# Integrating Heterogeneous Traveler Information Using Web Services

Shanzhen Yi[1],Bo Huang[2*]

[1]Department of Civil Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 117576
E-mail: cveys@nus.edu.sg
[2]Department of Geomatics Engineering, University of Calgary, Calgary, AB, T2N 1N4
E-mail: huang@geomatics.ucalgary.ca

**Abstract**

Various types of information, e.g. weather, road and traffic conditions, can assist travelers in making better-informed decisions about their trips. The information is widely disseminated by distributed data sources and web sites. The integration of such information would provide significant value-added services to travelers, and XML-related technologies have proven to be effective to achieve this goal. This paper aims to design and implement an integrated system to make use of widely distributed traveller information by employing the XML and Simple Object Access Protocol (SOAP) techniques. The prototype system adopts a three-tier architecture and is implemented using integrated Java technologies. The shared XML schema for geo-referenced data provides a foundation for heterogeneous information integration. The XML wrappers, the metadata schema, and the visualization tools were developed to provide information services based on the heterogeneous data sources. Two examples concerning travel information query and route selection, respectively, are presented to illustrate the applicability of the system.

## I. INTRODUCTION

The availability of various types of travel related information plays an important role in travelers' route planning and decision making. Traditional approaches for route planning suffer from significant weaknesses. For example, printed brochures with static information[1] are difficult to integrate with other information. With the increasing need to obtain "real-time" travel information, a significant effort of research has been dedicated to Advanced Traveler Information Systems (ATIS). ATIS are systems with features that assist travelers with planning, perception, analysis and decision-making for route choice. Most travelers make the route choice either before starting their trips or while driving en route. Hence it is important to harness the different technologies to obtain the most updated travel-related information.

In recent years, traveler information has been dynamically disseminated through World Wide Web. Traveler information is being exchanged among public sectors, private companies and individual citizens. Information, such as weather forecast, real-time traffic, and road conditions, can be obtained from various websites. However, most of such information on the Internet is represented in local schema and format, for example, HTML pages and tables generated from connected databases by gateways or private flat files [2,3]. These heterogeneous web resources cannot be integrated effectively, making the automated exchange and integrated processing of traveler information difficult. In other words, possession of the most updated information does not guarantee a well informed travel decision and value-added service due to the difficulties in integrating heterogeneous information. It is because most of the multi-objective route selections involve various factors like weather, traffic and road conditions. The comprehensive information about these factors requires an integration of heterogeneous data sources.

In order to provide an integrated travel information and service from distributed and heterogeneous data sources, several issues need to be addressed. These issues include:

- How to extract traveler information from distributed data sources?
- How to retrieve and visualize travel information?
- How to deploy travel service model for travelers?

Such technical requirements on the integration of traveler information and services have driven the effort in exploring new and emerging web technologies. Though it is not new, eXtensible Markup Language (XML) is increasingly being used as a mechanism for exchange and integration of information among distributed data sources, such as relational, object-oriented, hierarchical, and semi-structured data. World Wide Web Consortium (WC3) provides XML scheme definition languages for data exchange and integration, including DTD (Data Type Definition) and XML Schema. XML technologies for graphic representation, such as SVG (Scalable Vector Graphics) and X3D (Extensible 3D) are effective methods for visualization. Furthermore, XML rapidly becomes the foundation of fully automated web-based transactions[4]. Web Service using XML and SOAP (Simple Object Access Protocol) provides an interoperable protocol on the Internet [5].

In this paper we will present a prototype for traveler information and service using XML and Java technologies. To resolve the different integration issues as described above, we have developed a set of methods, including shared XML schema, an XML wrapper, metadata,

* Corresponding author

1082-4006/05/1101-50$5.00

visualization methods and a web service model.

In the prototype, the shared XML schema provides a global view for exchange and integration among distributed data sources. Because the data distributed in heterogeneous systems have discord schemas, types and formats, the mapping rules should be provided between the shared schema and private schemas. In this paper, the wrapper is designed to extract information from data sources using mapping rules and then XML document data and corresponding metadata are generated. Based on the shared XML schema and metadata, the generated XML document data can be integrated and processed. The methods of visualization and traveler service model are developed based on the XML Document Object Model (DOM). The service model is deployed as Java Applets for travelers. An implementation of the prototype uses Java components. The applications of the prototype system on information retrieval and route selection are also presented to illustrate the proposed methods.

In the next section, related work on XML-based traveler information is discussed. Different methods for developing distributed systems for traveler information are introduced. Following that, the design, main components, implementation and applications of the prototype are described. Finally, the paper is concluded with a summary and suggestions for future work.

## II. RELATED WORK

The review of related work consists of two aspects: prototype systems of travel information based on XML and the methods of integrating heterogeneous data using XML.

### A. XML-based traveler information system

While there are numerous web sites that use HTML embedded with images or figures for road conditions, weather, and road maintenance, it is difficult to use HTML-based information to automate the processing of large amount of HTML coded materials because HTML does not provide the meaning of the data, but just data format[4]. By contrast, XML can be used to provide a standard approach for defining domain specific tags that convey the meaning of the data. Several proposed languages are developed based on XML to represent travel and transportation data, such as Road Web Markup Language (RWML)[6,7], LandXML for Land Development and Transportation Professionals[8], Geography Markup Language (GML)[9], and Traveler Information Markup Language(TIML)[4].

Several XML-based traveler information systems exist. A typical example is the Mitretek System prototype[10], named TripInfo, as well as its application in information-provision based on RWML[6]. TripInfo is a demonstration prototype of a route-based traveler information system, allowing a traveler to access the system using a standard web browser. TripInfo generates a route and determines the list of relevant web sites found along the route. It then contacts each site, and extracts relevant travel-related information from these sites. It organizes the information into a summarized traveler report for the specified route.

TripInfo uses XML to manage a number of internal data sets, and also provides the route's driving instructions in an XML format. With prior knowledge of road construction activities or forecasted severe weather condition along the intended route, a traveler may choose an alternate route, an alternate travel time, or at any rate be better prepared for expected delays. This service could be valuable for individual travelers, for transportation service providers like bus or trucking companies, or simply for government agencies compiling information.

