

## SPATIAL SEARCH ENGINE – ENABLING THE INTELLIGENT GEOGRAPHIC INFORMATION RETRIEVAL

Yuqi BAI<sup>a</sup>, Chongjun YANG<sup>a</sup>, Donglin LIU<sup>a</sup>, Lingling GUO<sup>b</sup>

<sup>a</sup> Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, 100101  
(yqbai, cjyang, sxldl) @digitalearth.net.cn

<sup>b</sup> Dept. of Urban and Environmental Sciences, Peking University, Beijing, 100871  
joveidolon@yahoo.com

Commission II, WG II/3

**KEY WORDS:** Spatial Search Engine, Geographic Information Retrieval, Web Mapping, Web Service, UDDI

### ABSTRACT:

In this paper we discuss a particular type of search engine, Spatial Search Engine (SSE), aimed to search the World Wide Web for geographic information to meet the users' needs. We investigate the architecture of General Purpose Search Engines (GPSE), such as Alta Vista, Google, Infoseek, etc. We then analyse the mechanism of web mapping systems, which are geographical information providers on the web. After comparing SSE with GPSE, we point out the great differences between them and then argue that without the support of web mapping systems, it is impossible to fulfil the SSE. We propose a conceptual design of SSE and then propose a possible solution, a combination of Geospatial Web Services and UDDI. With Web Services, a new technology framework, formally monolithic geospatial information providers turn out to be self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. With UDDI, those Geospatial Web Services can be discovered dynamically and then be accessed automatically. At the end of this paper, we discuss the advantages and shortcomings of our prototype implementation, as well as future work building upon our experiences with SSE prototype.

### 1. INTRODUCTION

From Timothy Burners-Lee's innovative idea about Hyper Text in 1989, the World Wide Web (WWW) is becoming the most popular publishing media and the biggest digital library in the world. Lawrence and Giles estimated that the number of publicly indexable pages on the Web at that time was about 800 million pages on about 3 million servers [1999a] and the number of web pages currently rise up to about 7700 million and 9.5 million servers are now running day and night on the Internet.

At the same time, searching for useful information on the web has become increasingly difficult due to its dynamic, unstructured nature and its fast growth rate. Currently, going to known sites, following the links, browsing the maintained sites directories and using search engines are four common ways for us to do it. Among them, search engines are the most useful ones. Different Search Engines have different system designs and implementing strategies, but they have almost the same system architecture.

#### 1.1 General Purpose Search Engines (GPSE)

In this section, we look at the system architecture of general purpose search engines.

Generally, a web search engine contains three important parts: a Crawler, an Indexer and a Query Server. As shown in Figure 1, the crawler is full in charge of collecting useful materials, such as html pages, images and even music from the web. The indexer's role is to index what the crawlers has collected and then store them into some particular databases, according to the type of the materials. The Query Server aims to get the exact query string via the user interface, analyze it, query the

classified and indexed documents and then give the query results, usually in the way of links.

In order to return as many relevant results as possible (Recall) and as more accurate results as possible (Precision) to the user, different search engines have their own characteristics and employed their preferred algorithms in indexing, ranking and visualizing web documents.

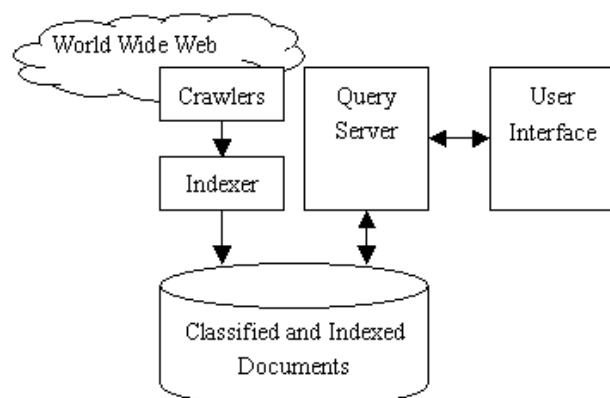


Figure 1. Architecture of General Purpose Search Engines

#### 1.2 Web Mapping System

For Geospatial search engine, in order to fetch needed geospatial information on the web to the end user, the first important thing is to identify where the geospatial information is. Currently, most of the geospatial information is provided via a web mapping system, such as ESRI ArcIMS (www.esri.com),

GeoMedia WebMap ([www.geomedia.com](http://www.geomedia.com)), MapInfo MapXtreme ([www.mapinfo.com](http://www.mapinfo.com)), GeoBeans MapServer ([www.digitalearth.net.cn](http://www.digitalearth.net.cn)), GeoStar GeoSurf ([www.geostar.com](http://www.geostar.com)) and MapGIS MapWeb ([www.mapgis.com.cn](http://www.mapgis.com.cn)). These web mapping systems now have a similar system architecture.

