GoI SEARCH as a SERVICE (GoI-SaS)

Government of India (GoI) Search Platform initiative has enabled the citizens and other stakeholders to quickly access the data on Government websites. Under this initiative, Search as a service (GoI-SaS) facilitates the website owners in the Government to define their own search interface on their websites/portals and it helps then in analysing the user search queries and click behaviours on Search results.

KAMALAKANNAN M.
Technical Director
kannan@nic.in

RITU GHAI
Senior Systems Analyst
ritu@nic.in

NARENDER KUMAR JAIN
Systems Analyst
nk.jain@nic.in

GoI Search as a Service (GoI-SaS) is a managed Search service on the GoI Search Platform that makes it easy to set up and manage search interface on any website/portal or online web application. SaS enables you to search large collections of data such as web pages (HTML documents) and document files like PDF, Open Office documents(.odt, .ods, .odp) & MS Office documents(.doc, .xls, .ppt) etc. With GoI-SaS, the website owners can quickly add search capabilities to their website even without becoming a search expert or worry about hardware provisioning, setup, and maintenance. With a few lines of code and few clicks in the GoI-SaS Admin Console, you can create a search interface, and the GoI-SaS search service automatically provisions the required technology resources and deploys the highly tuned search index. You can easily change your search parameters, fine tune search relevance, and watch the user search queries and behaviours on the website.

GoI-SaS INTERFACE

Search as a Service (GoI-SaS) enables web site owners to define their own search interface on their websites/portals. GoI-SaS provides a deceptively simple form-based interface for building a domain-specific search interface on top of the GoI-SaS search platform. This means that the builder gets to focus on selecting valuable content and tuning the ranking criteria, while GoI search platform performs the immense task of crawling, indexing, ranking, and display of results.

HOW IT WORKS?

The main objective of building a GoI-SaS is to determine which

sites/URLs (including flexible URL patterns) are searched, and to define a set of rules that guide the ranking of results. Once the GoI-SaS is defined, the site owner places a search box on their site.

The GoI-SaS users can also include onscreen Virtual Floating Hindi Keyboard along with the search box to help the visitors in submitting the search queries in Hindi language.

When the users perform a search, they are brought to a web page that looks much like the traditional search results page. However, there are two important differences:

1. The site owner can choose to have the search results appear in an iframe on their own site.

2. The site owner can customize the look and feel of the page to make it look more like their existing site.

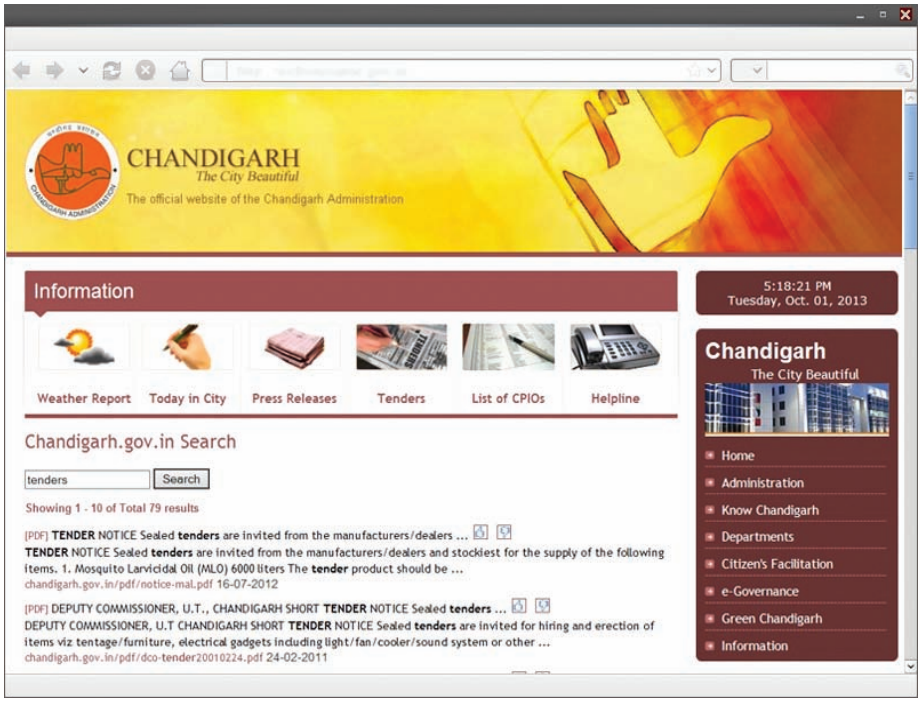
REQUEST FLOW IN GoI-SaS

Initially, the search page is downloaded from the respective website and depicted in the visitor's browser. Once the search query is submitted on the search form, it is directly sent to the GoI Search Platform, http://goisearch.gov.in as Ajax request. The GoI Search platform then returns the search results in JSONP format and finally the visitor's browser renders the search results.

The steps are explained in the sequence diagram. To customize the look and feel of the search results page, the site owners can modify default CSS to suit their respective website design.

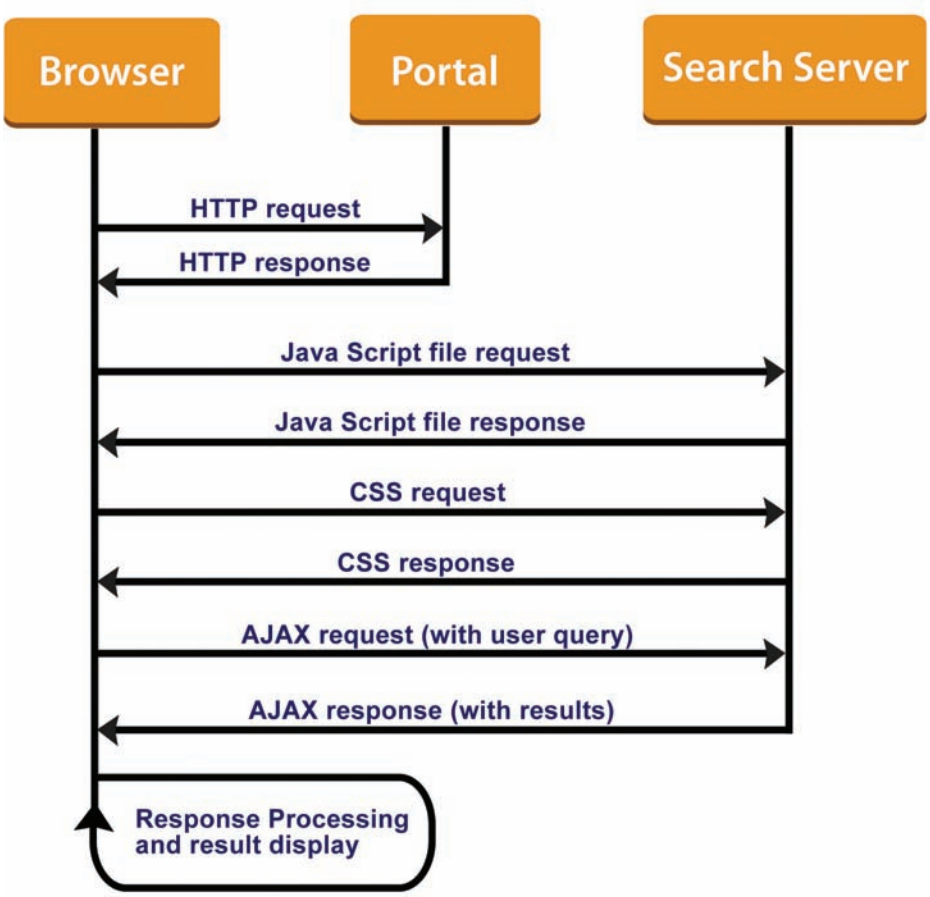
HOW TO REGISTER YOUR WEBSITE WITH GoI-SaS?

If your organization is functioning under the aegis of Union/State Government, you can contact the respective NIC Web Co-ordinator to register your website with GoI-SaS. Once the registration process is completed successfully, the website will be further registered with GoI Search Platform for periodical crawling/indexing. The GoI-SaS Admin Console will help you in generating the search code and implementing it on your website search page. You can also login to the SaS Admin console using NIC LDAP NIC/Email account. The GoI-SaS admin console is available at http://searchservice.nic.in.



NEETA VERMA
Deputy Director General, NIC

With exponential increase in the number of Government sector websites, the need for enhanced accessibility for a better user experience is being increasingly felt. GoI Search as a Service (GoI-SaS) is envisaged to provide quick, easy and relevant search results to the users, thereby substantially improving the user experience of the website . Website proprietors can use the GoI-SaS service to enable their site Searchable and also define their very own customized search interface. GoI-SaS also enables the owner of the website to monitor the incoming traffic on their websites, search queries specific to their reports, relevancy of search results specific to their website through a common dashboard. I am sure that the initiative shall establish in milestone in enhancing the usability of Government websites.



Request Flow in GoI-SaS

GoI-SaS ADMIN CONSOLE

The SaS admin console helps users in generating/customizing the search code for their website/s. It also helps the users in registering a new websites with GoI-SaS. Once the site is registered with GoI-SaS console successfully, the site will be added in the crawl/index list. The crawl/index status of the site will also notified to the concerned official.

GoI-SaS ADMIN REPORTS

Search Statistics section of the admin console allows the users to monitor search traffic onto their website. It generates a variety of reports which helps the website owners to analyze the user search queries, user click behaviour, target URL frequency etc. The following are the list of reports available in GoI-SaS Admin console:

1. Query Frequency:

This report lists the search keywords/queries on the site for the selected period and their frequency. It also gives the details of target URLs selected for the query and the visitor's host IP address.

2. User Click Behaviour on Results:

It helps in getting the details of user click frequency on search results. It gives the number of user clicks on a particular result position for the search queries i.e. How many users have selected the results at first position? How many on second? and so on.

3. User Navigation Behaviour on Result Pages :

It lists the details of the user navigation in first five pages of search results. It gives page-wise user click behaviours like How many have chosen the first page results?, How many on second page and so on up to first five pages.

4. Exceptional User Navigation Behaviour on Result Pages :

It lists the page-wise user click behaviours between the results pages six to ten. i.e. it reports how many results are selected in sixth page, seventh page and so on up to tenth page.

5. Result Returned Status:

The search queries by number of results returned are sorted in this report. It helps in finding out the search queries which returned no results, maximum results and so on.

6. Target URL Frequency:

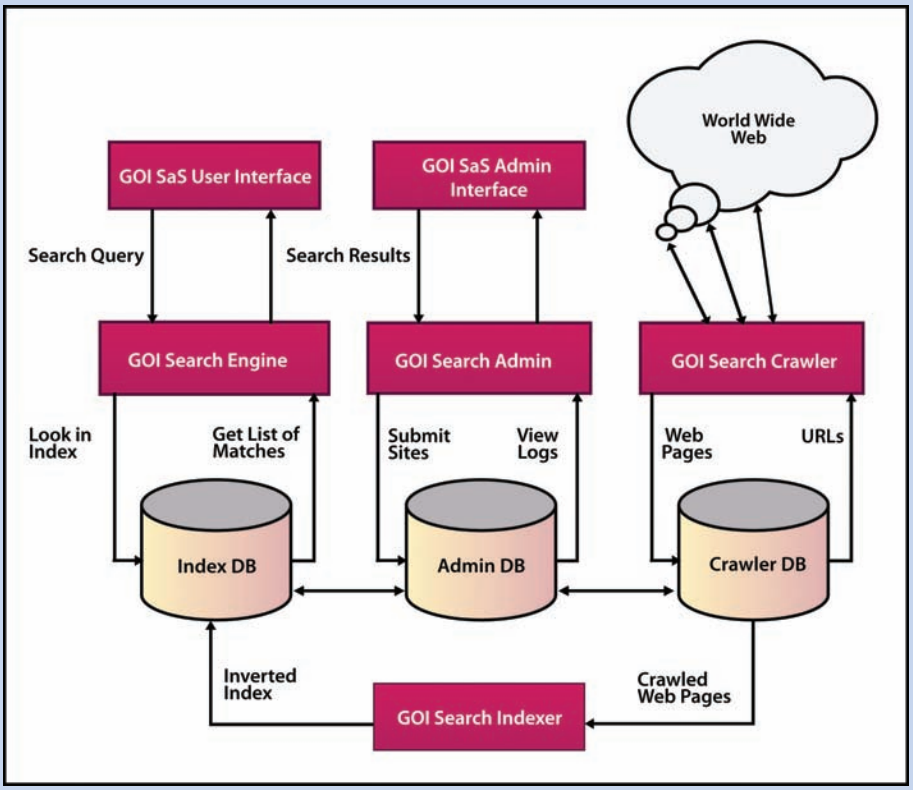
It sorts the target URLs of the website by query frequency and also lists the search queries used to reach on the page.

The results of the report can be exported as PDF/CSV files and system has the facility to print the reports. It also has the facility to graphically present the report data either as Bar or Pie Chart for better clarity.

WEBSITES USING GoI-SaS FACILITY

Currently, more than 100 Union/State Government websites and district administration websites are using the SaS facility. Some of the important sites which are using the SaS are:

GoI-SaS Architecture



The GoI-SaS components are deployed above the GoI Search Engine components. The main components of GoI search engine are crawler, Indexer and Search Engine. The Crawlers look at web pages and follow links on those pages. The crawl process begins with a list of web addresses from past crawls and websites submitted by website owners. The crawl configuration determines which sites to crawl, how often, and how many pages to fetch from each site. GoI Search Engine crawls and indexes the sites through the file called "robots.txt". With the robots.txt file, site owners can provide more specific instructions about how to process pages and links on their sites.

GoI Search engine essentially gathers the pages during the crawl process and then creates the index using the Indexer module. The indexer generates and maintains the inverted indexes of the crawled pages in the Index DB. The Search Engine evaluates the search query to quickly locate documents containing the words/phrases in the query and then rank the documents by relevance and returns the Search Results.

The GoI-SaS Admin console facilitates the users in submitting the websites to GoI Search Engine and also helps them in analysing the search queries and results. The user Search queries and search results along with the user profiles are logged in the Admin DB for further analysis. The GoI-SaS Admin console generates various reports like top queries, results returned, user click behaviour, target URL frequency etc. The SaS Search Interface facilitates the users in submitting the search queries and then it returns the search results as XML file which can be further rendered on the respective websites using the customized XSLT.

In the GoI Search Engine platform, the Search Engine Crawler & Indexer is powered with open source Apache Nutch & Solr. The Search portal as well as Search Admin Interface is developed in Java and deployed on Apache Tomcat server. The search engine infrastructure is housed on the load balanced and highly available environment for better performance and high availability.

- http://pmindia.gov.in
- http://india.gov.in
- http://chandigarh.nic.in
- http://www.mp.nic.in
- http://tnrd.gov.in/

WHO CAN AVAIL GoI-SaS ?

If your organization is under any Union/State Government entity or District Administration or Judiciary/Legislative body, you are eligible for the service. Eligible organizations can contact their respective NIC web coordinator to avail the SaS or login to the SaS Admin console using your NIC LDAP/NICEMAIL user account to register your website for GoI-SaS.

FOR FURTHER INFORMATION:

Support, GoI Search Engine,
Data Centre & Web Services Division,
3rd Floor (A4B4),
NIC, A – Block, CGO Complex,
Lodhi Road,
New Delhi -110 003.
Phone: 011-24305376, 5373
Email: support@goisearch.gov.in