Another application system was developed by using RWML [6]. In this experiment, information on roads, local regions, weather, and others were integrated and transmitted to drivers. The objective of the experiment was to propose new ways of utilizing information that can contribute to increased road safety and comfort. The experiment proved that information provision by RWML was able to increase the demand for circular touring in local regions and alleviate winter transportation problems in cities. The experiment also demonstrated that the use of XML as a data format of RWML enables the information provider to efficiently construct systems that collect data distributed over the entire Internet, compile the collected data according to user needs, and provide users with a clear report of the compiled data. This greatly increases the flexibility of the system in adapting to specification upgrades. In light of these encouraging results, they proposed the development of a comprehensive mobility-support service that incorporates the web service concept.

These systems demonstrate the methods of using XML to develop travel information web sites. However, travel information could be heterogeneous and distributed over the Internet. It may include weather information, road information and more. At present, users of these systems must call up each web page individually, making the decision process even more tedious. Even value-added information service providers (ISPs), while providing a portal of multiple information sources, do not fully integrate the information. This is partially due to the difficulty in integrating information with different formats from distributed sources, including databases and flat files, automatically. While most travel related information is based on geo-referenced spatial data, integration and visualization of these data are not provided in these prototypes.

### B. Methods for integrating heterogeneous information based on XML

The methods of wrapping heterogeneous databases based on XML are well discussed in the literature[3,11]. An ideal method is to follow the same standard or schema. For the integration of travel related information from distributed sources and multiple formats of geo-referenced data, the most significant contribution is the effort by the Open GIS Consortium (OGC). OGC has published its abstract specifications (Version 4) and implementation specifications for OLE/COM, CORBA and SQL[12]. The practical direction is to develop a set of APIs for each data format. Geo-referenced information can then be integrated by calling APIs locally or in a distributed environment.

Metadata[13] also play an important role in traveler information applications, such as travel data retrieval and management. OGC and other standardization organizations at the national level (such as the Federal Geographical Data Committee (FGDC) or international level (such as ISO /TC211) have contributed significantly to the research of metadata[14,15].

Among the newly emerging Internet-based technologies, "web service" is a promising solution to the interoperability of traveler information. It can be used to exchange data and execute functional interactions among incompatible systems via the Internet. More importantly, its application can significantly reduce the cost of system integration and migration. Yang and Wei[5] provided an analysis on how this technology would be beneficial to the transportation-related communication solutions and integration.

Kajiya and Yamagiwa[6] discussed in depth the concept of "Smartway web service XML". In their discussion, standardization of services on the web-based system and the combination of multiple arbitrary services are leveraged to offer new value-added tourism services. Since web service-based traveler information systems allow freedom in combining different services to satisfy the user's needs, they can be flexibly customized according to users' current locations and preferences, as well as the time of the day. Various ideas and local characteristics can be incorporated into the construction of a flexible system.

Although various XML technologies have been discussed for traveler information integration and possible on-line services, the techniques of wrapping and extracting information from heterogeneous data sources using XML remain obscure. Furthermore, the methods on visualizing and providing value-added services based on the integrated data repository still need to be addressed.

## III. SYSTEM DESIGN

In order to integrate travel related information, a prototype system has been designed. The system takes into account the following aspects:

- Characteristics of distributed and heterogeneous geo-data sources
- Possible integration methods using innovative XML technologies
- Visualization

The adopted three-tier architecture includes distributed data sources, the application server and user client (Figure 1). The travel related distributed data sources have the following features:

- Travel related information may be distributed among different organizations and repositories. For example, the weather information may reside with the meteorological agency and the traffic conditions with the transportation management agency.
- The distributed data sources present information in heterogeneous syntax, structure and semantics. The syntactical heterogeneity indicates the differences of data formats (e.g. text, HTML, binary file and XML) and data definitions. The structural heterogeneity indicates differences of schema, e.g. XML schema, relation database schema or object database schema. The semantic heterogeneity indicates the differences of meaning or interpretation for similar geographic entities.

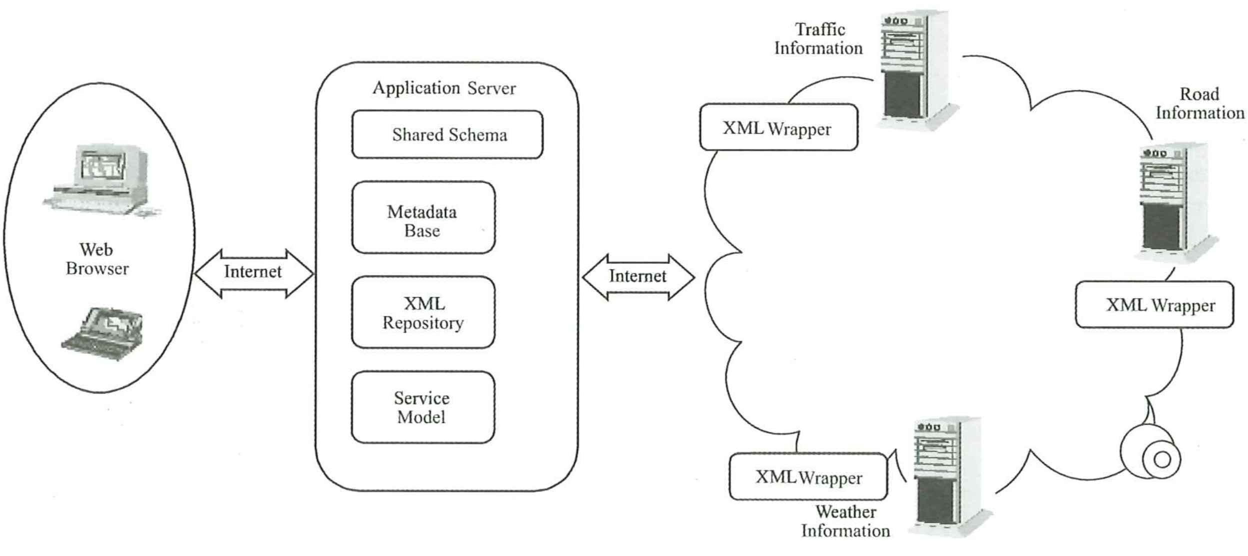


**Figure 1.** Prototype architecture

In reference to the characteristics of distributed and heterogeneous travel related information, the prototype encompasses:

