Shopping Search Engine Technology Based on Services

LIHAILONG, LIUYU
Beijing Institute of System and Engineering
P.O.B. 9702 19#, Beijing
CHINA

Abstract: Focused on shopping website and e-commerce area, this article discusses the technology of specific search engine service platform. The overall goal of the platform is to provide a new web shopping pattern that is more convenient, powerful, and swift for consumers. This effort promotes the flat structure development to e-commerce platform. At the same time, the search engine service platform will strive to provide a Service-based Integration Bus, which combines the shopping-information services and shopping-process services together for consumers, suppliers and delivery companies.

Key-Words: E-Commerce, Shopping, Search Engine, Web service

1 Introduction

With the development of Internet, Search Engine technology has got the fast fierce development and applications. It has already succeeded to apply to the areas of education, academic, research, business, manufacturing etc. Most research focuses on topic-oriented, crawler-based search engines [1, 2] and Spider [3]. There have many successful search engine website, such as Google, Yahoo, Baidu, etc.

The characteristics of the general search engine websites is its large information capacity, the wide scope. The information source includes the website, web page, data, document, etc. the searched information contains website/ web page information summary, the document summary, etc. Because of the diverse of the information source, the searched information is non-structure.

Different from general search engine website’s concern point, this article specifies the shopping website exclusively. Along with the gradually mature of e-commerce, this kind of shopping mode has already been accepted. At present, the shopping website becomes more and more, including the electric appliances, clothing, book, IT product, etc. The on-line shopping has its characteristics, fast, labor-saving, and easy contrast. This is a good complement to traditional shopping methods. Compared with general Search Engine website, our Shopping Search Engine Platform can provide structural or semi-structural information. As a result, the searched information is more precise, detailed and explicit. This shows the basic different from general Search Engine.

For general shopping website, its information can satisfy the basic needs of the on-line shopping consumer, including the goods browsing, choosing, payment, delivery, etc. However, every independent shopping website has its natural limitations, for examples:

(1) From the viewpoint of the consumers: the suppliers are limit, the goods are limit (each website mostly lays particular emphasis on some goods), no overall price contrast, and fixed goods shipping location, etc.

(2) From the viewpoint of the website: Most shopping website’s visitation is limit, because the market has already been divided the line, every consumer will pay attention to particular shopping websites only.

This caused to the needs of the shopping search engine (a platform). The purpose of building the platform is to emerge all essential e-commerce websites’ shopping information together, to provide more accurate, complete information, to make consumers faced to numerous websites and suppliers directly. As a result, consumer can carry on the depth search and contrast analysis. This platform is benefit for both consumers and suppliers. It extends the visitation net of website. On the other hand, it provides the more all-directions shopping information for the consumer. These functions exceed the capabilities of general search engines.

2 Basic solution

In order to build the shopping search engine service platform, a few problems must be resolved, includes:

(1) Build up the shopping information model: establish the information model according to the characteristics of the shopping information. That
model is the basic structure of the shopping information object (SIO).

(2) The data source collect: Get the details information of each shopping website, resolve, identify, structure the information, and reflect to the information model mutually. By filtering and updating, bring the SIO into the information database center.

(3) Shopping Web service center: To storage the shopping information, provide the information organization, searching, subscription services.

(4) Shopping Web service center QoS service: Guarantee the integrity and renewal of information; guarantee not to extend the information error among information tidying up. Provide the auditing service.

(5) Shopping information integration bus based on services: The shopping search engine service platform should confirm with the SOA framework. It provides the meta-information registration center for the shopping information, the payment service, and the shopping process service includes registry, auditing, release, and search. Together with the publish service, query service, and integration service, the shopping information integration bus is built.

2.1 Shopping information model and the data collector

First, we should analyze the shopping information organization method of most big shopping website, and then establish the shopping information model (SIM). The model is a basic structure of the shopping information object (SIO). SIO is also the basic unit for shopping information’s organization, transporting, and accessing across network.

Information Model( SIM) is as bellows

![Fig.1 shopping information model](image1)

In the information model, the brand, name and price are the key attributes of the commodity object, and these are the foundation of the shopping search engine platform. Among them, the price has its particular scheme, including the market price, member's price, net average price, and history price.

**Data Collector:**

Data collector contains the core technique of the shopping search engine service platform. Its function is to grab, analyze, filter/clean the information of each shopping website, then reflect to the shopping information model and establish the shopping information object, and then deposits the SIO into centric information database. The data collector mainly composed by four parts: the network search spider, parser, filter, and updater. The network search spider is responsible for grabbing all of the commodity pages of the website. The parser is responsible for identifying the commodity page information, and transforming the non-structure information into the structure/semi-structure information (namely SIO). The filter is responsible for evaluating the data’s quality and determining the classification. The updater storages the SIO into service center information database. The framework of Data Collector is shown as below:

![Fig.2 the framework of Data Collector](image2)

Because of every shopping website has different design style, development language, information organization method, data semantic, commodity taxonomy etc. Before searching, the information parsing rule must be established for different website. Includes:

1. **Website search strategy.** Either use Depth-first or breadth-first to grab the pages.
2. **Commodity classification reflecting rules.** Each website has its own Commodity classification method, the classified layers is also different. First, we should reflect the classification of website to the classification of information model and build up the reflecting relation rule. Thus, in the course of website...
data parsing, the particular Commodity category can be transformed into the unify category.

