

For centuries, people have struggled to understand the meaning that's hidden in large amounts of data. Over 2000 years ago, tax collectors for Emperor Qin Shihuang used the abacus—a device with beads on wires—to break down tax receipts and arrange them into categories. It's something else to classify what species of trees they are, how they cluster at different altitudes, and what could be built with the wood they provide. From this, they could determine how much the Emperor should spend on building extensions to the Great Wall of China. To sort out unstructured data, humans have created many different calculating machines. After all, it's one thing to estimate how many trees grow in a million square miles of forest. That information can be difficult to extract from a very large amount of data. Because it's hard to see without help, scientists call this dark data. An abacus