

Analysis of “Against Cleaning”

Full Reference Citation:

Authors: Katie Rawson

Date Publish: N/A

Title: “Against Cleaning”

Source Type: Article/Textbook

This Article seemed unique at first glance. It is supposed to make a case against data cleaning and how it can inhibit the diversity and uniqueness of data itself. However, during my reading, I found myself consistently getting lost in the sheer amount of information. It genuinely started to make me have a headache, like an overload in my brain, but I will try to annotate as best I can.

As previously stated, Its purpose was likely to make the case against data cleaning and the use of raw and diversified data. They used the terms Scalability and Non-Scalability, and in simple terms, Scalability is essentially the mass production of a collection of objects or concepts, like a teddy bear that was made the same each time. Non-scalability is the concept of every said object being its thing, like two handmade teddy bears, they will always be different. I say all this to draw a connection to their stance on data cleaning, If you clean all the data from all its imperfections, It runs the risk of becoming a matched set along with similar data sets. However, leaving the data relatively raw, the data will all be its own thing.

Just by looking at it, I can tell that this was researched carefully and with an overflow of love from the authors, such as the concepts of scalability and Non-scalability, and how it explained against something that researchers have been doing to data for several years. The results showed that Data Cleaning can actively diminish diversity by actively hiding it in the scrubbing of raw data, something that the theory of non-scalability wants us to recognize. One

limitation is that this article, as educational as it is, is essentially one big conglomerate of information that was admittedly very difficult to follow.

This differs from all the other reading annotations in that it was just so arduous to get through, to the point I just skimmed through it and jotted down only the points I thought were important. Most of the other sources I have encountered have also been relatively long but at least I could get what I needed to know from the abstract at the top of the article to get the important information, I sincerely and desperately hope that this is the last annotation with an Article of this caliber.