

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

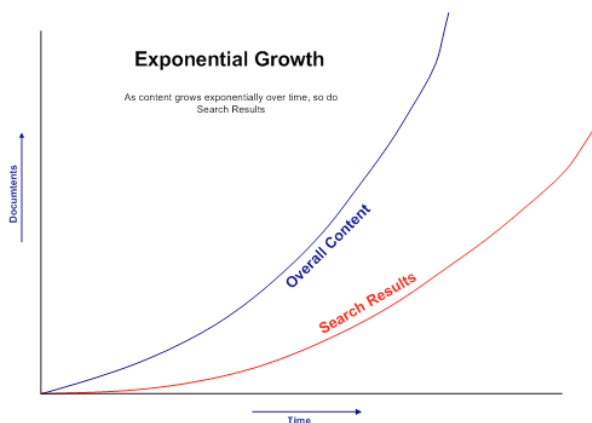
By Mark Bennett

Chief Technology Officer, New Idea Engineering

Exponential Growth Swamps Results Lists

We've all heard the hype about the exponential growth of the Internet, but there are a few related developments, not so widely covered, that Enterprise Search people need to keep in mind.

First, as the overall number of documents grows, so does the average number of documents returned by each search. Think about it. If a broad search matches 10% of the documents, then it would match 100 documents out of 1,000 total. If the same search is run later when there would now be 2,000 total documents, it would return on the order of 200 documents in the results. This may not be the case for every search; the overall makeup of the documents may change over time, or specific terms may come and go, but on average *results list sizes will tend to grow exponentially*, generally mirroring overall data growth.



Second, enterprise data is also growing exponentially. Of course the growth curve may not be as steep, and will vary from company to company or agency to agency, but exponential growth at any reasonable rate adds up surprisingly fast.

You'll recall that in the 1990s, the public Internet crossed certain boundaries. The headlines

reported when the web surpassed 10 million pages, 100 million pages, 1 billion pages, and more. We are now at the point where many private networks are as large as the entire public Internet was in the late 1990s. In fact, there are now private *datasets* that have crossed, or will soon cross, the 100 million and 1 billion document marks. This means that the Internet search engine problems of the 1990s are now hitting corporations and other institutions; this is a "big" problem.

The problem with "Single Shot Relevancy" -- The Problem even Google Can't Fix

"Single Shot Relevancy" is the idea that when a human types in a search, the search engine returns the "best" most relevant documents on the first page of results. One query goes in, and one amazing results list comes back. Early vendors' relevancy was based solely on the content of the documents themselves; later vendors added external techniques such as Google's link ranking. Now, when vendors talk about relevancy, this is usually what they are referring to, but their engines will all eventually fail because of exponential data growth. It's not that there aren't any good relevancy algorithms around – there certainly are. You can even employ "query cooking" to improve them further.

But imagine two fictitious search engines, "Average Engine" and "Good Engine." We give each engine 1,000 documents to index, then run a reasonable search. The search matches 5% of the content, returning 50 documents. Both engines are probably going to put a few decent documents on the first page. Now imagine increasing the docset from 1,000 to 10,000 – this would raise the result set for the previous search to 500 documents – and at that point "Average Engine" has a good chance of not putting reasonable documents on the first page. Meanwhile, "Good Engine," through better ranking algorithms, manages to still get decent results on the first page.

But let's keep going with this. Now think of a similar dataset that is two orders of magnitude

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

(100 times) larger, and we're now indexing 1 million documents. The results set has mushroomed to 50,000 hits. At this point "Good Engine" is going to have some problems populating that first page with relevant results – or at least what the user would think of as relevant. We are unhappy with the results, so now for this 1 million document dataset we consider a couple of new vendors, "Cool Engine" and "Great Engine." Let's assume that both of these engines do a better job than "Good Engine," and even with 50,000 hits, they can both usually put the correct documents at the top.

One more time – this is important – consider another dataset, 100 times larger again; that puts the raw data at 100 million documents, and a results list on the order of 5 million documents. At this lofty number "Great Engine" and "Cool Engine" are going to be straining, and even slight problems with ranking will be greatly amplified.

Even if one of the vendors could miraculously improve its engine's relevancy, those gains would likely be wiped out by an even larger result set.

Don't Be Confused: We're not talking about simple "performance"

A note to the reader... There are two paths this type of discussion usually follows which are distractions, neither of which is pertinent to the discussion of "Search 2.0."

Distraction 1: "OK, so you're talking about search engine performance – how long it takes to index and search documents?"

No. While that is certainly an important factor, the big players in the industry can scale up, using multiple servers to handle vast quantities of data. If you've got 100 million documents and a big enough budget, you will be able to get them indexed and searched.

Distraction 2: "Ah, so you're talking about Relevancy – getting the right documents to show up at the top of the results list?"

No, not exactly. Better relevancy, or "document ranking algorithms" will certainly buy you some

time, but each time the number of matching documents increases, the effectiveness of even the best algorithms will eventually fail. This type of relevancy which we call "Single Shot Relevancy," will eventually fail when the dataset gets large enough – it's only a matter of time.

The mistaken "HAL9000" Assumptions about Users and Questions

The assumption that a human, given enough time, could identify the most relevant document out of 5 million hits is suspect. We do, however, expect a computer to be able to do it.

Many early computer scientists were inspired by the computer "HAL9000" in the science fiction epic "2001: A Space Odyssey" from the late 1960s. In the movie, this highly-evolved intelligent machine, created in the 1990s, could converse, reason, and question like a human being. Many people still hope and assume that computer technology will ultimately get there; a system that could truly understand what a human is asking, then research the question and come back with a concise answer would be wonderful to have and would satisfy the "HAL9000" benchmark. Sadly though, 10 years after this fictional technology was supposed to be available, it doesn't exist. Still, many users subconsciously assume that their search engine knows the difference between a baseball diamond and the diamond on a wedding ring, or the difference between President Bush vs. a "bush" that you plant in your front yard. We are not there yet, our search engines are not "HAL9000" capable.

In the movie, the human operators asked well thought-out questions, and those questions do have "correct" answers. A number of articles cite the average query as only 1 to 2 words in length -- humans don't usually ask well thought-out and complete questions. Even with a longer, more specific query, the "correctness" of answers is still debatable. Imagine the discussion that would ensue grading the results of the lengthier query "effectiveness of tax cuts to stimulate the economy." As you can imagine, opinions would vary widely over whether the answers from the

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

Republican National Committee or MoveOn.org would be more relevant.

The "Patches" to Postpone the Inevitable

As mentioned earlier, search engines started including external data in their relevancy engine. Google used "link ranking," a methodology that takes into account the number of other web sites that link to a particular page, with the assumption that a page with more links to it is more authoritative and therefore more relevant.

Other engines started taking click-through rates into account. For example, if everyone who enters a particular search clicks on the third document on the results list, then presumably that document is more important, and should be bumped up to the #1 slot.

Still other engines worked with "context," trying to look at the user's previous surfing and search activity, or job function within a corporation.

We laud all these efforts; they have certainly moved us further along in our Internet searches, but they have not been as effective inside private networks. Link ranking, Google's "secret sauce," doesn't work as well inside corporations because page links are generally based on organizational charts or subject taxonomies, rather than user-by-user. And of course the new big fad, paying for placement, isn't applicable. Nevertheless, this is fundamentally all still "Single Shot" logic.

In recent years, even Google's public portal has struggled to give decent results for many searches. If what you're looking for doesn't line up with what most folks are interested in, or if you can't get your search terms just right, Google can return page after page of garbage. Exponential growth has overtaken even mighty Google's ranking.

The Internet Search Meltdown of the Late 1990s is Now Visiting Enterprise Search

We have heard client after client complain that their enterprise search relevance is unacceptable. The most common quotes they relay from their

users are, "We can't find anything," and "Why can't our search be more like Google?" or simply "Our search sucks."

