An Alternative Method and Display for Content Analysis: Textual Visualization

Güllü Yakar
Necmettin Erbakan University, Faculty of Fine Arts
gyakar@konya.edu.tr

Abstract
The purpose of this study is to describe the usage of textual-based visualization methods as an alternative method in qualitative researches and especially in content analysis process. Qualitative researches –like social media analysis, linguistic or communicational studies that statistical methods can not be applied directly- necessitate usage of special techniques to determine patterns and structures. Textual based visualization techniques allows researcher to analyze the unstructured corpus, explore the inner relations and thus provide meaning to these texts. Decoding this structure can be achieved with content analysis method, and findings can be represented by using matrices or diagrams. However, cases of multidimensional data requires advanced visualization techniques. Data visualisation is graphical edition and common product of the fields of statistics, information technologies, broadcasting and visual communication design. Data of the qualitative research findings about determining the effectiveness and further user expectations of Kastamonu University Kuzeykent Campus wayfinding system were used for analysis and visualization. The assessments of 20 participants are collected by using a semi-structured interview form developed by the researcher. The findings have been subjected to content analysis and displayed with matrice table. These process steps were repeated with three different web-based/online visualization tool.

Keywords: Content analysis, textual data visualization, wayfinding, word cloud, word tree.

Introduction
Data visualization is regulating the numerical or verbal data with graphical representation methods such as charts, maps, graphs, diagrams and tables. With the help of visualization applications; it is possible to process the findings obtained from the researches and make them meaningful, to analyze the structure and relations of the contents, to question the premises and hypotheses. By doing so, it becomes easier to understand the data in complex structure. "When content is illustrated with pictures or figures, the information can be maintained in the mind of the viewer over a period of time" (Fountas and Pinnell, 2001). Arrangements such as timelines, Venn diagram, periodic table; are the best-known examples of data visualization. "Many useful graphical displays result from application of multidimensional statistical analysis techniques such as factor analysis (including principal components analysis) and clustering" (Zinovyev, 2011).

Visualization is perceived much faster than intensive text or numerical data can be. In addition to the advantages it provides to the viewer, it also allows its user to be able to analyze the data with all of its dimensions. In addition to the direct involvement of the user, the main advantages of visual data exploration over automatic data mining techniques from statistics or machine learning are:
  • visual data exploration can easily deal with highly inhomogeneous and noisy data,

• visual data exploration is intuitive and requires no understanding of complex mathematical or statistical algorithms or parameters" (Keim, 2002).

Textual Visualization

Social media analysis, linguistics/communication studies, open-ended interviews; are unstructured textual data involving subjective interpretations of people. Analyzing the words in the text, their repetition frequency and word sequences; leads to an understanding of the structure and thus, to make the text meaningful. Meirelles (2013) states that a text is a categorical type of data and making graphical representations of such data are more difficult because there is no ordering/proportionality in the construction of a text. For this reason, special methods are needed to visualize the qualitative research results.

Cao and Cui (2016) classify text visualization techniques according to their purpose as follows: (1) for visualizing document similarity, (2) for revealing content, (3) for visualizing sentiments and emotions of the text, (4) for exploring document corpus, and (5) for analyzing various domain-specific rich-text corpus, such as social media data, online news, emails, poetry, and prose.

In the application phase of the study,it is aimed to analyze syntactic and lexical structures of the visualized text.For this reason, an explanation is given about visualization tools that provides syntactic and lexical analysis of text.

Relational decoding of a word (Visualizing syntactic structure)
Selected word constitutes the root of visualization and this root diverges to branches enabling the analysis of word with the text. That word tree connects the root word and related sentences with arrows. Word frequencies affect the sequence of branches, their position and size. Punctuation, repeated words and conjunctions together constitute the context, for this reason, reducing the content of the text creates negative results. "A word tree is a graphical version of the traditional *keyword-in-context* method, and enables rapid querying and exploration of bodies of text" (Wattenberg and Viégas, 2008).

2.2. Word's frequency in text (Visualizing lexical structure)
   This visualization shows words in a cloud form. Words' proportional size depend on their frequency in text. For word cloud, the determining element is the font size of words. Words are categorical data and unlike quantitative data, they can not be classified with color values. Typeface and composition provide additional effects. The number of words can adversely affect legibility, therefore eliminating less frequent words can be considered. "Word clouds are useful when you are trying to get a quick and rough sense of some of the dominant keywords used in the text. They can be an option for working with qualitative data during the data exploration stage, more so as a means for reporting analysis to others" (Kirk, 2016).