- **Shared schema**. Resource Describe Framework (RDF) and XML Schema Definition Language provide the definitions of vocabularies and structures used in the XML documents. The shared XML schema defined by the XML Schema Definition Language provides an integrated and global view. The roles of a shared XML schema in information integration include the building of mapping rules between the shared schema and the distributed data source schema, and the parsing of XML documents for user query services.
- **XML wrapper and mediation**. The XML wrapper is an extractor which extracts information from data sources and stores the extracted information as XML documents with the shared schema. In the process of wrapping, the differences in data structure and semantics are mediated according to the mapping rules between data sources and the shared schema.
- **Metadata**. Metadata are "data about data." They describe the content, quality, condition, and other characteristics of a data set. In our prototype, metadata include the contents and descriptions of the XML document data extracted from the distributed data sources.
- **Integrated spatial geometric data**. Most travel related information is geographically distributed, e.g. the road network and weather outlook of a given area are provided by different institutions. Geo-referenced data from different sources may have different coordinate systems and projections. Before the integration of all these data, they need to be unified in the process of wrapping.
- **Visualization**. Visualization allows users to appreciate the data in graphical forms.
- **Web service model**. The service model of traveler information, e.g. the route choice service model, is of significance to the development of a traveler information system. The web service model of W3C uses SOAP as a light-weight protocol for exchanging data with XML documents. The combination of XML and SOAP provides an opportunity to achieve cross-platform and cross-system interoperability. In our case, the Java applet and classes are used to deploy traveler service model.

Practically, the shared schema determines the extraction and mapping rules in the process of wrapping distributed heterogeneous data sources. For geo-referenced information, the GML specification[7] provides a shared XML schema using the XML Schema Definition Language. In our prototype, the distributed geo-referenced data sources, such as HTML pages, relational databases (e.g. Oracle Spatial) and flat files (e.g. ESRI Shapefile) are wrapped into XML documents based upon the extraction and mapping rules between the shared schema and the schema of data sources. The procedure of information extraction is shown in Figure 2.

Extraction and mapping rules between the shared XML schema and heterogeneous data sources may be built manually or automatically. Automatic methods can involve schema matching using statistics or machine learning approaches. Based on the mapping rules, the wrapper extracts data from these heterogeneous data sources and generates XML document data. In our case, the mapping rules between the shared XML schema and the Shapefile data sources are

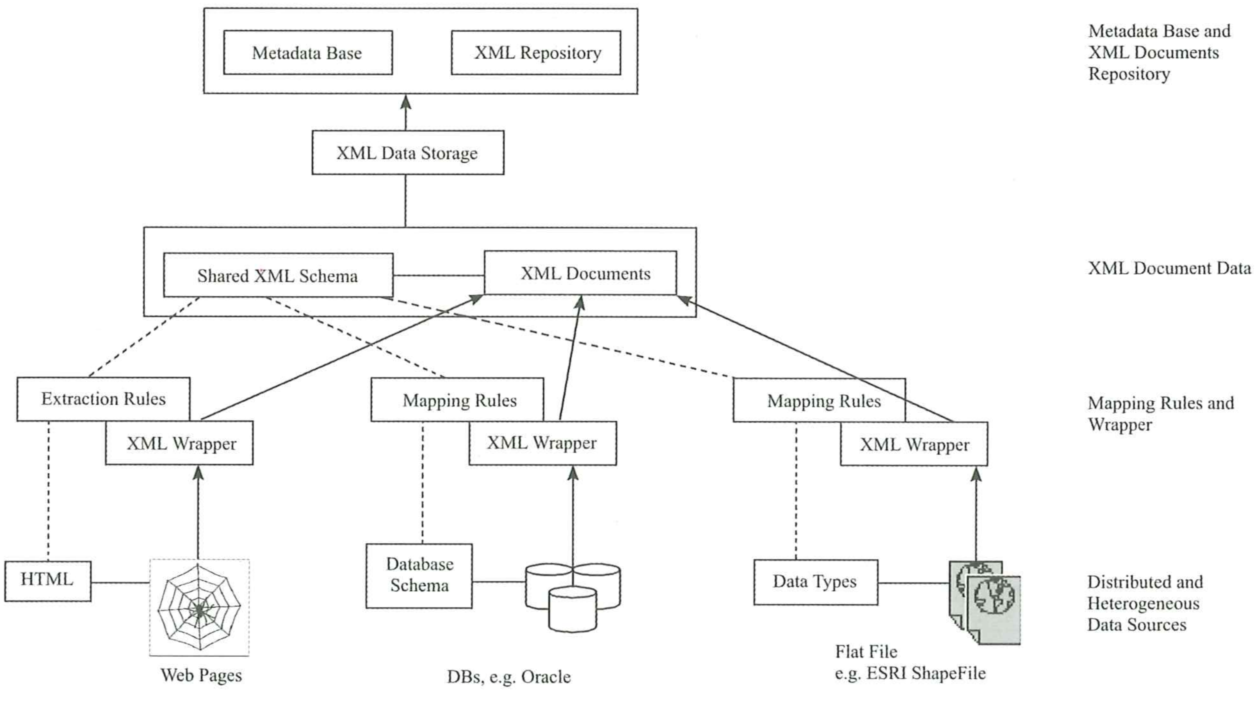


**Figure 2.** Data integration from heterogeneous data sources

generated manually, so does the mapping between the geometric and attribute data types.

