

The New York Times

Bringing Data Mining Into the Mainstream

By Steve Lohr

July 26, 2010 4:26 pm

5:50 p.m. | **Updated** to add link to the conference.

Plumbing the world's ever-growing pools of digitized information — on the Web, in corporate databases, generated by scientific research — for wisdom and profit is a growth industry today. The geeky field even has a shorthand name, “big data.”

So it is scarcely surprising that a four-day conference in Washington, organized by the Association for Computing Machinery, and focused on knowledge discovery and data mining, is attracting corporate researchers and university scientists in record numbers.

The papers submitted and workshops convened at the conference, which began on Sunday, point to the breadth of the field. The targets of big-data sleuthing range from behavioral targeting to cancer research.

Leading technology companies — I.B.M., Microsoft, Oracle, SAP — have all made large investments in the last few years in business intelligence and analytics software, to offer advanced data mining products to corporate customers. And they join the longtime specialist, SAS Institute, a private company in Cary, N.C.

But for all the excitement and investment, profitably probing large data sets remains an elite undertaking, costly for companies and difficult for users.

To bring modern data-mining into the business mainstream, two things are needed, Usama Fayyad, executive director of the conference, said this week.

The first, Mr. Fayyad said, is an institutional mind-set that recognizes the potential importance and payoff of data. “We need data to have a voice at the executive level,” he said in an interview on Sunday.

In the past, Mr. Fayyad noted, data was regarded as a byproduct of doing business, often a backward-looking record of little value. But today’s vast oceans of data, combined with tools for instant analysis and prediction, Mr. Fayyad said, are a “new strategic asset” that can be used to build “new revenue streams and new businesses.”

The largest new business, and big-data pure play, is Google, of course. Its search advertising business is based on collecting and probing big data — Web pages, blogs, social network chatter and users’ searches.

The other thing needed to democratize modern data mining, Mr. Fayyad said, is a translation layer of technology. It would take the underlying software for handling large data sets, often scattered across thousands of computers, like Hadoop and MapReduce, and link it to software that ordinary people can use.

For the user interface, Mr. Fayyad said, Microsoft’s Excel spreadsheet is “a very good metaphor.”

The sophisticated data-handling layer, he said, should be “built in ways that Excel can consume the data and people can browse it.”

Mr. Fayyad’s career itself has been a journey in bringing big data into the mainstream. An expert in data mining and artificial intelligence, he spent seven years as a scientist at NASA, working with astronomical data sets generated by observatories and spacecraft. He then joined Microsoft Research, and later became Yahoo’s chief data officer and executive vice president for research.

In the fall of 2008, he founded Open Insights, a consulting firm specializing in exploiting data in business. “Increasingly, companies all have this need for what is essentially a chief data officer — the ability to coordinate the computer science with new business models and organizational changes,” Mr. Fayyad said.

© 2014 The New York Times Company