IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

Improving Visual Search Results

Saichethan Reddy

Computer Science, Indian Institute of Information Technology, Bhagalpur, India

*Corresponding Author: m.reddy@iiitbh.ac.in Tel.: +91-9182520998

Available online at: www.ijcseonline.org

Accepted: 16/Dec/2018, Published: 31/Dec/2018

***Abstract*—** This paper introduces a new method to improve visual search results and understand structured data. While many online resources teach basics of web development, few of them are designed to help novices learn the web development concepts and design patterns used by experts to implement complex visual features. Professional web pages embed these design patterns and could serve as rich learning materials, but their metadata are complex and difficult for novices to understand. This paper presents Metadata.py*, a Metadata inspection tool that helps novices use their visual intuition to make sense of professional web pages/sites. We introduce a new visual relevance testing technique to identify properties that have visual search results, which Metadata.py uses to hide visually irrelevant code and surface unintuitive relationships between properties. In user studies, Metadata.py helped novice developers replicate complex web features 75% faster than those using Chrome Developer Tool and allowed novices to recognize and explain unfamiliar concepts. These results show that visual inspection tools can support learning from complex professional web pages, even for novice developers.

Metadata,py: Python Script using Beautiful Soup, a python library for pulling data out of HTML and XML files.

***Keywords*—** Visual Search, SEO, Optimization

I. INTRODUCTION

Novice programmers often rely on online resources while learning to code, particularly in the domain of web development. Metadata is data that describes other data. This means details about the services and types of content you offer to Search users

When tutorial examples do not meet developers' needs, they often turn to professionally-authored websites for design inspiration. Professional website embeds rapidly-evolving best practices and conventions not covered by tutorials and are continually updated as new solutions arise. Most importantly, all websites are freely inspectable using browser developer tools, which expose the Metadata responsible for a page's appearance and part of rich results. As a result, professional websites represent an appealing class of resources to help novices learn implementation approaches.
In practice, however, professional websites SEO's are too complex for novice developers to understand through inspection. In a need finding study, we observed that novices approach inspection from a visual perspective, asking questions about how the rich results on a webpage are implemented in code (e.g. "How does this website have a rating?"). This visually-driven approach leads to challenges. First, even a state-of-the-art web inspector might display over a hundred Structured data properties at a time, many of which have no observable effect on the page. As a result, novices struggle to locate the lines of code responsible. And although most websites use schema guidelines there are websites which use Open Graph Protocol and Twitter cards it's difficult to find them. The main purpose of this paper is to ease learning about visual search and improving it.

II. RELATED WORK

Search Results Clustering

Search result clustering has a long history in information retrieval area[2]. The main three issues are: first, the algorithm should take the document snippets instead of the whole documents as input, as the downloading of original documents on the Web is time consuming; second, the clustering algorithm should be fast enough for online calculation; and third the generated clusters should have readable descriptions for quick browsing by users.

Google Structured data tool

Structured data provides a way to standardize information about a page and classify the page content. We also use structured data to enable special search result features and enhancements. But Google structure data tool can understand only when written in JSON-LD, Microdata formats.

III. RESEARCH

Before we search, web crawlers gather information from across hundreds of billions of webpages and organize it in

the Search index. The crawling process begins with a list of web addresses from past crawls and sitemaps provided by website owners. As crawlers visit these websites, they use links on those sites to discover other pages. The software pays special attention to new sites, changes to existing sites and dead links. Computer programs determine which sites to crawl, how often and how many pages to fetch from each site.

For most sites, Googlebot shouldn't access the site more than once every few seconds on average. However, due to network delays, it's possible that the rate will appear to be slightly higher over short periods. Googlebot was designed to be distributed on several machines to improve performance and scale as the web grows. Also, to cut down on bandwidth usage, they run many crawlers on machines located near the sites they're indexing in the network. Therefore, logs may show visits from several machines at google.com, all with the user-agent Googlebot. Our goal is to crawl as many pages from your site as we can on each visit without overwhelming your server's bandwidth. Request a change in the crawl rate.
"A tool for metadata analysis" Web-based metadata quality tool that provides statistical descriptions and visualizations of Dublin Core metadata harvested via the OAI protocol. The lightweight nature of development allows it to be used to gather contextualized requirements and some initial user feedback is discussed. Published by the University of Waikato. it helps in visualization but does not help novice learner to learn or understand.

Google Search works hard to understand the content of a page. You can help us by providing explicit clues about the meaning of a page to Google by including structured data on the page. Structured data is a standardized format for providing information about a page and classifying the page content; Google uses structured data that it finds on the web to understand the content of the page, as well as to gather information about the web and the world in general. Google Search also uses structured data to enable special search result features and enhancements.

Name	time(s)
Hemanth	11.6
Pritam	14.8
kush	15.4
Krishang	12.5
Amit	13.2

IV. METHODOLOGY

First, I searched for API or tools from which I can get Metadata but they are not feasible for novice developers because of dependencies they might feel overwhelmed. Metadata.py is a simple python script which can be compiled and executed in any operating system (Linux or Windows) doesn't need any large dependencies or Docker.

In my research, I let 10 novice web developers (undergrad students) use it they felt it quite helpful they said it really help them understand and improved their efficiency. There is no proper SEO tutoring website available Metadata.py can help self-learners to learn and get hands-on experience.

We surveyed 10 undergraduate student web developers about their experience inspecting websites and learning from HTML and CSS tutorials. We then followed up with an in-person study with five of these developers. During the in-person study, we asked users to talk aloud while replicating a single structured data property and asked them to say whether it's using Schema or Open Graph Protocol or Twitter Cards.