There are many metadata standards, e.g. the Doubling Core Metadata standard for cross-domain information resource description[15], and the FGDC Content Standard for Digital Geospatial Metadata[14]. The FGDC metadata standard provides a common set of terminologies and definitions for documenting digital geospatial data. In our prototype, metadata content for an XML document data describe the content of the XML documents stored in XML repository, such as categories and attributes of elements.

XML document data and their metadata generated from data sources are stored in the XML repository and the metadata base, respectively. The user query service is based upon the XML documents and metadata.

## IV. SYSTEM IMPLEMENTATION

### A. System components

Java technologies are used to implement the prototype. These technologies include Java servlet, Java applet, Java server page (JSP), Xerces Java parser, JDBC-ODBC and Microsoft Access database. The system architecture is shown in Figure 3.

Spatial data, traffic and other data are extracted from data sources in ESRI Shapefiles format and other web pages. Shapefile is an open geospatial data format promoted by ESRI [16]. A set of Shapefiles stores non-topological geometry and attribute information for the spatial features.

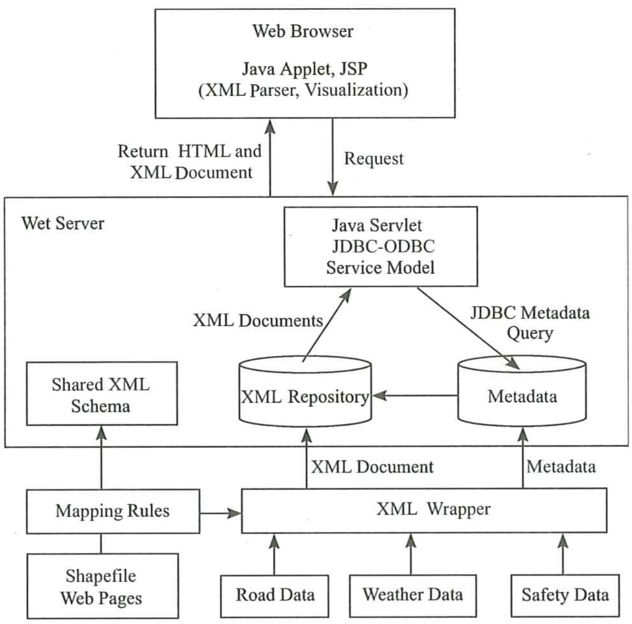


**Figure 3.** The components of the prototype system

The mapping rules from the Shapefiles spatial data types to the shared XML schema have been constructed. The XML wrapper reads shape files, extracts the spatial features and attributes data inside the shapefiles, transforms the data according to the mapping rules, and writes the transformed data into an XML document based on the shared XML schema. The XML document and its metadata are uploaded to the XML repository and the metadata database by the wrapper. In our prototype system, the XML repository is a file system which includes an XML data directory. The metadata are stored in a Microsoft Access table.

The web user interface is written in HTML embedded with Java applets and JSP for visualization and interaction with the web server. The web server communicates with a user through a Java servlet. After the user submits a request, the web server processes the request and uses JDBC-ODBC to access the metadata and manipulate the corresponding XML documents. The service model is deployed using Java class. The results are represented in HTML pages embedded with JSP scripts and Java applets, which are delivered to the web user. The Java applets allow users to interact with and visualize the parameter and result data. In our prototype, we used Apache Tomcat as the web server, which supports Java servlet, JSP, JDBC, and provides the Apache Xerces Java Parser package for processing XML documents. Two types of data sources, i. e. traffic data and weather data, are tested. The data formats include flat files (Shapefiles), access table files and HTML web pages.

### B. Shared XML schema and metadata

The data from heterogeneous data sources are transformed into XML documents with a shared XML schema. The shared XML schema in our prototype is based on a subset of the OGC's GML. GML is an XML encoding for transport and storage of geographic information, including both the spatial and aspatial properties of geographic features. The geometry schema and feature schema in GML define the geospatial types and elements. GML offers a number of advantages, including the ability to separate content and presentation. Furthermore, GML can display maps on a web browser. It is extensible, and can easily be linked with any other types of aspatial XML data.

A feature in GML is essentially a list of properties. Among the properties, some may be geospatial describing the position and shape of the feature, and some others may be of common numerical types. Features without geometry information can also be encoded in GML. This is suitable for traveler information, as there is no geometry, for instance, for the travelers (or consumers). Encoding a non-geometric feature in GML can simply be done by adding types *gml:_Feature* and *gml:AbstractFeatureType*. For example, the *Consumer* feature can be encoded as:

```
<Cosumer>
    <familyName>Smith</familyName>
    <age>42</age>
    <nickName>Smithy</nickName>
</Cosumer>
```

The element is declared as a GML feature:

```
<element name="Cosumer" type="ex:CosumerType"
substitutionGroup="gml:_Feature" />
<complexType name="CosumerType">
    <complexContent>
        <extension base="gml:AbstractFeatureType">
        <sequence>
            <element name="familyName" type="string"/>
            <element name="age" type="integer"/>
            <element name="nickName" type="string"/>
        </sequence>
        </extension>
    </complexContent>
</complexType>
```

A data set includes a collection of records or tuples. In XML, such a collection corresponds to the feature collection. The definition of a feature element and a feature collection element in the shared XML schema are:

```
<element name="FeatureCollection">
 <sequence>
  <element name="Feature" type="FeatureType" minOccurs="0",
maxOccurs="unbounded"/>
 </sequence>
</element>

<complextype name="FeatureType" >
 <sequence>
  <element name="ID" type=="Integer"/>
  <element name=" Geometry" >
    <xs:choice>
        <xs:element ref="Point" minOccurs="0",
maxOccurs="unbounded">
        <xs:element ref="LineString" minOccurs="0",
maxOccurs="unbounded">
        <xs:element ref="LinearRing" minOccurs="0",
maxOccurs="unbounded">
        <xs:element ref="Polygon" minOccurs="0",
maxOccurs="unbounded">
        <xs:element ref="MultiPoint" minOccurs="0",
maxOccurs="unbounded">
        <xs:element ref="MultiLineString" minOccurs="0",
maxOccurs="unbounded">
        <xs:element ref="MultiPolygon" minOccurs="0",
maxOccurs="unbounded">
    </xs:choice>
  </element>
     <element name="Attribute" type="attributeType"
minOccurs="0", maxOccurs="unbounded">
 </sequence>
 </complextype>

 <simpleType name="attributeType">
    <xs:choice>
      <element name="atype" type="xs:integer"/>
      <element name="atype" type="xs:float"/>
      <element name="atype" type="xs:string"/>
      <element name="atype" type="xs:datetime"/>
      ……………..
    </xs:choice>
  </simpleType>
```

