In this special lecture, first of all, the speaker Tova Milo introduced herself, she received her Ph.D. degree in Computer Science from the Hebrew University, Jerusalem, in 1992. After graduating she worked at the INRIA research institute in Paris and at University of Toronto and returned to Israel in 1995, joining the School of Computer Science at Tel Aviv university, where she is now a full Professor and holds the chair of Information Management. She served as the Head of the Computer Science Department from 2011-2014.

[Data mining](https://en.wikipedia.org/wiki/Data_mining) is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes, while [business intelligence](https://en.wikipedia.org/wiki/Business_intelligence) covers data analysis that relies heavily on aggregation, focusing on business information. In statistical applications, data analysis can be divided into [descriptive statistics](https://en.wikipedia.org/wiki/Descriptive_statistics), [exploratory data analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis)(EDA), and [confirmatory data analysis](https://en.wikipedia.org/wiki/Statistical_hypothesis_testing) (CDA). EDA focuses on discovering new features in the data and CDA on confirming or falsifying existing [hypotheses](https://en.wikipedia.org/wiki/Hypotheses). [Predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics) focuses on application of statistical models for predictive forecasting or classification, while [text analytics](https://en.wikipedia.org/wiki/Text_analytics) applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of [unstructured data](https://en.wikipedia.org/wiki/Unstructured_data). All are varieties of data analysis.

Modern data analysis combines general knowledge stored in databases with individual knowledge obtained from the crowd, capturing people habits and preferences. To account for such mixed knowledge, along with user interaction and optimization issues, data management platforms must employ a complex process of reasoning, automatic crowd-task generation and result analysis. In this talk, I will introduce the notion of crowd mining and describe a generic architecture for crowd mining applications. This architecture allows audiences to examine and compare the components of existing crowdsourcing systems and point out extensions required by crowd mining. It also highlights new research challenges and potential reuse of existing techniques/components. I will exemplify this for the OASSIS project, a system developed in Tel Aviv University, and for other prominent crowdsourcing frameworks.

This is my pleasure to attend this speech. The speaker named Tova Milo, who is very famous all over the world. Her research focuses on large-scale data management applications such as data integration, semi-structured information, Data-centered Business Processes and Crowd-sourcing, studying both theoretical and practical aspects.

In her speech she introduced what is data analysis in? Data analysis is a process of inspecting, [cleansing](https://en.wikipedia.org/wiki/Data_cleansing), [transforming](https://en.wikipedia.org/wiki/Data_transformation), and [modeling](https://en.wikipedia.org/wiki/Data_modeling) [data](https://en.wikipedia.org/wiki/Data) with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.

And in this speech, the main topic is “Crowd-Powered Data Management”. by searching papers from the internet, I found many researches based on data management. As we know, data is everywhere, the amount and diversity of data being generated and collected is exploding, for example the data in web pages, sensors data, the data in wearables and DNA sequences and so on.

However, if we only have data. But data could not tell stories, the raw data alone is not sufficient! Because for people, we can only make sense of our world by turning this data into knowledge and insight. And then buried in this flood of data are the keys to new economic opportunities, for example, to training these data in machine model to predict the stock market. And we can also get data from industries and companies to improving their productivity and efficiency. In this speech, the speaker will focus today on human knowledge, to think of humanity and its collective mind expanding. The research frontier can be divided into three steps, the first step is knowledge representation and the second step is knowledge discovery. The last step is knowledge collection, transformation, and integration, sharing.

Then the speaker also told us the background of crowd (data) sourcing. Crowdsourcing and human computation enable organizations to accomplish tasks that are currently not possible for fully automated techniques to complete, or require more flexibility and scalability than traditional employment relationships can facilitate. In the area of data processing, companies have benefited from crowd workers on platforms such as Amazon’s Mechanical Turk or Upwork to complete tasks as varied as content moderation, web content extraction, entity resolution, and video/audio/image processing. Several academic researchers from diverse areas ranging from the social sciences to computer science have embraced crowdsourcing as a research area, resulting in algorithms and systems that improve crowd work quality, latency, or cost. Given the relative nascence of the field, the academic and the practitioner communities have largely operated independently of each other for the past decade, rarely exchanging techniques and experiences. There has the engagement of crowds of web users for data procurement.

And through this speech, I also searched from internet. What is crowd mining? Crowd mining is a new business concept where solutions can come from crowds, and marketers are turning to them as a database. They are a handy source of information, especially with online advances and wireless technology. The changing economy brought in this concept to tap expensive or even unavailable information and it goes beyond surveys and crowd sourcing. Scientists are also scouring the internet for information and much of the pioneering work in this type of internet surveillance is in the public health field. At same time, we can know, human knowledge forms an open world. And we try to find out what is interesting and important in some domain area. In health care area, the folk medicine and people’s habitats and doctor’s intuition they all influence the health of people. And also in finance area, the people’s habits and perferences and consultant’s intuition they all influence the prediction the finance market, thus data analysis patterns are very important. When we back to classic databases we can find that the sifnificant data patterns are identified by using data mining techniques. We can consider a useful type of pattern, there are many association rules. For example, the relationship between diapers and beer. In our first sight, or in our first image, there has none relationship between diapers and beers. However, through market analysis, we find that, sometimes, fathers often forget buying diapers when they went to supermarket, they often feel more interested in wine. Thus, when we put diapers and beer in the same area, it is a good way for father to remember to buy diapers when they select their beers. In this way, the sales volume between beers and diapers will increase quickly.

In conclusion, crowd data management is very important. In this method, data row is not simply and sclient, we can through it to let data start to speak. Through several visualization methods, and combine these pictures with analysis, we can have a powerful data management mode.