
From: Udi Manber ^{Redacted}@google.com> on behalf of Udi Manber
Sent: Wednesday, August 26, 2009 12:50 AM
To: Hal Varian
Subject: Re: scale

I was concerned with the original interview for cnet http://news.cnet.com/8301-30684_3-10309375-265.html

where among other things you said

"So in all of this stuff, the scale arguments are pretty bogus in our view because it's not the quantity or quality of the ingredients that make a difference, it's the recipes. We think we're where we are today because we've got better recipes and we have better recipes because we spent 10 years working on search improving the performance of the algorithm."

You specifically mentioned search improvements, and it's absolutely not true that scale is not important. We make very good use of everything we get. UI experiments are done on a small percentage but ranking is using a lot more.

-- Udi

On Tue, Aug 25, 2009 at 5:43 PM, Hal Varian ^{Redacted}@google.com> wrote:

Hey, Udi, I hear you have some concerns about a statement attributed to me by *Time* magazine: "Hal Varian, the company's chief economist, has pointed out that most search engines look at only a small sample of their data in order to improve their results"

This is a garbled version of what I actually said. Here is what I actually said: "Most experiments with Google algorithms are initially run on small samples: one percent or less. As you tune your algorithms, you run on larger samples and, when all is working well, you deploy the improvement to the world at large."

I can understand why you might think that "most search engines look at only a small sample of their data in order to improve their results" is inaccurate. If he had substituted "test" or "experiment with" for "improve" it would have been a more accurate report about what I actually said.

This is how we do things in ads (using RASTA) and I have always understood that you do similar experimentation on small samples in search. If I'm wrong about this, and you deploy at 100% levels from the beginning, please let me know.

There's also another issue I wanted to mention about the term "economies of scale" as used in these discussions. Economists say that a technology exhibits *increasing* returns to scale if doubling its size **more** than doubles its quality. For example, Google is said to be twice as large as Yahoo in terms of queries served. Does this mean that Google has more than twice the quality of Yahoo? Or, to phrase the question differently: back in Feb 2007, Google queries were about half the volume they are now. Was Google's quality less than half as good then as it is now?

I would claim that Google's quality exhibits diminishing returns to scale --- as you double output, quality increases but by less than a factor of 2. Admittedly, I am being somewhat loose about how we're

measuring quality but that's the basic idea. Let me know if you think this makes sense.