A feature with geometric properties must be declared with the element *Geometry*. The types of geometry include primitive geometric elements and aggregate geometric elements. The geometric types defined in the shared XML schema are:

```
<!— Primitive geometry elements — >
    <element name="Point" type="gml:PointType"/>
    <element name="LineString" type="gml:LineStringType"/>
    <element name="LinearRing" type="gml: LinearRingType"/>
    <element name="Polygon" type="gml: PolygonType"/>
    <element name="Box" type="gml:BoxType"/>
<!— Aggregate geometry elements — >
    <element name="MultiPoint" type="gml:MultiPointType "/>
    <element name = "MultiLineString" type = "gml
MultiLineStringType"/>
    <element name = "MultiPolygon" type = "gml:
MultiPolygonType"/>
<!— Aggregate geometry elements — >
    <element name="coord" type="gml:CoordType"/>
    <element name="coordinates" type="gml:CoordnatesType"/>
```

The metadata for the XML document are stored in the relational Microsoft Access database. The schema for metadata involves the index (ID) of an entry, domain of the document content, feature category or categories contained in the document, spatial range (described by maximum and minimum coordinates), creator, file name of the XML document, and the attributes of the feature class. Metadata provide a mechanism for information retrieval based on properties.

### C. Mapping and wrapping of data from heterogeneous sources

Information extraction is performed in two steps. The first step is the schema mapping between the shared XML schema and the schema of the data sources, and the second step is wrapping data from the sources with XML documents. Three data formats, ESRI Shapefiles, MS Access table and webpage, are used to illustrate this process.

Based on the shared XML schema and the Shapefile structure, the schema mapping rules are placed between the shared schema and the heterogeneous data sources. The data types and the file format of Shapefile can be found in the ESRI Shapefile Technical Description document[16]. There are two data types: geometric types and attribute types. In the type mapping from the Shapefile to the Shared XML schema, the attribute type mappings take on the form of one-to-one, e.g. *String* to *String*, but the geometric type mappings take on the form of many-to-one, e.g. Shapefile's *pointM* and *pointZ* to

*<Point>* in the shared XML schema.

While the Shapefile geometric data are stored in binary flat files, the attribute data are stored in database files. The geometric and attribute data of a road network are shown in Figure 4. The travel related information can be extracted from this dataset, including the location and length of a road, and average speed on a road segment. The Shapefile XML wrapper opens the geometric and attributes data files, extracts useful data and transforms them and places them into an XML document according to the mapping rules.

The weather data, such as temperature, cloud, and humidity in a given area or a city, may be extracted from a webpage published by a public agency. The weather data as attributes of a polygonal area are encoded by XML as shown in Figure 5.

The XML documents and the corresponding metadata generated by the XML Wrapper are stored in the XML repository and the metadata-base. The management and operations of the metadata and XML documents are undertaken by the web application server.

**D. XML parser and client side visualization**

XML data processing is based on the XML parser deployed on both the server and client sides. This parser is a programming library used to generate and parse XML documents. The Apache Xerces Java parser[17] supports W3C XML Schema Recommendation 1.0, SAX1.0 and 2.0, DOM level 1 and level 2[18,19]. This level of support makes it possible for an application to read and write XML data.

The procedure of XML data processing and visualization is shown in Figure 6. Each XML data document is in compliance with the shared XML schema. When the XML parser processes XML documents, the structure and vocabulary defined in the shared XML schema are used for services, such as retrieval of elements and attributes. The XML parser uses the elements, attributes and hierarchical structure of XML documents to generate the DOM tree, upon which XML data operations (i.e. selection, insertion or update) become available, and the geometric and attribute data can be extracted for further processing, including analysis and presentation.

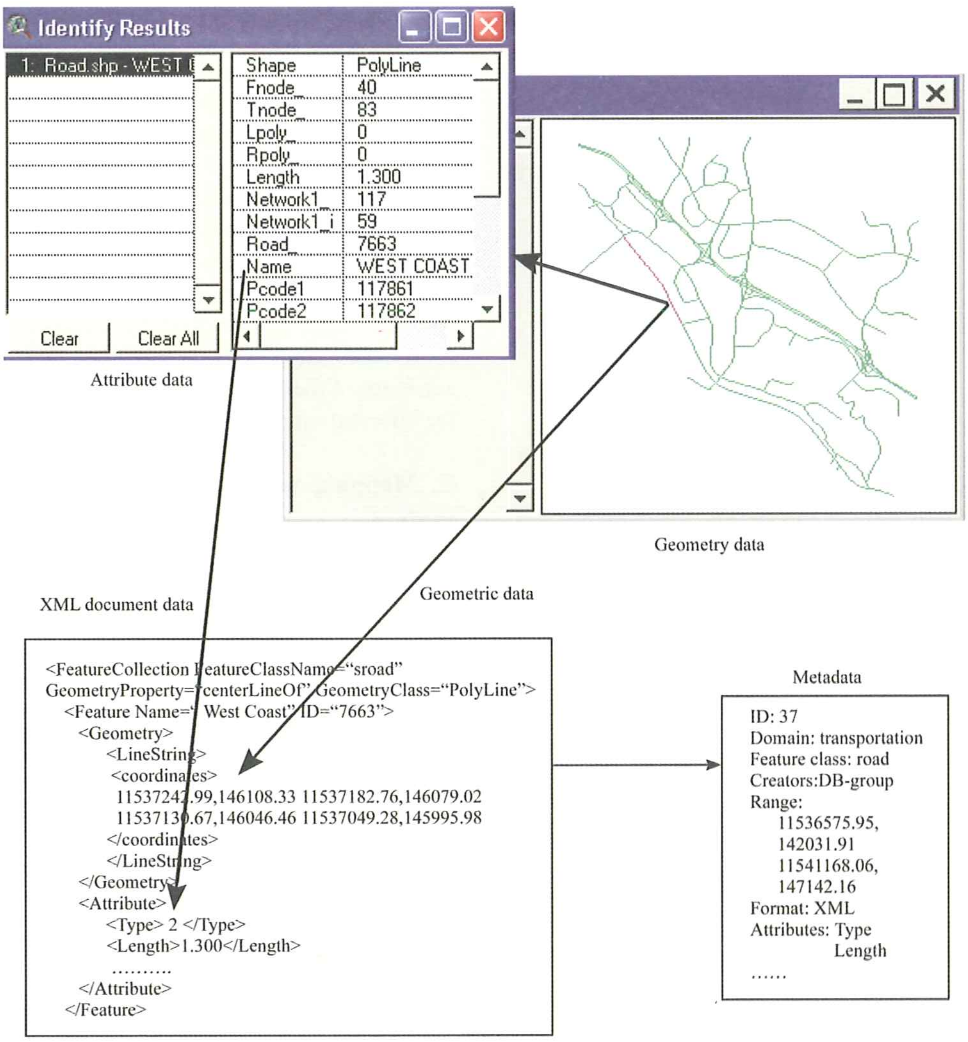


**Figure 4.** Data extraction by XML wrapper from shapefile to XML document

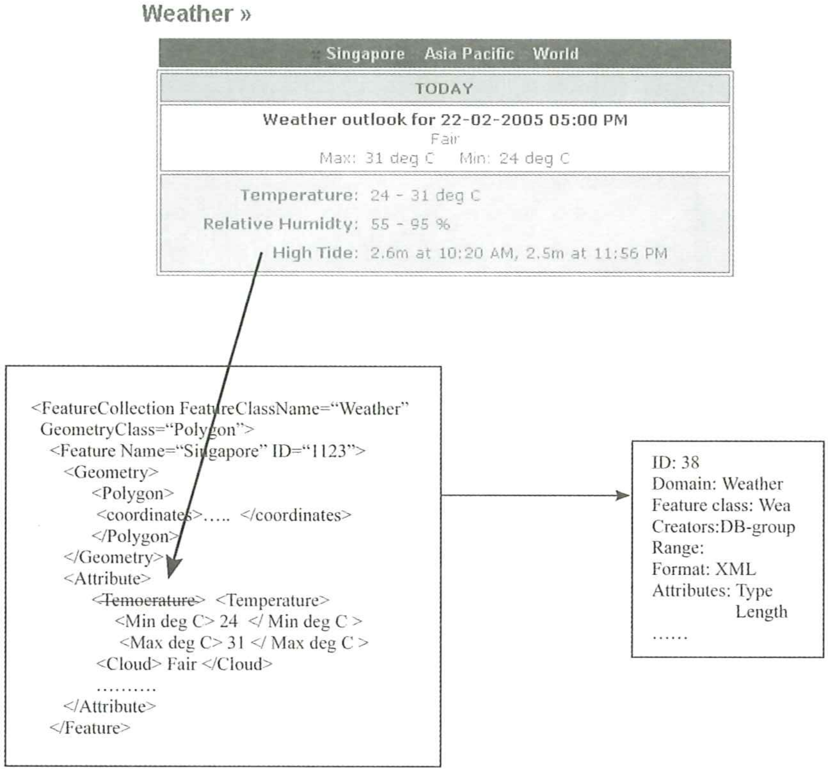


**Figure 5.** Data extraction from a webpage into an XML document

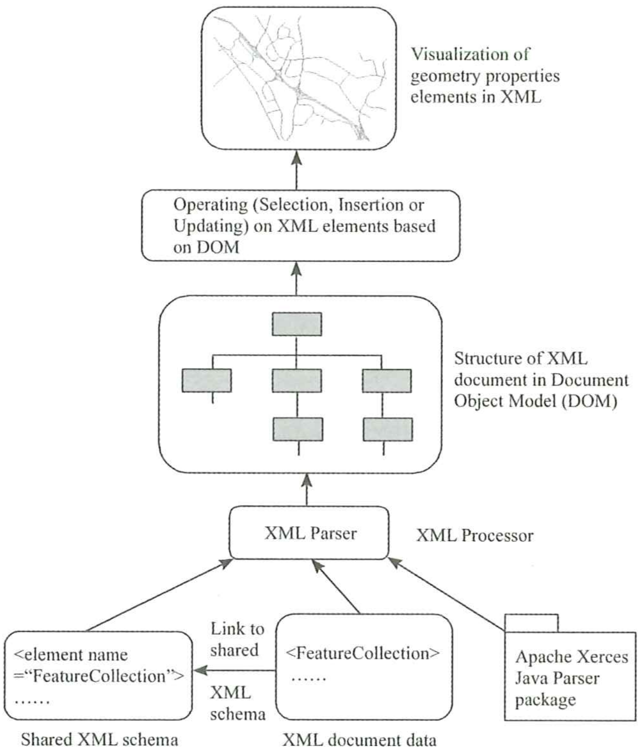


**Figure 6.** The procedure of XML data processing and visualization

## V. APPLICATIONS

The integrated XML document data and metadata provide a foundation to launch travel related applications. Two applications will be illustrated: one is concerned with information query and visualization and the other with route selection.

### A. Traveler information query and visualization

Metadata include items describing the properties of XML document data generated from heterogeneous data sources. Based on the metadata item, user information can be retrieved. The results of a retrieval are returned as an XML data set and can be visualized in the web browser. In our example, the primary metadata items include category (represented as domain), feature class, spatial range of feature collection, and attributes.

