Future Tense

on ABC Radio National

The 'Deep' Web	11 March 2010

What lies beneath
full image

Think of an iceberg, the bit you see is only a small part of the total amount - it's the same with the Internet. On this program we explore the so-called 'Deep' Web, the part of the internet that evades most of us and our search engines.

Hide Transcript

Transcript

This transcript was typed from a recording of the program. The ABC cannot guarantee its complete accuracy because of the possibility of mishearing and occasional difficulty in identifying speakers.

Antony Funnell: As the crew of the RMS Titanic famously found out, there's often more than meets the eye ...

[Sound effects: *A Night to Remember*]

Antony Funnell: Oh dear, who'd have thought all that ice would be lurking below the surface.

Hello, Antony Funnell here and welcome to *Future Tense*.

Today's program doesn't really have anything to do with the Titanic, or icebergs, but the mental picture of an iceberg isn't a bad one to have in our heads as we talk about the so called 'deep' web. I'll explain in just a minute.

Also in today's program, we'll hear about the age of metrics, and we'll meet a man who wants us to rethink the way we think about facts.

Sam Arbesman: When people think of knowledge, they generally think of two sorts of facts, facts that don't change at all like the height of Mount Everest or the number of continents, or facts that change constantly, like the temperature or the stock market close. But in between there's a third kind of fact, facts that change slowly on the scale of a human lifetime, and these are mesofacts, and I've created a website mesofacts.org to understand these facts and I understand these facts and I'm Sam Arbesman, a post-doctoral Fellow at Harvard Medical School and the Department of Healthcare Policy.

Antony Funnell: Now back to that iceberg analogy. And it goes like this. Just as the visible part of the iceberg is only the tip, so too the part of the internet that you and I deal in, and which we traverse through search-engines, is only a small fraction of what's out there online.

In effect, there are two webs, the visible one and the invisible one, the Deep Web. Now what we want to do today is to just briefly introduce you to the concept of the deep web.

About a decade ago, Mike Bergman, a web scientist, entrepreneur and now the CEO of a company called Structured Dynamics, became the first person to try to estimate the extent of the invisible internet.

Mike Bergman: Well in the original papers that we published, we thought it was going to be perhaps two or three times larger than the standard web that people access via search engines. Depending upon the metric views, we were shocked. At that time the numbers came out to be 100 to 500 times larger either by virtue of number of documents or the size of those documents, depending upon the metric 100 to 500. Frankly, so that's now ten years ago. My personal belief today is that much of that that had been hidden ten years ago actually is being surfaced up and exposed. Search engines like Google are being very aggressive about ways that they oftentimes don't tell us about, about how they bring that material up.

Antony Funnell: So the deep web, according to Mike Bergman, was up to 500 times bigger than the conventional web at the turn of this century. Now as

we've just heard, he believes it's probably shrunk since that time as search engines have become more efficient at trawling for information. But has it?

Chris Sherman is the author of a book called *The Invisible Web* and he has a contrary view.

Chris Sherman: The big challenge that search engines have, they are very technologically advanced and they really can find a lot of information, but they have a trade-off, because mostly what they're trying to do is find answers to popular inquiries, to things that people are looking for a simple answer for. They're not doing deep research, they're not doing sort of investigations that might take time. They want a quick answer. And so the search engines really have to make a trade-off here, between really kind of flexing their technological muscle, if you will, and coming back with good enough answers for most of the people, most of the time.

Search engines over time have become much better at accessing that kind of information. They've gotten smarter, they've gotten more powerful and so on, but we still have that problem of the web expanding at just incredibly exponential rates, and it's a race that they simply can't keep up with. I started with people in Google and they've acknowledged the problem, and they've said 'Yes, we're working on it'. I asked them when they think they can really catch up with it, and the answer that I had actually from their Chief Technology Officer was 'Approximately 50 years'. So it's a big problem and it's one that's not going to go away soon.

Antony Funnell: So who uses the deep web? Well, it is sometimes called the 'dark web', so that should give you an idea.

But according to Chris Sherman, it's important to remember that not everyone who makes their activities and their files invisible on the internet is up to no good, a suggestion that's often inferred by governments and policing authorities.

He points out that two of the biggest publishers of deep web material are universities and governments themselves.

But how invisible is invisible? Well, according to Chris Sherman, if you know what you're doing, it is possible to dig down and pull out at least some of

what's hidden.