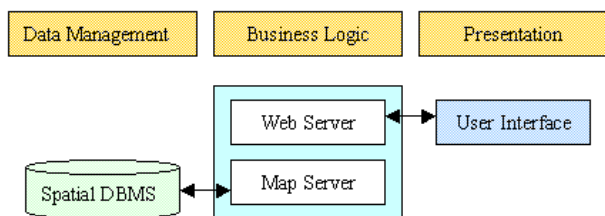


Figure 2. Architecture of Web Mapping Systems

These web mapping systems are usually organized into three major tiers: presentation, business logic and data management.

The presentation tier, mostly taking the form of web browser, contains the programming that provides the graphical user interface (GUI) and spatial-specific entry forms or interactive windows for users to specify query conditions.

The business logic tier contains business rules that govern interaction with data management and acts as a server for request of client. Typically, it consists of a web server and a map server. The web server is responsible for the HTTP connection between the client and server, and the map server is in charge of all spatial processing, corresponding to the each user's request.

The data storage tier typically consists of spatial database management systems that enable spatial data to be stored, managed, and quickly retrieved from commercial databases.

### 1.3 Comparing SSE with GPSE

Both SSE and GPSE are all web search engines, and the only difference is that the former focuses on the spatial information and the latter deals with the general type of content on the web, such as html page, images and even music. It's seemed that to extend GPSE is a convenient way to realize SSE. But the truth is that this mere difference is a radical difference, because of the distinct characteristics of the spatial information.

Since digital spatial information usually is very massive and also has various storing data formats and can only be stored in spatial DBMS, it's impossible to put them directly on the web and they can only be accessible on the support of the Business Log tier application. On the contrast, the quarries of GPSE are html pages, images and music, etc and they are all just on the web and can be accessed directly via HTTP, FTP or other standard network protocols.

Secondly, query for GPSE is very simple, only a key words input form is enough. But for SSE, versatile input interface is supposed to give end users good way to make a spatial related query conveniently and precisely.

Thirdly, in GPSE, the response to a query generally is satisfying links of relevant html pages, which are easy to show for the end user. On the contrary, spatial information is location-referenced and is supposed to be rendered as a map.

The following table summarized the compare between GPSE and SSE in four aspects: system architecture, difficulty of deploying, way of fulfilling and others.

Comparing Items		GPSE	SSE
1	Need data tier	Yes	Yes
	Need business logic tier	Yes	Yes
	Need presentation tier	No	Yes
2	Difficulty of collecting information	Easy	Hard
	Difficulty of manipulating collected information	Middle	Hard
	Difficulty of presentation of query results	Easy	Hard
3	Way of collecting information	Direct	Indirect
	Way of manipulating collected information	ASCII, Indexing	Digital, Spatial Indexing
	Way of presenting query results	HTML, links	Map, tables
4	To be information provider-aware	No	Yes
	Face to different information structure	No	Yes

Table 1. Difference between GWSE and SWSE

## 2. CONCEPTUAL DESIGN

### 2.1 Design Goals

In this section, we examine the design goals of Spatial Search Engine and then propose a conceptual design.

#### 2.1.1 Automation

The spatial search engine should work transparently and automatically. After getting the command of query, it can search the web or some local caches on the server rapidly and correctly, and then renders the query result as maps to the end user, with preparations for the feedback from them.

#### 2.1.2 Interoperability

There are countless web mapping systems available around the world. Spatial web search engine should be able to connect to anyone of them dynamically, no matter how they organized the spatial data or how they will react to the queries.

#### 2.1.3 Integration

Spatial search engine, as a middle ware, should be able to be integrated with other existing applications seamlessly and easily.

### 2.2 Conceptual Design

The conceptual design consists of three parts: a modularisation of spatial information providers, a standard description of them and a global registration of them, as shown in figure 3.

#### 2.2.1 Modularization

As the section 1.3 has revealed, the business logic tier is needed for map mapping system, otherwise there is no another way available for massive volume of spatial data to be published on the web. For the goal of automation described above, the business logic tier should be modularised and can be accessed via some standard interfaces exposed out to spatial search engines, without people involved.

**2.2.2 Description**

In support of the dynamic invoking spatial information providers, a standard of descriptions of what each web mapping system can offer and how to connect to them dynamically is needed. With this description, the spatial search engine can distinguish each of them and then can connect to the candidate automatically. Further more, a detailed description of technical information about connecting is also needed to enable dynamic interacting.

