a *process* not an object.

So, the big challenge for you and your company is actually **not innovation**.

It is not the single, first learning event.

The challenge is *spreading innovation*

●●● trying to recapture and re create many times over that **one moment**…that process…that learning event where a person, or people and information interact while doing some kind of activity. Drawing information and energy from all around - **POW** - an AHHH happens!!! Learning happens.

So, the process of implementing innovation is **trying to make that learning event more likely to happen.**

That means getting the **right information** to the **right person/people** at the **right time** in the **right place**, when the **energy is highest**.

Why Searching Sucks
Looking Behind Problems

Pick your favorite search engine. Enter a word or phrase. How many hits do you get? Is it over a million?

Is that useful?

Remember Weick and Sensemaking? Remember one of the principles is that information that is convenient will be used even if it is not as good as information that is better? (A bird in the hand…)

So, when searching, when an engine returns several million documents, most of us look at the first few and grab the closest thing that will work.

We are awash with information and it makes us feel that any piece of knowledge we need is out there, *if we were only smart enough to find it.*

The implication is that we are wrong, not the search engine.

WRONG !!!!

Searching is fundamentally flawed because of

decisions made over forty years ago that are so intertwined with what we do when we search that neither we nor the people developing search engines think about them.

Looking at this is useful in itself, but is also useful to demonstrate the importance of looking at how **underlying, often hidden, decisions shape our actions.**

Why is searching so bad?

Is it because there are boat loads of sites out there with huge piles of information? Well, that helps. But a large part of the reason can be found in a few, flawed underlying assumptions that drive the way search engines work. However, the assumptions are so far immersed in the search tool, almost no one is aware of the assumptions and how they mess with us every time we search.

Here's a quickie lesson in how search engines are evaluated. **You have to understand this to understand the problem. You have to understand the problem to see a solution** (see the chapters "Don't Search - Filter and Evaluation for the Information Age in Part 2).

The "classic method" of evaluation needs three things:

1) a bunch of **questions** well, actually queries.. because when you are trying to find information you don't type in the whole question. (If I want to know the closest star to earth, I type star closest earth - which returned 673,000 documents last time I searched).

2) a bunch of **documents** that the engine is going to search for answers to those questions (during the testing, they are just going to search those documents.. they don't search the entire web).

3) a bunch of **relevancy judgments,** which are **-** this document answers that question**...** so you might say that in *this* bunch of documents, there are 50 documents that are relevant to the first question, 12 for the second.. like that.

If you are asking, how many angels can dance on the head of a pin, a document that talks about shoeing a horse is NOT relevant, neither is a document about pin the tail on the donkey. Got it?

Three things - questions, documents, judgments.

The way a search engine is tested (and yes.. there are bunches and bunches of search engines out there) is this- the testing people write a little program that uses the search engine to search the documents for answers to the questions (or documents retrieved by queries…if you want to sound confusing).

Then, they *evaluate* the documents retrieved.

The first question is, "**How many of the relevant documents were returned?**" If there were 20 documents that were relevant and the search engine found 10 of them, then it found 50% of the relevant documents. **Make sense?**

That measurement is called **Recall**.

You can use recall to compare different search engines. So, if one search engine manages to find 50% of the documents that are relevant to a particular question and another finds 80%, then the 80% one is better.

Here's a way to cheat.

Instead of actually doing some interesting kind of search, which involves all kinds of matching and programming, why not just have the search engine **retrieve every document** in the test bunch of documents?

That way, the Recall will always be 100%. If you get **all**

the documents, you are certain to get all the **relevant** documents.

To keep anyone from being tempted to cheat, there is a **second measure**. This is a measure of how **messy** the results are.

Say I did a search and got 200 documents back. Of that 200 documents, 50 were relevant. Then, 25% of the documents retrieved by the search engine

were relevant.

This measure is called Precision.

In the **traditional method,** in order to **compare two different search engines, you use a <u>pair</u> of measures** - Recall **and** Precision.

Usually, the search engines that retrieve more relevant documents (in other words, they have higher recall), ALSO have more garbage in the results (the lower the precision).

The higher the recall, the lower the precision and the lower the recall, the higher the precision...usually.

So far so good? It sounds great. What's the problem?

There are two big nasties hiding in the bushes.

The first one is the reason you get several million hits every time you search. This way of measuring **assumes** that a "good" search engine will retrieve as many good documents as possible. **More is better.**

Look at the measures - if you had a perfect search engine, then it would return every relevant document - right? Then you would have 100% recall. It would not return anything else, so you would also have 100% precision. Sounds good.. right?

Well, that was nice when there weren't very many things out there on the web. **However, now that there are billions and billions of documents out there, returning as many "good" documents as possible is no longer a good idea.**

Wouldn't it be better to get <u>a few</u> really good ones?

Because search engines are measured by the number of hits - a measure of **quantity,** of course we are going to have search engines that bring back a boatload of hits**,** even if it means that we are going to ignore most of em.

Our measure creates our problem (next

chapter - You Get what you Pay for).

The **second nasty** in the bushes is the idea of "relevant" documents.

Remember, we measure search engines by how many relevant documents they return. **The problem is - who decides what is relevant?** If Veronica Vegetarian and Meaty Mike are both searching

for good restaurants in their area.. well, who is deciding what a relevant document is?

 In fact, relevancy depends a great deal not only on **who,** but on **when.**

If I need to get a shipment of ice cubes through customs, where they are busy melting and I need some help filling in the forms, I don't want a document that explores in depth the best methods for setting up an exporting business. At other times, that may be EXACTLY what I want. In both cases I might be entering the same words in my search.

Relevance is a very squishy term, but in order to test search engines the way it is currently done, someone or something has to make the decisions as to **which documents answer each query**. So, an engine that does a great job for one person, may do a terrible job for another even if it is the same engine.. evaluated as having the same "quality"..

Why do you get piles of interesting, but often useless tripe to sort through when you search? It is because of the **way quality is judged,** and that is **so ingrained in the way search tools are designed <u>and evaluated</u> that no one can see beyond it.**

This has bearing on you, not only because you can…we all can…begin

to demand search engines that respond to our wants and needs, but it also points to **two problems which are common within organizations:**

1) **you get what you pay for,** so pay
   for what you want ... and
2) **one size does NOT fit all**

More on these in chapters to come.