**An Interactive Digital Newspaper Visualizing the Spiritualist Movement in 19th century**

Group 16: Wangjun Qian, Jinrui Wang, Rong Jin, Youwen Zhang, Jingyu Zhou

**1 Introduction**

As people are so used to targeted reading, we feel reluctant to read faced with large textual data. The social movement records are buried in the dust of history. How to preserve these stories and present them to people in the original style has become our focus. Our group used the NLTK toolkit in python to analyze textual data and used quantitative data as evidence to guide and support our reading in the complete dataset. We combine our findings and present an interactive digital newspaper to break down the barrier for the general public to get an overview of the 19th century Spiritualist Newspaper. Through the coursework we unpacked how visualization techniques can help humanity studies and deepened our understanding of the visualization process.

**2 Research Background: The Spiritualist Newspaper dataset**

The Spiritualist Newspaper dataset is provided full access by the National Library of Scotland(NLS), covering newspapers from 1869 to 1882, providing a rich document to explore what happened in the early stage of the Spiritualism movement. The Gothic headlines bear a sense of Victorian history, with few photos but fine illustrations. In our design we would like to preserve the style in a most vivid way.

The Spiritualist is more like a forum journal. Each issue starts with an article directory, followed by commercial advertisement. Followed are notices of upcoming meetings and events organised by multiple associations. The rest contains interview records, event news and latest academic findings. Most activists are from the middle and upper class, including scientists, politicians, dames and mediums. Stories act as their evidence to voice their beliefs in Spiritualism.

**3 Data Exploration Process: Using NLTK toolkit in Python**

**NLTK in OCR corpurs**

In this project we mainly adopted the NLTK toolkit to conduct data analysis. OCR digital scan files require data cleaning to improve accuracy. Then we extract the word tokens for next step categorization. Newspaper data has strong time series characteristics. The inventory of the text file helps us to build dictionaries of different year corpus, ready for next step data analysis.

**An iterative process to find a theme**

We have tried many statistical variables including the highest occurrence of words, lexical dispersion of major cities. Nltk also helps to locate specific key words in the text to quickly understand the context.

The process of data exploration is more like a treasure hunt. It is impossible to know the full picture of the data in a short time, which is even more difficult for our audience. The scattered data provide many possibilities, and we finally determined that key figures are the main story line. We search for stories worth telling from their data abnormalities.

**Locate our narratives and raise questions from data**

We sorted by the frequency of the phrases to find which characters are worthy of our discussion. NLTK provides us with a library of methods. Many names appear frequently. Who are they? Why do they appear frequently? What relationship do they have with this newspaper and the spiritual movement? The results of phrase extraction have brought us unknown content. At this time, we began to comprehend the meaning of programming tools to the humanities database. The calculation makes the patterns become clear at a glance. But at the same time, numbers cannot explain the different meanings of the same word, nor the whole chapter. We follow these clues to explore the story.

**Mining the stories from charts**

We have identified five protagonists of the narrative representing the main promoters of the newspaper: two scientists, a medium, and a social activist. The line charts of their name occurrence throughout the years shows possible story transitions: 1) sharp increase point, 2) most active period, and 3) the sharp decrease point. Programming tools help us locate the point of specific newspapers of their reference where we need to return to the original text. We conduct a reading survey of that specific period supplemented with literature. Clues started to make sense, connections revealed and we have stories to communicate.

**4 Visualization & Communication**

**Audience: Visitors in the National Library of Scotland**

Our audiences are usually general readers who do not have much knowledge of Spiritualism. The original newspaper’s presentation is direct and complicated. For the visitors, it is difficult to find an entry point into this vast dataset, let alone generate interest in exploring the newspaper. We aim to engage the audience to explore the dataset through interesting visualisations that will allow them to build a foundational understanding of the spiritualist newspaper. Also, as the data is at a certain threshold for the general reader, our visualisation needs to quickly build up a general understanding of the data in a clear and concise way for the viewer.

**Design: Interactive digital newspaper**

The product form of our data visualization is an interactive digital newspaper. The newspaper style restores the original newspaper. It can visually enhance the theme, so that the audience can be attracted and immersed in it. The interactive way of exploring the data gives the audience the freedom to interact with the story he/she would like to explore.

Our data visualisation is based on the logic of data exploration, with data charts complemented by relevant historical information to present the whole story to the viewer. Such an approach helps the user to better understand what the data chart is saying. In addition, interesting interactions such as illustrations in which characters speak can engage the viewer in active exploration.

**Interaction process: Communicate with the newspaper**

We want to create a communication between the audience and the newspaper. The whole process can be divided into three parts: before, during and after the interaction. Before the user interacts, the interface scrolls through basic data about the newspaper (year of publication etc.) to give the viewer an overview of the dataset. Interaction buttons can guide the user to the next interaction. Once the viewer interacts with the main page, the content displayed is a newspaper cover built around the figures identified above. We used deepfake technology to make the characters in the illustrations talk, which can engage the user in exploration. When the user is drawn to the person page, they can see a description of the information connected by the newspaper's data. This allows the user to understand the general content of the newspaper through the data.

**5 Conclusion**

The interactive digital newspaper we presented, provided general visitors to the library a chance to explore an overview of the Spiritualist Newspaper dataset by communicating with key figures’ data visualizations and stories. In the process we’ve unpacked how to use programming tools to aid targeted reading, which can lead to our future studies of visualizing texture data.