Chris Sherman: One of my favourite websites is actually Geosciences Australia. It's just a wonderful site that has all kinds of information about natural resources in Australia. You can search for the various things that it has information about. Mining, it actually has a database of all the nuclear explosions that have happened in the world since 1945. You search for those things, and what will happen is you'll come to the search forum for the databases, because those are very straightforward, no hidden aspects to them at all. Once you get to them, that's when you type in search terms that you're interested in, and then retrieve the information.

A telltale sign when you're using an invisible web database, is when you get a return resolve, and you've got either a list of search results that look like a search engine, that's a pretty good indicator that you've hit a database result. Or if you actually look in the address of the page that's returned, at the very top of your browser in that address window, if you see the question mark symbol, that almost always is an indicator that a database query has been run and it's giving you back dynamically assembled results based on your search terms alone. That's what makes the deep web so appealing, not that it's hidden or something mysterious, but the quality of information quite often tends to be quite a bit better than what you find on just the normal or visible web.

Antony Funnell: But of course if you go to the trouble of trying to make the material you store on the web inaccessible to the uninvited, you don't really want someone like Chris Sherman probing about.

And that's where a service like TOR comes in.

Andrew Lewman: The TOR project is an official non-profit, dedicated to the research and development of tools and technologies for online privacy and anonymity. So what that basically means is that we are funded by individuals, by governments, by other non-profits, and by for-profits, to do research into how to stay private and anonymous, or and/or anonymous online.

Antony Funnell: That's Andrew Lewman, the Executive Director of TOR, and he says it's time to get past the idea that being secretive on the web is

necessarily a negative or bad thing.

In fact he and his organisation question the notion that the more open and transparent the internet is, the better we'll all be.

Andrew Lewman: Over the past nine years we've grown into building a far more anonymous system so that people who really need anonymity, whether you're a whistleblower, whether you're an abuse victim or just a general person, have the utmost control of what they give out. A growing number of people are concerned that everything they do online is tracked and recorded and there's very little control over who gets to record what and how often they're recorded. A fine analogy is if you walk into a store to buy something, if someone came up to you and started asking you where do you live, how much money do you make, who are your friends, where else have you visited, how much other things have you bought, people would understandably be a little freaked out. Yet, this happens on the internet every time you go on.

Antony Funnell: Now some people, particularly governments would say that this is a dangerous thing, that if people are allowed to hide their identity and the content that they have online, that a large number of them will be doing for nefarious reasons, you know, be they paedophiles or violent fanatics of some sort, what would you say to that?

Andrew Lewman: I would say that many governments also use and fund TOR. Their law enforcement people are very concerned about what information they give out when they go to do online stings, and in many cases the criminals are more sophisticated than the law enforcement is. So sure, do jerks use TOR, or all the definitions of jerks you just gave? Yes, of course they do. But they also use the general internet, they also use emails, they also use digital cameras and cars and highways and everything else in the world.

We build the strongest anonymity system we can build for all the good positive uses of it. We do work with law enforcement to help them understand what TOR is, and is not, what our software should look like on a computer and what you should find and not find, depending on how the particular investigations. They go to confiscate a PC we'll generally tell them, look this is what our software does, and this is what it does not do, let them

go from there.

Antony Funnell: Andrew Lewman, from the online privacy service called TOR. And just for the record, like Chris Sherman, he also believes the deep web is expanding rather than contracting.

Add your comment

- Delicious
- Facebook
- Digg

what are these?

x close
Fields marked * are required

email *

(Multiple addresses? Put commas between them.)

your email *

your message

Send

We don't collect any of the information you're putting in. It's between you and the person, or people, you're sending it to.

By clicking 'Send' you agree ABC Online is not responsible for the content contained in your email message.

Guests

Mike Bergman
Entrepreneur, web scientist and CEO of Structured Dynamics.

Chris Sherman
Executive Editor of Search Engine Land and co-author of 'The Invisible Web:

Uncovering Information Sources Search Engines Can't See'.

Andrew Lewman
Executive Director of TOR.

Further Information

The TOR Project

Chris Sherman's profile

Mike Bergman's 2001 paper on the 'deep web'.

Guardian article on the 'deep web'.

Movie clip: 'A Night To Remember'. 1958. Dir: Roy Ward Baker

Presenter

Antony Funnell

Producer

Andrew Davies

Radio National often provides links to external websites to complement program information. While producers have taken care with all selections, we can neither endorse nor take final responsibility for the content of those sites.

Thursday 8.30am
repeated **Friday 12.30am**
Presented by
Antony Funnell

In This Program

(full program)

(full program)

- **8.30**: The 'Deep' Web
- **8.45**: Age of Metrics
- **8.55**: Mesofacts

Future Tense | Radio National | Programs A-Z   © 2010 ABC | Privacy Policy | Conditions of Use