With many private networks now as large as the public Internet was in the 1990s, it was inevitable that the public search crisis of the late 90s would hit the enterprise world.

And what technology finally rescued the public Internet? Many say Google did. They used the links between web sites to judge which sites were "better"; in effect they overlaid external data to improve results – an early type of social network voting.

"Search 2.0" is really "Search 1.5"

To our minds, "Version 2.0" sounds revolutionary, whereas much of what we're seeing has been tried before, sometimes with limited success. That's OK – software often goes through several cycles before it's completely usable, so if some of this looks familiar, you're not imagining things.

Ironically, even the Enterprise version of Google, their "Google Box," can't use its link-rank trick to rescue Enterprise search. Without all that extra data, it performs similarly to other engines. Google's brand name is certainly leveraged in the enterprise space, but its relevancy mechanism isn't. So the search meltdown of 1990s is coming ashore in the enterprise and the Google link-rank lifeboat is nowhere in site. It's OK! Even patches to single-shot relevancy were only temporary; now it's time to rethink the entire search process.

Search 2.0: Rethinking the Results List - The Power of Well-Implemented Drill Down Search

It is unrealistic for search engines to guarantee the "best" answer at the top of every results list. The best they can do is empower the human user to see what information is available and give him convenient power tools to navigate through the clutter.

Author John Battelle said it best in his book [*The Search: How Google and Its Rivals Rewrote the*](#)

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

Rules of Business and Transformed Our Culture

when he compared the Google results list to the infamous DIR command from the DOS operating system. Search results are evolving from simple lists of documents to interactive tools that show the type of data that is available, and allow users to efficiently navigate to their answer. In some cases the results list may even answer the question directly.

helpful; in the past some bad vendor implementations have frustrated users. Good quality drill-downs often require at least some human supervision (though a majority of the actual work can be automated).

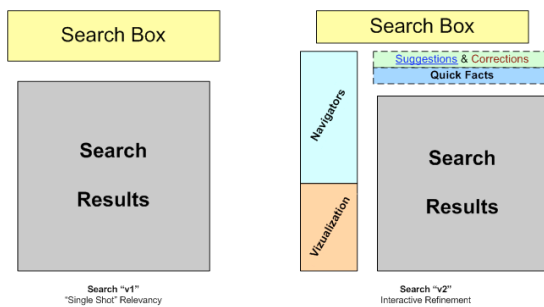
The following are some illustrated examples:

- **Entity Extraction based on proper name** (*upper left*)

Clicking on any of these terms will drill down into articles that include that person's name, along with the query term "bush."

Enterprise Search Results List

Block diagram: "Search v1" vs. "Search v2"



Textual Navigation

This block shows several examples of textual navigation elements.

- People:** A list of names: George W. Bush, Laura Bush, Dick Cheney, Condoleezza Rice, Colin L. Powell, Scott McClellan.
- May we suggest:** A link to 'Welcome to the White House' with a small icon.
- Did you mean:** A suggestion for 'Bush'.
- Places:** A list of locations: Washington, DC; Crawford, TX; Baghdad, Iraq; Brussels, Belgium.
- Subject Taxonomy:** A hierarchical list of categories including US Government, Executive Branch, Commander and Chief, President George W. Bush, President George H. W. Bush, Biology and Life Sciences, Botany, Shrubbery and Bushes, Food and Drink, Alcoholic Beverages, Beer & Ale, Anheuser-Busch, Bush Beer, Dubuission, Entertainment, Music, Popular American Music, Kate Bush, and Adult Content (results filter settings).
- Related:** A list of related terms: White House, State of the Union, Air Force One, UN resolution, Congress, 2007 budget, exit strategy.

Types of Results List Navigation

Here is a quick tour of the types of "Search 2.0" features that enterprise search vendors are starting to offer.