Methods and Procedures
This study was designed to provide an alternative analysis and presentation method recommendation to the content analysis process. For this purpose, the data gathered from the opinions of the users of Kastamonu University Kuzeykent Campus about the wayfinding process and signs are used. Descriptive research data gathered during 2014-2015 Fall-Winter Semester are evaluated with content analysis, and then, organized with application of visualization methods. Two main terms about gathered data, *wayfinding* and *signage/wayfinding system* can be explain as:
   - Wayfinding refers to information systems that guide people through a physical environment and enhance their understanding and experience of the space (SEGD, 2014).

- Signage/wayfinding system consist of assistant tools such as maps, signs, arrows and pictograms, kiosks, floor/building plans, landmarks which designed and combined to enable people to reach their destinations independently, easily and without wasting time.

Content Analysis Procedure

Interviews, as a qualitative method, are conducted in order to compile campus users' views on their way finding process in terms of their perspective, experiences, feelings and perceptions. The data were collected by using the interview form developed by the researcher in Turkish. As a result of the literature review made during the question preparation phase; signage/way finding systems have seen to be evaluated on the basis of syntactic, semantic and pragmatic levels (Teker, 2002). Designated concepts are transferred to everyday speech; the qualities such as noticeability, legibility, and sufficiency of the signs have been questioned. Moreover, by asking opinions on experiences such as *getting lost, finding easily, loss of time, anxiety* and on points of the system deficiencies/corrections; indirect effects of signs have been tried to be detected. An interview form approach has been adopted with the anticipation that the responses of the participants could change the course of the study.Preliminary studies have been carried out for analysis by selecting important parts such as meaningful words and sentences in the texts derived from the interview records. This kind of analysis, which Yıldırım and Şimşek (2013) characterized as induction, is called *content analysis*. *Codes* as the key words that correspond to the responses of the participants and *themes*as the headings that categorize and group the codes are developed. Matrix notation is used in the presentation of the data. Data entry in the form of this table representation is possible in various forms such as short text blocks, expression quotations, classifications, abbreviations, symbolic figures, marked lines and arrows (Miles and Huberman, 1994). Participants are coded with numbers such as K1, K2, K3 and placed horizontally on the line, themes/ codes forming themes are placed vertically in the column. User joining the statement marked with *x* in the cell where the user and the code coincide, while unmarked cells mean that they do not participate. Cells covered with different colors belong to the participant who has not made assessment because they did not notice the direction signs on the campus. Yıldırım and Şimşek (2013) indicate that the percentage calculations in the systematic digitization of interview data are important.In this direction; the total number of participants who expressed their opinion on the subject was taken and the percentages within the group were calculated and added to the last columns of the matrix. By doing so, a digitized visualization of the content analysis was created.

Table 1. Matrix visualization of content analysis

THEME	CODE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	TOTAL	%
Behavior	Easily finding	x						x						x	x	x		x				6	30
	Noticing signs	x		x			x	x	x		x			x	x	x	x			x		11	55
	Noticing signs later									x		x	x									3	15
	Getting help	x	x	x	x	x	x	x	x	x	x		x	x		x		x	x	x		16	80
Negativity	Having difficulty / failure		x	x	x	x	x		x	x	x	x	x			x	x					12	60
	Anxiety				x			x	x	x	x	x					x					7	35
	Loss of time			x			x				x	x			x	x						6	30
Evaluation of signs	Sufficient in quantity /content	x						x						x	x							4	20
	Insufficient in quantity /content			x			x		x		x	x	x			x	x			x		9	45
	Clear message/ understandable	x					x	x	x			x	x	x	x		x					9	45
	Illegible except from short distance	x		x			x			x	x	x	x			x				x		9	45
	Easy to notice	x					x	x						x		x				x		6	30
	Prominent in color	x					x	x			x									x		5	25
	Prominent in size	x					x	x						x						x		5	25
Expectation	Bigger signs/ text	x		x	x		x	x			x	x	x	x	x			x		x	x	13	65
	Prominent colors		x	x	x	x	x			x			x	x		x	x	x	x	x		13	65
	More indoor signs		x						x	x											x	4	20
	More outdoor signs	x	x				x			x			x	x				x	x	x	x	10	50
	Campus map		x	x	x	x	x	x	x	x	x		x	x			x			x	x	14	70

16

Shaped by personality and experiences, individuals' views on events and phenomena are subjective and unique.The basic principle of qualitative research is to describethis unique form of expression in depht from the point of view of the people. These surveys "provide a very detailed account of how people perceive what is happening in their surroundings and what they mean in the given circumstances" (Frey, Botan and Creep, 1991).In order to reflect in a more striking way the users' experience of navigating the campus and the effects of these experiences; the data were examined through descriptive analysis, with direct cues from responses.