Name	time(s)
Shushant Samir	44.6
Abhishek	48.4
Abhro jeet	46.0
Anuranjan	49.7
Hitesh	47.3

V. RESULTS AND DISCUSSION

Both groups are told to check metadata of

1. Google
2. Pinterest
3. LinkedIn

Group A (used Chrome developer tools)

Mean = 47.2.

Median = 47.3.

Standard Deviation = 1.994;

Group B was told to use Metadata.py
Mean = 13.5, Median = 13.2, Standard Deviation = 1.581;
Each of these five users was given three professional example websites to inspect using Chrome Developer Tools (CDT). Users could freely search the web for documentation or additional resources. We observed users and analysed their progress using informal milestones for each feature they took nearly 45-50 minutes.

The next five students were told to use Metadata.py to solve and inspect the same three websites 12-15 minutes, although both groups are new to Search Engine Optimization the group that used Metadata.py, it helped novice developers replicate complex web features 75% faster than those using Chrome Developer Tool.

For website: https://www.google.co.in

1. **Using Metadata.py**

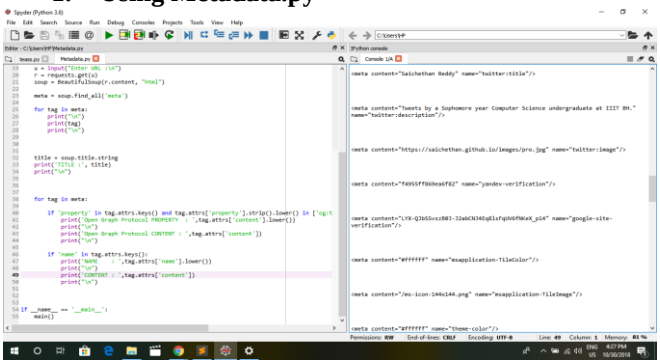

Image URL: Using Metadata.py

2. **Using Chrome Developer Tools**

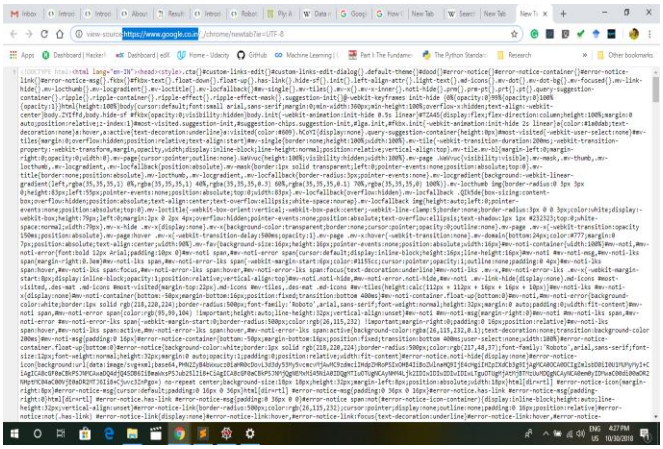

Image URL: Using Chrome Developer Tools

VI. Conclusion and Future Scope

As users' devices and context evolve beyond the desktop into the world of mobile devices and everyday tasks, Search engines are also evolving by providing rich search results that support users beyond the desktop: from making a reservation, to calling your business's customer service number, to selecting the perfect chocolate cake recipe while at the grocery store.

As of now Metadata.py is written in python using Beautiful Soup, a python library for pulling data out of HTML and XML files and executed on Anaconda environment. This can be improved by making it available online without downloading dependencies.

Google Search can enable a rich set of features for your page in search results if it understands the content of the page and, in some circumstances, if you explicitly provide additional information in the page code using structured data. These features fall into two general categories:

- Content type
- Enhancements

Visually rich snippets are currently not supported in every market, although annotations will be processed and validated in every market.

Since this is a new and an actively developing area Metadata.py could help novice learners to learn more about Visual search results. These results show that Metadata.py can support learning from complex professional websites, even for novice developers.

References

[1] M. L. Bernard. Examining the effects of hypertext shape on user performance. Usability News, 4(2), 2002.]]

[2] An Attribute-Assisted Reranking Model for Web Image Search, Isroset-Journal (IJSRCSE) Vol.4 , Issue.3 , pp.16-19, Jun-2016

[3] L. H. Armitage and P. G. B. Enser. Analysis of user need in image archives. Journal of Information Science, 23(4):287--299, 1997.]]

[4] P. Borland and P. Ingwersen. The development of a method for the evaluation of interactive information retrieval systems. Journal of Documentation, 53(3):225--250, 1997.]]

[5] R. C. Veltkamp and M. Tanase. Content-Based Image Retrieval Systems: A Survey. Technical Report UU-CS-2000-34, Dept. of Computing Science, Utrecht University, 2000.]].

[6] Greenberg, Jane, M. Pattuelli, B. Parsia, W. Robertson, Author-generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization, Journal of Digital Information, volume 2 issue 2 (November 2001)

[7] Heery, Rachel and Manjula Patel, Application Profiles: Mixing and Matching Metadata Schemas, Ariadne, Issue 25 (September 2000)

All related to this paper can be found at:
https://github.com/Saichethan/Saichethan.github.io/tree/master/projects/Improving_Visual_Search_Results

Authors Profile

Saichethan M. Reddy is pursing Bachelor of Science from Indian Institute of Information Technology, Bhagalpur, Bihar, India with Computer Science as his major. He is currently in Sophomore year. Currently working as student researcher at Indian Institute of Information Technology Bhagalpur. This is his first research paper. His main research work focuses on Cryptography, Algorithms, Number Theory, and Pattern Recognization and Machine Learning