Disclaimers: In the examples that follow we use a US president and recent events as an example, and show data that you might get back on a public search. Political figures and world events are standard examples when discussing search engines. Your users will search for data specific to your company, so try to think of financial or technical terms that might be more applicable. Also keep in mind that all of these examples are mockups; actual technology isn't quite up to performing some of these miraculous tasks yet. And finally, it's unlikely that any single site would implement all of these – it would be too cluttered.

- **Entity Extraction based on geographic location** (*center left*)

Performs similarly to the People navigator.

- **Noun Phrase entity extraction** (*lower left*)

Noun phrases are becoming very popular in the industry. They can use language rules or dictionaries to find noun phrases. Noun phrases are typically defined as a noun with all its surrounding qualifiers, or a sequence of words previously defined to be a single subject of interest. Vendors have been stuck on one and two word phrases in the past, though lately some vendors are experimenting with longer phrases.

- **Directed Results** (*upper right*)

This feature can also be called "Best Bets," or

Textual Drill-Down Navigators

These are examples of the most common types of iterative search results elements offered today. When implemented accurately, they can be quite

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

"Webmaster Suggests," depending on the vendor. In this case a specific suggestion for a likely web page is being offered; these suggestions are likely to have been created or approved by a human editor. We believe this is one of the easier "bang for the buck" ways to improve enterprise search; efforts can be spent on the top 100 or top 1,000 searches and related areas. We have also referred to this type of activity as Behavior Based Taxonomy.

- **Spelling Suggestions** (*center right*)

This mechanism can also be used to suggest commonly related searches such as accessories for a product or related instructional links. Implementations vary from completely automated to completely manual.

- **Subject-Oriented Taxonomy** (*lower right*)

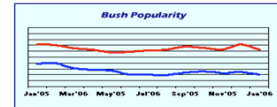
This may look the most familiar and remind you a bit of Yahoo. Clicking on any of these links would have re-issued the user's search to content in that specific area. If you think about it, "b u s h" could relate to politics, beer, gardening, or other diverse subject areas. The biggest challenge with this type of system is getting your content into a well-organized taxonomy. Beware of vendors who promise complete automation! At a minimum, you should be able to edit or override the system's groupings and rename the node labels.

Non-Textual Suggestions:

In the first example, three links are offered to multimedia items related to "b u s h." The first is a link to a video of a President Bush State of the Union address; the second links music by the pop artist Kate Bush; and the third links to a gardening show on planting a bush in the back yard. The second graphic presents (fictitious) polling data. In an enterprise, a graph might show sales or customer satisfaction related to a particular query term.

Including non-textual Content

Video, Music and Pictures:



Fact Aggregation:

In what may be very close on the horizon, these examples show the system attempting to bring together snippets of data related to a query. Though this may sound like it borders on Artificial Intelligence (AI), vendors are taking a more mundane approach. One step beyond entity and noun phrase extraction is simple fact extraction. If enough articles mention a fact, and some reasonable percentage can be normalized down to a common form and tabulated, the system builds evidence to support a fact. In this case, we're not asking the system to understand every sentence in every document. Instead, the system is simply breaking down all documents into smaller units, normalizing them structurally, and then tabulating statistics. In the first example, such a system would have already broken down George H. W. Bush, George W. Bush, and United States by dictionary matching.

Aggregating Data

Quick Facts:

- George Herbert Walker Bush was the **41st President** of the United States. [\(info\)](#)
- George W. Bush was the **43rd President** of the United States. [\(info\)](#)
- **US landscaping revenues** are approximately \$40 Billion annually. [\(info\)](#)



"41st president" and "43rd president" would be recognized as noun phrases. Landscaping revenue would also be identified as a noun phrase. It's not too much of a leap to have pieced together the sentence structure necessary to assemble the facts. Keep in mind that the system may have seen thousands of documents with almost identical information; some of the documents may

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

have even used this exact wording with other variants used as statistically supported evidence.