In the beginning of the interview, question of "Did you find the places you did not know/visited for the first time at Kuzeykent Campus easily?" was asked, 30% of the respondents answered positively and stated that they found it easily, 60% answered negatively and stated that they could not find it/had difficulty. The answers are:

  - *I can find it with the help of my friends who are sitting here, I couldn't find when I was alone, but now I have learned, I can find it.* (K5)
  - *I found with the signs and the help I got from people I asked. I found thanks to them. I have not had much trouble.*(K7)
  - *I could not find it, when I first came. Then I get used to it gradually, but by asking, not by looking at the signs.*(K9)
  - *I had problems in some places. For example, I am a student, I have morning classes, it is going on until the evening. I go to cafeteria at lunch breaks. I find it very difficult to find cafeteria alone. The cafeteria is not shown with arrows, so I had to find it by asking.*(K10)
  - *We found it easily because it is a small city. There was no problem. Because we asked to people, to students.*(K18).

The behaviours of the participants to find their ways/directions in the campus and their expectations related to present signage/wayfinding system are seen in the content analysis matrix and interpretations.

Visualization Procedure
It is possible to determine the codes and themes in the text prior to data collection. If codes and themes are to be developed after data collection, researchers can make use of manual methods or can perform analysis automatically by using computer programs. In addition to Atlas. ti, Maxqda and NVivo programs for professional users, there are also web-based/online visualization tools for the average user group. "Many of these visualisation tools and techniques include a welcomed aim of bringing corpus analysis to students, language learners, and researchers from other disciplines who use corpora (Allen, 2017).

Word cloud and word tree visualization applications preferred in the study; since they offer an intuitive use (DePaolo and Wilkinson, 2014) and are compatible with the Turkish language. In addition, both visualizations create static images instead of dynamic images. It is possible to use visualization applications separately to the interview text of each participant. However, the texts have been combined to reach the most common topics of all twenty participants. Thus, a corpus consisting of 4241 words was formed.

Word cloud visualization
The study is based on the hypothesis that high frequency words in the text express important topics for users. Word cloud visualization is a technique suitable for this purpose because it shows the relation between words frequency and size. Color is not a determining/separating element. Similarly, the location of the words in the composition is not a determining factor, too.

Figure 1. Word cloud visualization

The word cloud visualization tool at www.wordle.net is used in the study. Since they do not make any contribution to the analysis, words such as *I think…, partly, another, many, some* have been removed from the visualization. The effect of most descriptive terms such as *find, couldn't find, trouble, help, hard, easy* which is important for research has been increased. The similar words shown in Figure 1 are regulated and number of words is reduced in Figure 2. This visualization can be presented in a different color combination or black and white colors.

Figure 2. Word cloud visualization with reduced words (max. 50)

Word tree visualization
The descriptive terms obtained by the word cloud visualization method summarize the concepts that are most frequently referred to by people. In order to understand the participants' choice of words in context, it is important to examine the sentence setups. Digitization is a relatively inadequate method for understanding social phenomena and especially human behavior. "In order to understand why and how inter-variable relations are established, to elaborate and deeply examine the phenomena" (Demir, 2010), appropriate methods are needed for qualitative data.

"Human dialogue is so complex that definitively analysing patterns of conversation may well be impossible. Within a conversation, all the complexities and ambiguities of natural language exist and each speaker will have his/her own speech characteristics and moods. Examining these characteristics through text dialog can be a demanding cognitive task…This task can be made more convenient if there is a way of visualising all this information at once through graphical patterns. Graphical patterns can revolve around the conversation, creating an abstract piece of artwork. From these patterns, one can guess at the speaker's emotion and how he/she is connected to another speaker during a conversation" (Tat and Carpendale, 2002).

As Table 1 demonstrates, the most common behavior of campus users in their orientation process is getting help from other people (80%). The percentage of participants who stated that they had problems like having difficulty to find or not being able to find was 60%.Word tree visualizations made by using the online tool on www.jasondavies.com/wordtree/for the given themes are as follows:

Figure 3. Word tree visualization for the word *yardım (help [while finding the way])*

Signage/wayfinding systems are assistant tools designed to enable people to reach their destinations independently, easily and without wasting time.People needing help despite the existence of signs, points to deficiency of the system.For this reason, the word *help* was detected in the answers and its place within the structure was examined. The visualization presented in Figure 3 represents the individual expressions of the persons who stated that they had received assistance. The root term *help* and the related statements/branches are:

- *We got help (Yardım aldık),*
- *I found by getting help (Yardım alarak buldum),*
- *I got it done with help (Yardım alarak halletim),*
- *…helped me (…yardım etti),*
- *…isn't helping me (…yardım etmiyor),*
- *I wanted to get help (Yardım almak istedim).*

Visualization presented in Figure 4 contains the answers of people who have stated that they had difficulty in the wayfinding process. The verbs in the Turkish sentence sequence are at the end of the sentence.In order to meet this requirement; visualization was prepared in reverse version.