Results of metadata retrieval and XML data visualization for road networks are shown in Figure 7. On the left side of Figure 7 is the interface for retrieval of XML dataset based on metadata. When a user submits the query for metadata, the suitable XML datasets are then listed in the left of the front window. When the user picks a dataset item, for example, *ID37-road*, the geometry properties of the selected dataset are then displayed in the middle window, and the metadata of the

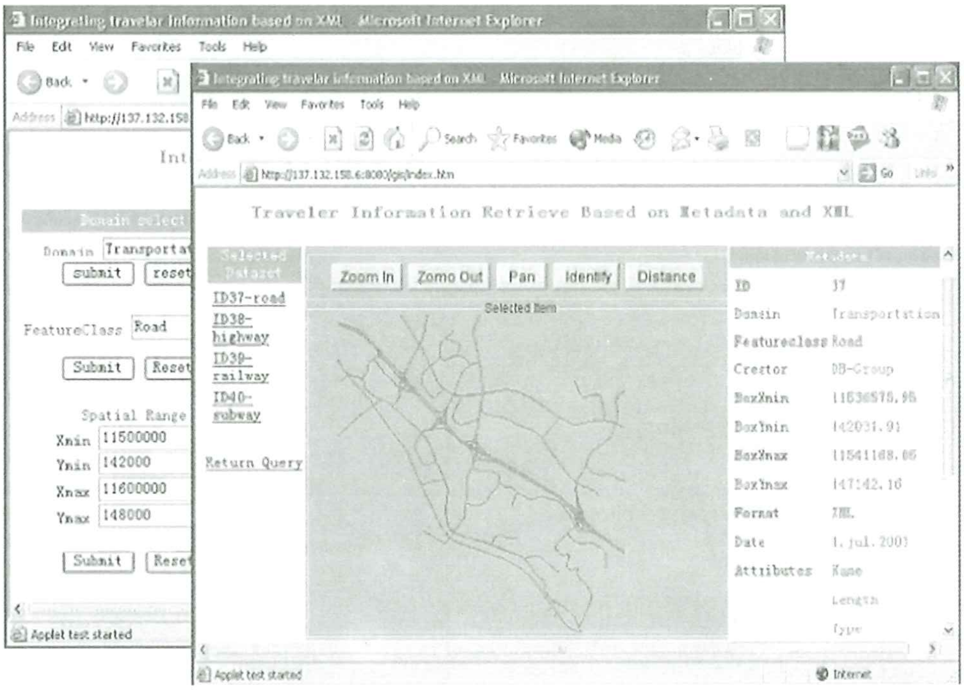


**Figure 7.** Information retrieval and visualization of XML document data

dataset are displayed in the right window. The visual e-map allows a user to perform simple operations, such as zoom in and identify a feature object.

### B. Route selection service

Transportation users make route choices using their preferential factors, such as travel time, cost, scenic quality, comfort and weather. The service models based on these factors are seldom deployed in any web site.

One of the benefits of the XML data repository constructed from the distributed heterogeneous data sources is that the development and deployment of service models and decision support models can be easily done. It is because the shared XML schema specifies all XML documents with a consistent schema and structure, avoiding the conflicts of schema and semantic heterogeneity.

In our prototype, the service model is located at the local application server. It can easily be extended similar to that the service models are distributed on the Internet using the Web Service technology, because the SOAP technology utilized is also based on XML. Hence the XML documents in the XML repository can satisfy the transmission requirements of transportation parameters in SOAP without significant changes.

A service model of route selection for travelers involves multiple factors. The data capturing these factors are extracted from distributed data sources and stored in XML documents. The route selection model based on the generalized cost in our prototype considers 3 factors, which is represented as:

$$C(i,j) = \sum_{k=1}^{n} \alpha_k F_k(i,j) \tag{1}$$

where $n$ equals 3 in our example. The three factors are travel time, weather condition and safety, represented by functions $F_1$, $F_2$ and $F_3$, respectively. $F_k(i,j)$ denotes the function value of the $k$-th factor between points $i$ and $j$, such as the length of travel time, weather condition (i.e. sunny, cloudy, or rain) and safety. The weights $\alpha_k(k=1, 2, 3)$ attached to the factors show the degrees of emphasis on the factors by a user. The weights satisfy the constraint in equation (2):

$$\sum_{k=1,\cdots,n} \alpha_k = 1 \tag{2}$$

The generalized cost $C(i,j)$ is the cost of traveling on the path between points $i$ and $j$. The weights are specified by a user. In our prototype, the weight of each factor carries a value representing the level of significance: insignificant, significant, very significant (Table 1).

**Table 1.** Weight levels of factors

| Factors ($k$) | Weight levels ($l$) | | |
|---|---|---|---|
| | Insignificant | Significant | Very significant |
| 1 | $S_{11}$ | $S_{12}$ | $S_{13}$ |
| 2 | $S_{21}$ | $S_{22}$ | $S_{23}$ |
| 3 | $S_{31}$ | $S_{32}$ | $S_{33}$ |

We define the value of each weight level $S_{kl}$ in the following manner: 0.1 for "insignificant", 0.5 for "significant", 1.0 for "very significant". For each factor, only one weight level is selected using the single selection. For example, when user selects the first factor of travel time with "significant" (second weight level) in the route decision, then $S_{12}$ is selected and has the value of 0.5. Other weight levels of this factor (e.g. $S_{11}$ and $S_{13}$) have values of 0.0.

The weight calculation for each factor follows equation[3].

$$\alpha_k = \frac{\sum_{l=1}^{3} S_{kl}}{\sum_{k=1}^{n}\sum_{l=1}^{3} S_{kl}} \quad (3)$$

When the background dataset, e.g. road network XML document, is selected, the dataset is then presented graphically on the user interface (Figure 8). The factors for route selection and weight levels described in Table 1 are shown on the left part of the window. A user can select factors and weight levels for route selection decision according to his/her preferences. He/she then needs to determine the origin and destination of the trip by clicking directly on the map. When the parameters are submitted to the web server, the service model will calculate the generalized travel cost on the paths between the origin and the destination, and display the route with the minimum cost.