**2.2.3 Registration**

Since there are many web mapping systems online, going directly to one of the well-known ones is easy. But when you want to find out which web mapping system meets your needs, the ability to discover the answers can quickly become difficult. One option is to visit each of them manually, and then try to find the right one to query. Another way is through an approach that uses a description file on each web mapping system’s web site. But this approach is dependent on the ability for a crawler to locate each web mapping system and the location of the description file on that website. This distributed approach is potentially scalable but lacks a mechanism to insure consistency in description formats and for the easy tracking of changes as they occur.

One true possible way to solve this problem is to make a logically centralized, physically distributed registry centre with multiple root nodes that replicate data with each other on a regular basis. With this mechanism, all the advertised web mapping systems can be discovered dynamically and then be accessed at any time.

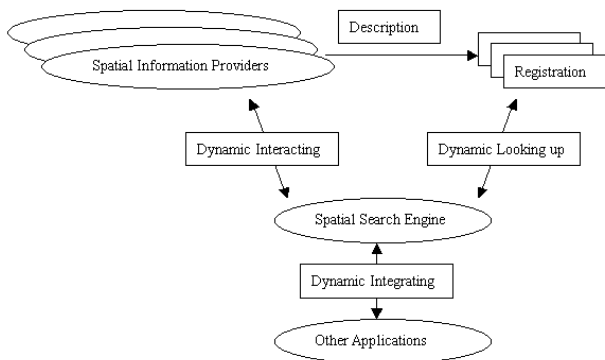


Figure 3. The Conceptual Architecture of SSE

**3. PROTOTYPE SYSTEM**

During the last year, a new technology framework called “Web Services” has emerged as a viable model for Internet based applications. As a matter of fact, Web services are a new breed of Web application. They are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. Web services perform functions, which can be anything from simple requests to complicated business processes. After a Web service is deployed, other applications (and other Web services) can discover and invoke the deployed service.

Meanwhile, the maturing technologies such as Extensible Markup Language (XML), Web Service Definition Language

(WSDL), Universal Description Discovery Integration (UDDI) and Simple Object Access Protocol (SOAP) make our conceptual design possible.

Based on these technologies, we designed and deployed a prototype system, which was introduced as follows.

**3.1 Scenario**

For the prototype system, we pose the following scenario: when one is reading news on the web, he meets a place name, for example Afghanistan, and he wants to know where it is. Can he just simply select this place name; right click the mouse and then can get the corresponding map, showed in another window?

**3.2 System Design**

The architecture of prototype system is showed in figure 4, as well as the work flow.

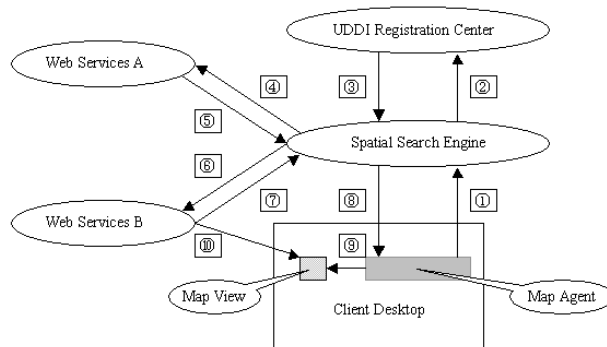


Figure 4. The System Design of Prototype of SSE

**3.2.1 Spatial Web Services**

In the prototype system, we chose GeoBeans, a Internet GIS platform, and deployed two demonstrating web services, one is responsible for Geo-coding and is showed with the name of ‘Web Service A’ in Figure 4. The other is responsible for map providing and is showed with ‘Web Service B’.

**3.2.2 UDDI Registration Centre**

For UDDI registration centre, we choose IBM WebSphere UDDI Registry Version 1.1. It is a UDDI-compliant registry for Web services in a private intranet environment. It supports the SOAP-based APIs defined by Version 2 of the UDDI specifications, and provides persistence for published entities through a relational database.

**3.2.3 Spatial Search Engine**

This engine is currently developed using Microsoft Visual Studio, and migration from Windows to Linux is now ongoing.

**3.2.4 Map Agent**

The Map Agent is a specific modular of Spatial Search Engine. It is installed on client machine, acting as an agent for end user. It can be invoked dynamically and then connected to Spatial Search Engine automatically.

### 3.2.5 Map View

The Map View is a particular window running on the client desktop. It is an Graphical User Interface, usually used to show maps to end user.