Figure 4. Reverse word tree visualization for the word *zorlandım (had difficulty [while finding the way])*

The root term *I had difficulty* and the related statements/branches are:

- *Ihad difficulty while finding student administration office (Öğrenci işlerini bulmakta zorlandım),*
- *I had difficulty finding my classroom (Sınıfımı bulmakta zorlandım),*
- *I had difficulty finding while I'm at outdoor (Dışarıdayken bulmakta zorlandım),*
- *I had difficulty to find the building when I came for registration (Kayıt yaptırmaya geldğimde binayı bulmakta zorlandım),*
- *I had difficulty to find the faculties (Fakülteleri bulmakta zorlandım),*
- *I had difficulty although I'm from Kastamonu(Kastamonu'lu olmama rağmen zorlandım),*
- *I had great difficulty the day I came for registration (Kayıt için geldğim gün çok zorlandım),*
- *I had difficulty to read while driving by (Arabayla geçerken okumakta zorlandım).*

It is envisaged that, among the other key terms that can be examined in this direction, the word *more* has significant connotations.The percentage of participants who indicated

that *larger signs/letters* or *more noticeable colors* would be useful was 65%. A visualization made with this term expands with terms *bigger*, *more striking*, *more colorful*. Thus, statements that show the expectations of the users about the signs can be reached.

Conclusions

While there are a large number of visualization methods that can be used in the analysis of quantitative data, there are relatively few options in qualitative data display. It can be stated that these data are usually presented as quotations from the text (in plain text form) or digitized. Text visualization applications with advantages such as easy access, user centricity, intuitive, fast and flexible usage, brought an alternative dimension to the analysis process.

The qualities of the data to be analyzed and the format of display are factors for determining the appropriate type of visualization practice. Word cloud visualization, which provides an explanatory and inductive structure, is example of a commonly used type. Word tree visualization is exploratory and deductive in nature, allowing researchers to search for new root terms in the text. These techniques, which researchers can use simultaneously with content analysis, will contribute to the interpretation and display of content in easy and understandable manner.

References

Allen, W. L. (2017). Making Corpus Data Visible: Visualising Text with Research Intermediaries, *Corpora*, *12, 3*.

Cao, N., Cui, W. (2016). *Introduction to Text Visualization*. Atlantis Press.

Demir, O. Ö. (2010). Nitel Araştırma Yöntemleri. (Ed. Kaan Böke), *inSosyal Bilimlerde Araştırma Yöntemleri*. İstanbul: Alfa.

De Paolo, C. A. and Wilkinson, K. (2014). Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data, *TechTrends*. *58, 3*.

Fountas, I.C., and Pinnell, G.S. (2001). *Guiding Readers and Writers (Grades 3–6): Teaching, Comprehension, Genre, and Content Literacy*. Portsmouth, NH: Heinemann.

Frey, L. R., Botan, C.H. and Kreps, G. L. (1991). *Investigating Communication: An Introduction to Research Methods*. New Jersey: Prentice Hall.

Keim, D. (2002). Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, *7, 1*.

Kirk, A. (2016). *Data Visualization: A Handbook for Data Driven Design*. UK:Sage.

Meirelles, I. (2013). *Design for Information: An Introduction to the Histories, Theories and Best Practices Behind Effective Information Visualizations*. Massachusetts: Rockport.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis*. London: Sage.

SEGD (Society for Experiential Graphic Design) (2014). What is wayfinding? Retrieved from https://segd.org/what-wayfinding onOct. 12, 2017.

Tat, A. and Carpendale, M. S. T. (2002). Visualising Human Dialog, *Proceedings Sixth International Conference on Information Visualisation.*

Teker, U. (2002). *Grafik Tasarım ve Reklam*. İzmir: Dokuz Eylül.

Wattenberg, M., Viégas, F. B. (2008). The Word Tree, An Interactive Visual Concordance. *IEEE Transactions on Visualization and Computer Graphics*, *14, 6*.

Yıldırım, A., & Şimşek, H. (2013). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*. Ankara: Seçkin.

Zinovyev, A. (2011). Data Visualization in Political and Social Sciences. *International Encyclopedia of Political Science* (Eds. Badie, B., Berg-Schlosser, D., Morlino, L. A.), UK: SAGE.