

Name: Chun-Chi Huang
Course: Data Modeling
Professor: Michael McKee
Date: 01/15/2018

Reflection: Needle in a haystack

Nowadays, data is everywhere. Our struggle is not hard to find the information. The real issue is how to organize our information. Consequently, "Metadata" is created for this purpose. Assyria people used tags added on clay tablets to put them in different places. Now, this mechanism becomes more complicated than before. Currently, Google is the most popular search engine company, and it creates abundant metadata to organize the information people want to know. Based on the relevance, this metadata can help Google to achieve the excellent job. Google deals with unimaginable data on the public internet every moment, and this search engine has the capacity of completing around 35000 queries every second. Consequently, Google is one of the most profitable companies in the world because it knows how to find the information by its effective search engine. Metadata is also applied in pictures and videos. It not only stores the time and the place people take but also the model of camera. This unorganized information, such as pictures and videos, is tagged by internet users. Sometimes the tags are reasonable, but sometimes people distort the real meaning of media information. Physical items such as paper and films are only stored in one place. However, digital data can be spread out everywhere at the same time. In conclusion, digital metadata makes searching easier and harder in the meanwhile.

Finding a needle in a haystack is almost impossible. However, searching a book in a library is possible due to metadata. Instead of physically searching for each floor, each block or each shelf, librarians only need to look up the index and then show us the exact location of the book. Searching the information on the internet is the same concept with library books. My favorite searching tool is Google. If I use the correct keyword, I can find the right thing I need. Nevertheless, when I search a photo or video, the wrong items data are usually shown in front of me. Although we know that the search engine looks up the metadata first and then get the information for us, why do we still receive the wrong one? Because of the correctness of this metadata, like the author's description, some people use chaotic words to describe. It will lead to an ineffective search. I agree with the author's statement, "Digital metadata makes things more complicated and simpler at the same time." Correct and accurate metadata can result in a valid search; however, wrong and distorted metadata will lead to a terrible mess. It is indeed an issue. Recently, Artificial Intelligence and Machine Learning are significant studies. Although I don't entirely understand how they achieve, the example, AlphaGo AI won the world's best Go player, showed that AI and ML could have the capability of fixing this issue. We can feed abundant existed photos and videos into AI's machine with the correct metadata to train the algorithm between them. It also needs a mechanism to judge whether the search is valid or not. Once the successful rate is satisfied, it also means that the issue is solved.