In the second example, dates are a well-understood entity that many systems can now extract, even when they appear in the text of a document. Combing this with the fact process previously mentioned, a timeline might be culled after tabulating thousands of documents. In some cases, the results list will actually contain the answer the user is looking for. If the "bush" query had been to find out quick presidential facts, a user might not have even needed to click any further; in these cases the results list *is* the answer. While this is nice when it happens, users should not expect this on a regular basis or for more complicated searches.

This is different than the "HAL9000" model. No specific question was asked – the query was just "bush" – and the system didn't need to understand these facts from only 1 or 2 documents. Furthermore, the "factoids" that are auto-generated for popular searches could be reviewed by human editors before being published.

An approach that makes these tricks possible, while not relying on fancy AI, is that vendors are approaching simple sentence parsing from two angles. On one hand, instead of single words, they are searching for multi-word phrases using various techniques; these multi word entities can be used as the seeds in straightforward statistical analysis. At the same time, they are breaking down documents into smaller chunks, analyzing statistical correlations at the paragraph and sentence level. None of these simple methods are attempting to identify broader concepts. For example, when these systems encounter the pronoun "he," they are generally not backtracking to previous sentences trying to find the antecedent; a sentence with "he" or "she" as a subject will probably not contribute much to any simple statistic tabulation of facts.

Whether this round of implementations will be able to achieve this any better than previous attempts will be decided by the market. We suspect in some cases that some of these fact extraction techniques will yield satisfactory

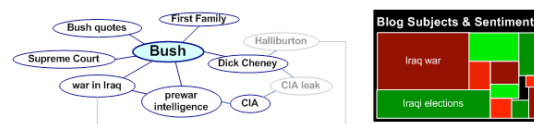
results.

Visualization and Sentiment Extraction:

By far the "sexiest" demos that vendors give use automatically-generated graphics to give a high level view of perhaps thousands of matching documents. These techniques gather data in similar ways to the previous items we described, namely applying statistical methods to extracted words, phrases, and entities. The first mockup, the "floating cloud of interconnected nodes," has been show by many vendors for over 10 years in different forms. Are they useful? Do they work at all? And once the newness wears off, would users ever get in the habit of navigating results in such a novel way? We just don't have the answer. We have seen many implementations that were not productive in practice, so we are a bit skeptical, but many programmers have been spending a lot of time refining these methods, and are at a point where customers could potentially evaluate different implementations side by side.

If real competition heats up in the enterprise space, it could spur a lot of improvements. As an example, if editors could conveniently adjust these meshes interactively and in real time from the native interface, it would probably make a big difference. Improving user interfaces, while not trivial, doesn't require any AI leaps.

Search Results Visualization



We've seen some even more impressive graphics-based visualization of results. If you are interested, talk to your vendors and ask for a demonstration. There are some very interesting techniques being packaged for the enterprise.

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

Other Ideas Bubbling Up

There are quite a few less sexy, but potentially still quite useful ways of improving search results.

Leveraging "Context"

If a system knows something about you – your job function, searches you've done before, or even the first link in a results list that you clicked on – then perhaps a search can be rapidly tuned.

Vendors are talking a lot about this.

Social Networking / "Voting"

Another general technique that is being re-implemented in many different ways is to identify similar users or to use the activity of other users to gauge relevancy. Some of these methods are similar to Google's link ranking; some vendors are counting 'click through' as a gauge of relevancy. This may work better inside the enterprise where titles are not often intentionally misleading. Other systems feel more like Amazon and Netflix's suggestion engines.

The User's "Frame of Mind"

Previously we talked about sentiment extraction, trying to gauge the opinions of content authors. Yahoo has tried something different, allowing folks to push a slider between "Research" and "Purchase," re-weighting results towards pure authoritative information or towards vendors offering specific products. We laud this experiment. In an enterprise we could imagine sliders that move between "marketing" and "technical." Searches for product names often bring back a mix of press releases and troubleshooting / FAQ articles. This type of slider would help select one type of data versus another. Some vendors would even say that the user provides additional "context" by moving the slider.

