Did you go go?

An Altavista-slayer

Imagine a web search service of googol proportions that:

- doesn't use javascript or cookies
- doesn't track you
- is built completely on open-source code
- offers both a public GUI and a API, both equally open, allowing you to query in your favorite language and read/write in your favorite format
- through the GUI and API provides a way to query both in natural language and in structured query languages
- has a ranking model that encourages a good ratio between content and markup (less markup/script is better)
- is community-driven, enabling the public to:
    - add to the public index
    - create personal indexes
    - create queries that join between all indexes
    - curate indexes
    - add to the code that runs their favorite search service
- encourages good nettiquette (e.g. small page size)

This is such a service.

I'm Marcus, author of the software behind Didyougogo. For the past ten years I've been trying to improve my programming and math skills to get to a level where I could write a proper web search engine for the written word using absolute cutting-edge IR methods. The final result is something I have not seen or read about: a language represented as a 65K wide vector-space, serialized into a binary tree that is balanced according to node's cosine angle between them and their closest neighbours. Querying is very fast, even for long phrases. Fuzzy, prefix, suffix and wildcard type queries comes for free with the vector-space model. The system uses relatively little resources and can run on as little as 1 CPU and 1GB RAM.

I hereby present my (0.1 alpha MVP) Altavista-killer *: didyougogo.com

* Because calling something a Google-killer makes you sound really wierd. And because the Altavista search engine came as a standalone installation (which is also the case with the Didyougogo software).

What to do now?

Me, personally, I would like to spend the next ten years building the strongest web search service in the world. Here's what I have in mind:

- build a public analytics view that shows num of pages, num of sites, top 100 semantic topics and other interesting metrics
- create a way for on-prem installations to talk to public installations (to interchange data)
- build a tool for the public to create queries that act as "instant answers" on the didyougogo site
- add voice-capabilities
- create a strong sales strategy:
    - approach companies that would benefit from the indexed data and ask them to integrate it into their day-to-day business of driving traffic and sales

- entice them to try an on-prem installation enabling them to query and join between private and public data so that they are able to see trends within their market place
- gamify the search result page in a tastefull manner that serves both the public and companies trying to market their ideas
- build a strong, knowledgable and curious team and a community

Check it out please (feedback greatly appreciated)!

Submit a page, do a search.

Didyougogo
Source code
Give feedback here

Investors

This service needs computing, storage and a small-ish team of developers. If you are willing and able to offer sponsorship, reach out to me at marcuslager at the biggest email provider in the world * dot com.

* It's a real dragon. Someone should slay it.