

Search Replacement

Background and rationale for change

Our current search appliance — the device that determines what results search.missouristate.edu will display — is returning less relevant results than our customers expect. The voices of the campus community have asked for something better, and central funding has been approved.

In addition to lackluster results, poor vendor support and unexpected outages are big problems for our current search appliance.

Search is the most important means of navigating a large site like ours, particularly for prospective students. Students should easily be able to find a page about academic programs, applicable scholarships or cost information. Our Web editors have built the content and Web and new media has optimized it for mobile and desktop, but without good search, good content goes unseen. Good search advances several strategic objectives, but chief among them is Access to Success.

We need a search product that can gloss over spelling errors, suggest synonyms and intuit relevant results from ambiguous searches. That is a difficult problem to solve, but we live in a world where we are used to a ready solution to that problem on hand at all times.

Proposal

Replace our current search appliance with a Google-powered search system, aided by ElasticSearch for our People Search. Google's new pricing model should bring their services within reach of our campus.

This proposal has the support of the Student Government Association and Staff Senate. Administrative Council has also expressed interest in improving our search site.

Web and new media would implement the new search backend, in collaboration with our liaison in computer services, with the goal of having the new search site live by fall 2016.

Annual, on-going cost

Google Site Search	\$15,000
ElasticSearch + hardware	\$1000
Total	\$16,000

Detailed Proposal

Background and rationale for change

The best case against our present search appliance is simply to use it, especially compared against Google results that are limited to our site.

Campus Calendar - Note that our calendar site doesn't even make the first page of results

<http://search.missouristate.edu/?q=campus+calendar>

<https://www.google.com/#q=campus+calendar+site:missouristate.edu>

Email - This example showcases one way that we've worked around our current search appliance's lackluster performance--promoting every result that's relevant for a particular keyword. Yet with basically no coaching, Google managed to find the most relevant results on its own.

<http://search.missouristate.edu/?q=email>

<https://www.google.com/#q=email+site:missouristate.edu>

Library - This example hurts more because our attempts to promote and include more relevant results have pushed the thing customers are most likely to be looking for (namely, the main libraries.missouristate.edu site) off the front page entirely. Google has our libraries.missouristate.edu site as the top result without any massaging.

<http://search.missouristate.edu/?q=library>

<https://www.google.com/#q=library+site:missouristate.edu>

Papa John's - Granted, this example might seem trivial, but consider: information about our food vendors in the PSU didn't even make the first page of our search results

<https://www.google.com/#q=papa+john%27s+site:missouristate.edu>

<http://search.missouristate.edu/?q=papa+john%27s>

Behind the scenes, the appliance has its own problems. Multiple times in the past year, people and sites have unexpectedly gone missing from search results. Normal occurrences, such as a large new site being added to our domain, have broken the search appliance in the recent past. Vendor support has been sub-par at best. In one instance, the vendor installed updates without notifying us, breaking the search site for more than 24 hours, despite being aware that our failover appliance was down. Since then, they have continued to install updates without prior warning when we contact them about problems.

Their product has a solvable technical bug in how it handles Chinese characters, and other non-Latin character sets. We have built (and maintain) workarounds for these, and suggested solutions to the vendor, but they have yet to resolve the issue.

Detailed recommendation

Google Site Search + ElasticSearch

Google Site Search lets you embed Google's search results into your website without the Google branding for a fee. The license is flexible, allowing us to include search results from other data sources alongside web search results, which would let us maintain and improve our current search experience while still providing more relevant results.

ElasticSearch is a free and open source tool that we can use to provide results from predictable, structured data feeds like our People Search. Serving those results from ElasticSearch instead of Google Site Search would reduce the cost of a Google Site Search subscription (we would use fewer queries), without requiring us to solve the much harder problem of searching our entire website and returning relevant results.

Detailed cost and pricing model

The price of [Google Site Search](#) scales up as customers submit more queries to your site. You could think of it as buying queries by the bushel.

Based on our current analytics, we expect that we would need three 500,000 query bushels per year, each of which costs \$2,000. We would not have to pay all of this at once. Basically, we could buy one 500,000 query bushel, use it up, then buy the next 500,000 query bushel when we needed it. We could continue doing this up to the point that we enter the next pricing tier, which is three million queries for \$15,000. Each bushel is good for one year, so we expect that we would continue buying the cheaper 500,000 query bushels, even if our search load did brush with three million. Only if our search site was getting extensively more use would we need to consider that level of subscription. Even so, \$15,000 per year represents a good "ceiling" price, and is still cheaper than the Google Search Appliance, which is \$50,000 for a three year contract.

ElasticSearch is free and open source software, but hosting it does require server hardware. The \$1000 per year would help fund hardware infrastructure required to sustain the implementation long term.

Other institutions

The majority of our benchmark institutions are using Google Custom Search, the free version of Google Site Search. This severely limits what they can do, and we could not offer our current

search site's features with Google Custom Search. That said, those that have responded to our inquiries are happy with their search results and the low maintenance burden. The real power of both Site Search and Custom Search is that Google does most of the work for you. They host the hardware, they optimize search relevance; we can still promote results and (in the case of Site Search) decide how we want to display them.

ElasticSearch

Very large companies use the ElasticSearch open source product to power their site search, including eBay, Wikipedia, and OpenTable.

ElasticSearch has also proven useful in smaller applications. Ohio State University has used ElasticSearch in a course catalog search. University of Missouri also uses it in a few applications.

Missouri State University is exploring ElasticSearch for use with our Blackboard Learn implementation, and we have some in-house knowledge of it.