Multiple "Databases"

In some cases searches are being run against multiple data sources and the results are being combined. In enterprise software this is called "federated search." By itself it only adds to the exponential growth of the results lists, but when it is combined judiciously with other data and

presented in a logical or non-intrusive way it can provide additional facts. Even reports and graphs can be thought of as a "search."

Since vendors focus on their own technology, this trend is not being driven by them. Instead, smart users, programmers, and IT folks are doing it. For example, a system may check every search issued to see if it looks like an employee name. In the 95% of the cases when it doesn't, nothing extra is added to the results list. But in a few cases where a high confidence of an employee match is seen, the system will give a one line suggestion like "Satish Jones, QA/Mountain View, x1028, email: statishs." If a query like "sales figures for last quarter" is seen, a callout to a graphing package can be included in the results. These are guesses – the user might have actually intended a full text search looking for a document – but there is a chance that this was the answer he was looking for.

Overuse of these features, or really "fuzzy" matching which turns out to often be incorrect, have given these types of add-ins a bad reputation. This time around, discretion may make a better impression.

The Biggest Pushback

The biggest pushback we've heard from clients on interactive search is that their users simply won't use it, and this has led to long discussions about things like usage rates and measurement techniques. Some customers have even presented reports showing low click through rates for initial drill down links that were offered.

In the end we believe that *well-implemented* drill down search will be used, especially by knowledge workers whose livelihood depends on finding answers.

Here we list some factors that affect usage rates of advanced interactive search:

1. *Providing quality suggestions*

Though this seems obvious, we suspect some bad implementations have soured users.

2. *A corollary to item 1: "When in doubt, leave it out."*

Search 2.0 in the Enterprise: Moving Beyond "Single Shot" Relevancy

If the system doesn't have high confidence on statistically generated suggestions, consider not displaying them for that search. Bad suggestions are often *worse* than no suggestions.

3. *Less is sometimes more when it comes to suggestions.*

Building on item 2, many systems can generate dozens of suggestions, but you should consider only presenting a handful of them at a time. Adding a small "more..." link will give users the option of getting more help.

4. *Avoid overall clutter*

Don't put too many different types of navigators on the results list at once.

5. *Placement matters*

It is generally accepted that links along the top of results will get more attention than suggestions on the left or right, though items at the top should be kept very compact.

6. *Revisit your measurement of new-feature adoption rates*

When trying to measure the use of these new features, simple click-through percentages may not tell the whole story.

By far our biggest argument in favor of interactive search is that, yes, even Google does it. When you type in what looks like a zip code, address, or a well-known personality, Google suggests various links and images at the top of its results, and its "Did you mean" spell checker is widely known. Google has deployed these extra features in a conservative, stylish way, and suggests things only when it has a fairly high confidence. We suspect Google would not continue to offer these items if nobody used them.

In Summary

Exponential growth will eventually break even the best search engine's relevancy. Private data sets are now as large as the public Internet was when single shot relevancy failed there, and Google's workaround of link ranking won't work in the enterprise.

Instead of relying on "Single Shot Relevancy," start planning for a well-implemented interactive search process.

Be a bit open-minded to new ideas, there are a lot of new features headed your way. At the same time, don't be afraid to ask for on-site proof of a concept or pilot project using your own data.

To improve the efficiency of expensive knowledge workers any further, you'll probably need to look past the old-school "search dial-tone" functionality. It's a buyers' market for search right now and most vendors have new offerings – check them out!

New Idea Engineering helps companies make search work right. We focus on search best practices to help companies select, design, and deploy advanced enterprise search applications. Our methodology includes search 2.0 interactivity, periodic review of search activity and ongoing search data quality monitoring to ensure great relevancy and user satisfaction. To contact us, call 1-866-IDEA-ENG or see our website for more information at www.ideaeng.com.

Copyright 2007. All rights reserved.

new idea
ENGINEERING