### 3.3 Working Flow

As shown in Figure 4, the working flow of the prototype system for the proposed scenario is the following:

1. User selects a place name in html page, then invoke 'Map Agent' from the right mouse button menu. Map Agent connects to Spatial Search Engine, if successful, it send a request with the selected place name and it's encoding type as parameters.
2. The Spatial Search Engine searches the UDDI registration centre to look for suitable advertised Web Services, both of its descriptive information and technical information.
3. For our scenario, Spatial Search Engine is supposed to find Web Service A (Geo-Coding), and Web Service B (Map Service).
4. According to the technical information of Web Service A, the Spatial Search Engine send a SOAP request with two parameters: the place name and encoding type.
5. The Web Service A converts the place name to latitude-longitude-based coordinates and send them back to Spatial Search Engine.
6. Spatial Search Engine then send a request for map to Web Service B with the latitude-longitude-based coordinates.
7. Web Service B analyse the request and renders an temporary Map file (In JPEG or SVG) and then send a link to them to Spatial Search Engine.
8. Spatial Search Engine forwards the link to the candidate map file to Map Agent.
9. The Map Agent invokes the Map View and forwards the link to it.
10. The Map View fetch the map file from Map Service B and then renders it to user.

## 4. CONCLUSION

Spatial Search Engine is an innovative idea about the sharing of Geographical Information. In the last several decades, many new technological innovations are allowing us to capture, store, process and display an unprecedented amount of information about our planet and a wide variety of environmental and cultural phenomena. But in spite of the great need for that information, the vast majority of that information has never fired a single neuron in a single human brain. Instead, they are stored in electronic silos of data.

In order to promote the full use of these geo-referenced information, several Web Mapping Systems have been developed to make some of these massive information accessible via the Internet. But unfortunately, there is little interrelationship among them. It is still difficult for the spatial information to be fetched.

Spatial Search Engine aims to search the World Wide Web for geographic information to meet the users' needs automatically and conveniently.

We analyzed the many differences between General Purpose Search Engine and Spatial Search Engine and argue that the existing Web Mapping Systems are the backbone of SSE.

Based on these ideas, we propose a conceptual design of Spatial Search Engine: a modularisation of spatial information providers, a standard description of them and a global registration of them. The possibility of this design is proved by a prototype system, which was implemented following the way of Web Services, and UDDI.

## ACKNOWLEDGEMENT

We thank Prof. Li Qi for useful discussions related to this work; Shunxing Dang, Jinping Li, Peng Dong, Zhenguo Qian for their help with implementation of prototype of SSE; Theresa Vallese, a foreign teacher in the Chinese Academy of Sciences, for her kind help on proofreading this paper; and the reviewers, Xingling Wang, Yuxing Wang, Liqiang Zhang, Xiaoping Rui, Huaji Zhu, Yahui Lu and Qimin Cheng for helpful comments on a draft of this article. We also thank the GeoBeans Research Group at the National Engineering Research Center for Geoinformatics of China for providing GeoBeans that we have extended. Finally, we gratefully acknowledge the financial support of the National High Technology Development 863 Project of China.

## REFERENCES

- FGDC, <http://fgdcearhs.er.usgs.gov>
- OpenGIS Specifications, <http://www.opengis.org/ogcSpecs.htm>
- Sergey Brin, Lawrence Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1-7), pp.107-117.
- Steve Lawrence, C. Lee Giles, Searching the Web: General and Scientific Information Access. *IEEE Communications*, 37(1), pp.116-122.
- Steve Lawrence, C. Lee Giles, Searching the World Wide Web, *Science*, 280(5360), pp.98-100.
- SOAP Version 1.2 Part 0: Primer. <http://www.w3.org/TR/2001/WD-soap12-part0-20011217/>
- SOAP Version 1.2 Part 1: Messaging Framework. <http://www.w3.org/2000/xp/Group/2/06/06/soap12-part1.html>
- SOAP Version 1.2 Part 2: Adjuncts <http://www.w3.org/2000/xp/Group/2/06/06/soap12-part2.html>
- UDDI 2.0 Specification, <http://uddi.org/specification.html>
- Venkat N. Gudivada, Vijay V. R Aghavan, William I. Grosky, 1997. Information Retrieval on the World Wide Web, *IEEE Journal of Computing*, pp.58-68.
- Web Services Description Language (WSDL) 1.1. <http://www.w3.org/TR/wsdl>
- Web Service Description Requirements. <http://www.w3.org/TR/2002/WD-ws-desc-reqs-20020429>