(3) The shopping information page identifying rules. All shopping website would have quite a few of page branches, but only parts of pages are the pages that includes the valid shopping information. Therefore, valid pages must be recognized before information extraction. The basic method is to recognize via page’s path regulation and information format. It needs building up the corresponding rule database.

(4) Commodity attributes identifying rules. The name, brand, price, specification of the commodity are all key attributes to be identifying. For different website, the method to identify these attributes would be different. The identifying rule must established, depending on page web’s attribute tags and the context relations.

(5) The Commodity information renewal rule. The particular Commodity information of website would change aperiodically, including the price fluctuation, new commodity on, sell out commodity off etc. The platform will statistics the frequency of website information changes within a long time, then establish and optimize the website information renewal rule.( mainly the renew frequency)

(6) The website information renewal strategy. For some websites, the page link is fixed, but for some other websites, the page may be dynamic. For different type of website, the information renewal strategy has some diversity.

(7) Commodity alias identifying rules. See the QoS service.

(8) Information validity rules. See the QoS service.

2.2 Shopping Web service center

The shopping Web service center is mainly composed by the service center information database, the shopping information Web service, and the QoS service. The shopping information service platform is realized based on Web service architecture, in order to provide better ability of system integration and information sharing.

The service center information database is a large-scale information database that is constitute by the commodity information of many websites. This database is collected by the data collector, and renewed in time.

The shopping information Web service mainly provides the powerful commodity information search service, search service can accept various query conditions, and will return the result via the compressed SOAP message to customers.

Correspond, search engine platform will provide the search service client program, the client program decode the feedback and emerge to the consumer. At the same time, the platform can also provide subscription/distributive service, with which consumers can subscription the information needed, and then the service center send out the information to consumers or via E-mail.

QOS Service: The Quality of Service is very important for shopping search engine platform. The shopping information services that platform provide must guarantee the performance of query, the information’s accurateness, integrity, comparison, and must fresh the information in time. For the factors above, the services bellows are needed.

(1) Concurrent capability. The platform should handle huge users’ access concurrently.

(2) Security Service. The platform should limit the access frequency of specific user and avoid the malice attacks. The information security of payment service and delivery service should be guaranteed too.

(3) Information filtering and auditing service. The platform should check the information collected strictly. First, the information sources must be credible. Second, the information must be integrated, and the information lack of name or price shouldn’t add into information database.

(4) Alias service. For the same commodity, different website should have different name. The difference of commodity name will influence the price comparing greatly. The platform provide alias identifying service, the service build up the math model based on commodity class, name and price, then compute the equivalence quotient (E(Price, Name))of commodities with same class and different website. If E (Price, Name)>e, then match the commodities farther more by computing the comparability factor by details parameters and specification of commodities. Finally, enumerate the alias pairs and construct the alias database, using the artificial intervention when needed.

(5) The information renewal service. It includes the renewal of price, new commodities, and the outdated information. The system should record the fresh time of every shopping information object (SIO). Father more, information renewal will cost more time and computing resource, so the system should optimize the renewal strategy according to the characteristics of the website.

2.3 The shopping information integration bus based on services
The shopping search engine platform is good at providing commodity information service, which is a basic part that can connect consumers, e-commerce websites, suppliers, delivery companies, and payment centers. Consequently, based on the platform, we can build up the shopping information integration bus providing entire process service for shopping.

Conform to SOA ideology, the platform can provide commodity information center, search service, information publish service, QoS service, and metadata center of shopping process services. The metadata center storage the delivery service metadata and payment service metadata of every e-commerce website, and provide services Meta information for platform and service integration.

The logic architecture of shopping information integration bus is as bellows.

Then, through the integration bus, how do integration partners participate in the shopping process? First, the consumers can search and compare commodity by search service, and call payment service and delivery service via the services that registered in the metadata center. Second, the commodity information can be built by pulling website data, on the other hand, website providers can also push their commodity information to data center via publish service. The payment center and delivery center of website can also register their service into metadata center, so that this service can be called by consumer directly. Consequently, the shopping steps are connected.

3 Conclusion
The search engine service platform strives to provide a Service-based Integration Bus. The central idea is to aggregate the heterogeneous shopping information together. It also combines the shopping-information services and shopping-process services together for consumers, suppliers and delivery companies. Now, we have developed the prototype of the service platform, and we have more work to optimize and test the platform.

References: