figure 6.

Koreans led to the subjective apprehension of the Cirrus result. Third, it was discovered that the Distinctive Words list in the Summary tool can detect a part of topics that are unique to the letter in question. This is evident from the three demonstrations which compared the list of Distinctive Words with the color coded categorization carried out by the researcher (Chapter V).

Second, it is proposed that when comprehending the signification of the Cirrus result, it is necessary to consider the possible discrepancy between the researcher's expectation and the actual content of the target text. This is because the visualized word cloud on a display panel is devoid of contexts that are critical in inferring meanings whereas a human scholar tends to apprehend the meanings of words at the moment they appear on the screen. Third, the study results imply that a humanities researcher can save time and energy through the use of the Distinctive Words in Summary, because he or she need not go through the whole texts line by line to extract key themes of each letter that are specific to the text, in comparison to the rest of the texts. Furthermore, the Contexts tool is expected to provide an in-depth exploration on concordances which can be easily refined according to a scholar's area of interest. This is possible through typing in a keyword to the search window embedded in the tool.

Given the fact that a missionary's letters are regarded as a valuable primary source of information for studying humanities, it is the hope that this research can provide a qualitative illustration on how a digital device can aid a scholar in performing text analysis on the vast amount of research materials.

**[주제어]** 디지털 아카이브, 내한 선교사 편지, 보이언트 툴, 텍스트 마이닝, 디지털 인문학



[Bibliography]

- Altick, R. & Fenstermaker, J. (1993). *The Art of Literary Research* 4th ed. New York and London: W.W. Norton & Company.
- Chang, D. (2021). *Ontology Practicum in Yonsei University* [PowerPoint slides]. LearnUs. <https://www.learnus.org>
- Chung, J. (2017). Early protestant missionary documents in Korea and their implications for further study - Focusing on missionary records of 'Methodist Episcopal Church, South' -. *The Journal of the Humanities*, 111(0), 5-31. [In Korean]
- Han, M. & Chang, Y. (2017). Letters by educational missionaries to Korea in North American archives: 1885- 1942. *The Journal of the Humanities*, 111(0), 33-61. [In Korean]
- Han, M. (2020). *Building a Digital Archive of Letters from Protestant Missionaries in Korea(1884-1942)*. Paju-si: Bogosa. [In Korean]
- Junkin, W. M. (1893, March). KOREA. *The Missionary*, 116-117.
- Junkin, W. M. (1897, May). Korea: an Alabaster Box. *The Missionary*, 223-224.
- Junkin, W. M. (1898, June). LETTER FROM MR. JUNKIN. *The Missionary*, 467.
- Junkin, W. M. (1902, April). Home-Coming to the Morning Calm Country. *The Missionary*, 194-196.
- Junkin, W. M. (1903a). [Letter to A.J.A. Alexander dated 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1903b). [Letter to A.J.A. Alexander dated Mar. 26, 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1903c). [Letter to A.J.A. Alexander dated Apr. 5, 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1903d). [Letter to A.J.A. Alexander dated Apr. 24, 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1903e). [Letter to A.J.A. Alexander dated Oct. 28, 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.

- Junkin, W. M. (1904a). [Letter to A.J.A. Alexander dated Jan. 11, 1904]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1904b). [Letter to A.J.A. Alexander dated Apr. 20, 1904]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1904c). [Letter to A.J.A. Alexander dated Jun. 3, 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1904d). [Letter to A.J.A. Alexander dated Sep. 8, 1904]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1905a). [Letter to A.J.A. Alexander dated Feb. 2, 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Junkin, W. M. (1905b). [Letter to A.J.A. Alexander dated Dec. 15, 1903]. "Alexander Family Papers, 1509-1969". MSS 93, Kentucky Historical Society, Frankfort, KY.
- Lynch, T. L. (2015). Soft(a)ware in the English classroom. *English Journal*, 104, 74.
- Maramba, I. D., Davey, A., Elliott, M. N., Roberts, M., Roland, M., Brown, F., Burt, J., Boiko, O., & Campbell, J. (2015). Web-based textual analysis of free-text patient experience comments from a survey in primary care. *JMIR Medical Informatics*, 3, e20. DOI:10.2196/medinform.3783
- Miller, A. (2018). Text mining digital humanities projects: Assessing content analysis capabilities of Voyant Tools. *Journal of Web Librarianship*, 12(3), 169-197, DOI:10.1080/19322909.2018.1479673
- Pang, S. Y. & Song, J. (2017). The Social impact of the Christianity in Gunsan area - Through the missionary correspondence of the Presbyterian Church in the U.S. -. *The Journal of the Humanities*, 111(0), 63-85. [In Korean]
- Presbyterian Church in the U.S. Executive Committee of Foreign Missions, and Presbyterian Church in the U.S. General Assembly. Committee of Foreign Missions. *The Missionary*. [S.l.: s.n.], 1868-1911.
- Sinclair, S., & Rockwell, G. (2016c). List of tools. Voyant Tools. Retrieved from

<http://www.voyant-tools.org/docs/#!/guide/tools>

Oxford Learner's Dictionaries. (n.d.). skim. In [oxfordlearnersdictionaries.com](http://oxfordlearnersdictionaries.com). Retrieved Nov 15, 2021, from

<https://www.oxfordlearnersdictionaries.com/definition/english/skim?q=skim>

Voyant Tools. (2012). Sinclair, S., Rockwell, G. & the Voyant Tools Team, [voyant-tools.org/?corpus=d6b21fe8b68d3b95443e5c265e34e250](http://voyant-tools.org/?corpus=d6b21fe8b68d3b95443e5c265e34e250). Accessed 19 Apr. 2021.

Welsh, M. E. (2014). Review of Voyant Tools. *Collaborative Librarianship*, 6, 96 – 97.

## [국문초록]

본 연구는 인공지능 텍스트 분석 도구인 보이언트 툴(Voyant Tools)을 통해 남장로교 선교사 W. 전킨의 편지 15통을 분석하여 디지털 인문학의 텍스트 마이닝을 시도하고 보이언트 툴의 텍스트 분석 역량을 고찰하였다. 본 연구의 결과 및 시사점은 다음과 같다. 첫째, 보이언트 툴의 요약(Summary) 기능은 기존의 전통적 읽기 방식을 통해 파악하는 데 많은 시간이 소요되는 정보와 지표를 효율적으로 처리하여 제공하는 이점을 보인다. 한 예로, 2차 연구의 분석 대상은 총 단어 수가 15,000단어에 이르는 텍스트였는데, 이를 보이언트 툴로 분석하는데 약 30초가 소요되는 것을 확인하였다. 둘째, 보이언트 툴을 통해 편지 텍스트의 내용을 파악하는 하나의 방안으로 단어 구름(Cirrus) 기능과 요약 기능의 두 드러지는 단어(Distinctive words) 목록을 교차 검토하는 것을 제안한다. 특히, 대상 텍스트의 단어 수가 많을수록 내용 파악의 정확성이 높아지는 현상을 관찰하였다. 셋째, 연구자가 편지 텍스트에서 확인하고자 하는 특정 맥락이나 표현, 주제가 있을 경우 맥락(Contexts) 기능을 통해 데이터를 선택적으로 추출하여 읽는 방안이 효율적이다.

선교사 W. 전킨의 편지 15통을 대상으로 실시한 연구자의 컬러 코딩 결과와 보이언트 툴의 내용 분석 결과를 비교 및 대조한 결과, 보이언트 툴의 텍스트 분석 역량은 신뢰할 만한 것으로 여겨진다. 내한 선교사 편지 텍스트를 대상으로 디지털 아카이브를 구축할 때, 파악해야 하는 편지의 양이 방대하기에, 보이언트 툴과 같은 인공지능 텍스트 분석 도구를 적절히 활용한다면 연구자의 시간과 같은 자원을 아낄 수 있을 것이라고 파악된다.

[Abstract]

**A Study on the Content Analysis Capabilities of Voyant Tools**  
– Targeting the Letters of the Southern Presbyterian Missionary W. Junkin –

Goo, Jiyeon (Yonsei University)

This study explored the content analysis capabilities of Voyant Tools with fifteen letters written by the Southern Presbyterian missionary W. Junkin between 1892 and 1905. The results and findings of the study are as follows. First, the Summary tool presented that the most frequent content words in the first seven letters are *little*, *Kunsan*, *church*, *good*, and *work*. Second, the major factor that influenced the interpretation of the outcome of Cirrus turned out to be the human researcher's subjectivity that is based on his or her worldview, prejudice and background knowledge. To be specific, the researcher's focus on the missionary's description of anonymous Korean kinds led to the subjective apprehension of the Cirrus result. Third, it was shown that the Distinctive Words list in the Summary tool can detect parts of topics that are particular to the letter in question. This is evident from the three demonstrations from this study, which compared the list of Distinctive Words with the researcher's color-coded categorization.

Next, it is proposed that when comprehending the meaning of the Cirrus result, it is necessary to consider the possible discrepancy between the researcher's expectation and the actual content of the target text. This is because the visualized word cloud on a display panel is devoid of contexts that are critical in inferring significations whereas a human scholar tends to apprehend the meanings of words at the moment they are displayed on the screen. Furthermore, the study results imply that a humanities researcher can save time and energy with the use of the Distinctive Words in Summary, because he or she need not go through the whole texts line by line to extract key themes that

are specific to the target text, compared to the contents in the rest of the corpus. Additionally, the Contexts tool is expected to provide an in-depth exploration of concordances, which can be easily refined according to a scholar's area of interest. This is possible by typing a keyword into the search window embedded in the tool.

Given the fact that a missionary's letters are regarded as a valuable primary source of information for studying the humanities, it is hoped that this research will be a qualitative illustration of how a digital device can assist a scholar with performing text analysis on a vast amount of research materials.

**【Keywords】** digital archives, missionary letters in Korea, Voyant Tools, text mining, digital humanities

논문투고일: 2021.11.17. / 논문심사일: 2021.12.4. / 게재확정일: 2021.12.23.

**【저자연락처】** pavanis@yonsei.ac.kr