Danny Eich

Dr. Daniel Anderson

ENGL 480

24 March 2020

Transcript of “Roberto Busa: The Father of Digital Humanities”

In 2020, we take search engines and digital concordances for granted. I can’t remember a time I wasn’t able to Google a question I had. Obviously, that wasn’t always the case. There was a time before iPhones, before Google, before personal computers, and even before business computers. Any research had to be done by hand and was extremely time consuming.

Roberto Busa, an Italian Jesuit Priest and scholar, began researching “What is the metaphysics of presence in St. Thomas Aquinas?” His dissertation focused on just the word “presence” as well as all peripheral words, such as “in the presence.” He created a handmade concordance that consisted of 10,000 cards – all focused on the appearance of the word “presence.” He defended his dissertation in 1946.

Following this project, Busa dreamed of creating a concordance of all of St. Thomas Aquinas’s works. This would require creating 10,000,000 cards. He realized the only way to accomplish his goal would be by finding a mechanical system. So, in 1949, he began searching for the right machine.

After touring 25 different universities, Busa met with Thomas J. Watson, the founding chairmen of IBM. Busa knew that Watson would initially say that IBM would not be able to achieve this task. While waiting to see Watson, Busa saw an IBM poster that read, “The difficult we do right away. The impossible takes a little longer.” Busa took the poster with him to the meeting.

When he sat down to meet with Watson, Busa said, “It is not right to say no before you have tried,” and then showed Watson the poster. Watson agreed to help – provided that he did not change IBM to “International Busa Machines.”

With the partnership established, Busa outlined the five stages to create his concordance. First, transcription of the text as phrases onto individual cards. Second, reproduce the cards so there was the same number of cards as in the phrase. Third, write the individual word on each of the cards. Fourth, alphabetize the cards. And, finally, print out the index for publication.

Let’s do an example to break down this process. For the nursery rhyme “Mary Had a Little Lamb,” the first phrase would be transcribed onto a card – so, “Mary Had a Little Lamb.” This would be duplicated five times, once for each word. Then, each individual word would be written onto the card additionally. So, “Mary” plus “Mary Had a Little Lamb,” “Had” plus “Mary Had a Little Lamb,” and so on. Then, the cards would be alphabetized. And, finally, the results would be printed. Thus, we would have the index of “Mary Had a Little Lamb.”

Rather than using a basic notecard, like in his dissertation, Busa used punched cards, a primitive for of data storage. The punch card system was developed in 1890 by Hollerith for recording the U.S. Census data. His punch card system was inspired by the player piano. Punch cards had a limitation of 80-characters per card, so he wouldn’t be able to fit more than a line per card. For reference, a Tweet’s limit has historically been a 140-characters, so it would require two punch cards to record a single tweet. That’s not much to work with when you have ten-million words to get through.

Due to the volume of punch cards, Busa had to move truckloads of them across Italy whenever changing locations. Today, the quantity of data could easily be stored on microSD card and mailed in envelope, or snuck in a shoe, or more realistically, shared over the cloud and avoid physical transfer all together.

With a theory and storage method in place, Busa had to search across the country to find the right mechanical components. By repurposing a variety of machines, primarily left over from World War II and machines used for accounting, Busa had found the necessary machines to begin his research.

Busa used six machines to complete his initial mechanized test: an automatic punch, a collator, a record interpreter, a reproducer, a shorter machine, and a tabulator. These six machines, along with human intervention, allowed Busa to create his concordance.

The first step was to type in a line from the piece. This was done using a keyboard and an automatic punch. The operator would type in each line twice.

The second step involved comparing the two entries in a machine called the collator. This step helped eliminate human error. In later systems, Busa would have two different people type out the line.

The third step was to print the letters that corresponded to the punches, so that the cards could be read. The record interpreter completed this task.

The fourth step was to replicate the cards and print out the first, second, and so on word on each copy. So, like before, “Mary Had a Little Lamb” plus “Mary,” and then another copy with “Mary Had a Little Lamb” plus “Had.” The aptly named reproducer did this.

The fifth step was to alphabetize the cards, accomplished by the sorter machine.

The final step was to count the total words and print out the sum, which was done by an alphanumeric accounting machine, also known as a tabulator.

Busa’s system worked. In 1951, he published a concordance of the poems of St. Thomas as a proof of concept. In 1980, Roberto Busa published his *Index Thomisticus* as a 56-print volume series. In 1992, a CD-ROM version with hypertext was released. In 2006, a web version of the *Index Thomisticus* was launched.

Busa had created the first working mechanical system to study language. While this system was not a computer, it certainly was an important step to bringing the humanities over to the computational world. Prior to Busa, tech at the time was only focused on using their machines for mathematical computations. Busa showed that the data from a text could help new scholarly study. With Busa’s research, innovative thinking, and persistence, an entire new field was born.