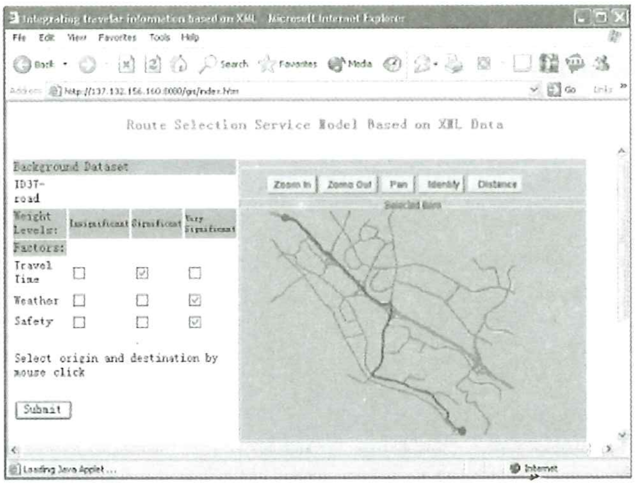


**Figure 8.** Route selection service

The XML documents of the three factors are integrated and processed in the route selection program. The road data include nodes, geometric lines of roads and attributes of road segments, such as name, length and average speed for a vehicle. Weather data contain weather conditions, such as rainy, cloudy, or sunny in the given polygons. The service model calculates the values of the factors on each road segment using the intersection method and then derives the minimum cost path.

## VI. CONCLUSIONS AND FUTURE WORK

Most of the travel related information is distributed in the form of heterogeneous data sources. The integration and retrieval of such information cannot be performed conveniently with traditional methods for database access. This paper has presented an integrated prototype system for retrieval and access of heterogeneous travel information and services based on the XML technology and wrapping methods.

The prototype system adopts an open scalable architecture to bring together geo-referenced information using a shared XML schema and wrapping and extracting methods. The XML wrapper transforms the complex geo-referenced data from flat files into XML document data, and builds the metadata of the XML documents in the XML repository. Furthermore, the parser for the geo-referenced XML data and client-side visualization of the spatial objects are developed. Lastly, a service model for travel route selection that involves information from multiple data types is also devised.

The prototype can be extended along several directions. For example, the wrapping and extraction of geo-inference data from relational database and HTML pages can be done automatically to build XML repositories. A diversity of XML-based service models for trip planning may also be investigated to satisfy various user requirements, especially with the increasing use of XML documents. A further challenge may include semantic interoperability between heterogeneous data sources. An automatic semantic matching method [20] is critical to accomplish this future task.

## REFERENCES

[1] Smith B. L., 2000, Using Geographic Information Systems and the World Wide Web For Interactive Transit Trip Itinerary Planning. *Journal of Public Transportation,* 3(2).

[2] Zhang J. T., Javed M. S., Shaheen A., Gruenwald L., 2000, A Prototype for Wrapping and Visualizing Geo-Referenced Data in Distributed Environments Using the XML Technology. *ACM GIS/CIKM*, pp.27–32.

[3] Zhang J. T., Gruenwald. L., 2001, A GML-based open architecture for building a geographical information search engine over the Internet. In *Proceedings of the 2nd International Conference on Web Information Systems Engineering*, December,Kyoto, Japan. pp.25–32.

[4] McGurrin M. F., Roberts D. L., Glassco R. A., 2001, TripInfo: Integrating Traveler Information Using XML. *ITS 2001*, June 2001, http://www.mitretek.org/its/TripInfo.pdf.

[5] Yang Q. Y., Wei H. G., Manring J. B., 2003, XML, SOAP and Web Service: A new option to ITS C2C communication and interoperability. *Transportation Research Board, 82nd Annual Meeting*, January, Washington, D.C.

[6] Kajiya Y., Yamagiwa Y., Hamada S., Shimano T., 2002, ITS BASED ON XML AND WEB SERVICE TECHNOLOGIES– The future of ITS in the ubiquitous network society.

*ITSWC2002*, Chicago, Illinois October 14-18. http://rwml.its-win.gr.jp/papers-pdf/RWML-ITSWC2002Chicago.pdf.

[7] RWML, 2001, Working Group. *Road Web Markup Language (RWML) Specification Ver.0.80*, http://rwml.its-win.gr.jp/eng/index.htm.

[8] LandXML., 2002, org Industry Consortium, *LandXML Schema Specification*, June, http://www.landxml.org/spec.htm.

[9] Cox S., Daisey P., Lake R., Portele C., Whiteside A.(Eds.), 2003, *Geography Markup Language(GML) Implementation Specification. OpenGIS Implementation Specification, version 3.0*, http://www.opengis.org.

[10] Mitretek Systems, 2000, *Development of the XML-based TripInfo Prototype*, http://www.mitretek.org/its/TripInfo/intro.html.

[11] Baru C., Gupta A., Ludäscher B., Marciano R., 1999, Papakonstantinou Y., Velikhov P., Chu V., XML-Based Information Mediation with MIX, exhibition program, *ACM Conf. on Management of Data* (SIGMOD'99), Philadelphia, USA.

[12] Open GIS Consortium, Inc (OGC), *Implementation Specifications*, 2002, http://www.opengis.org/techno/implementation.htm.

[13] Gobel S., Lutze K., 1998, Development of meta database for geospatial data in the WWW, *ACM GIS'98*, pp.94–99.

[14] FGDC-STD-001-1998, *Content Standard for Digital Geospatial Metadata*, http://www.fgdc.gov/metadata/contstan.html.

[15] ISO TC 46/SC 4 N515, *Information and documentation—The Dublin Core metadata element set*, 2003, http://www.niso.org/international/SC4/sc4docs.html.

[16] ESRI (Environmental Systems Research Institute, Inc.). *ESRI Shapefile Technical Description,* 1998, www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

[17] The Apache XML Project, *Xerces Java Parser 1.4.4*, 2002, http://xml.apache.org /xerces2-j/index.html.

[18] W3C Recommendation, *Document Object Model (DOM) Level 1 Specification,Version 1.0,* October, 1998, http://www.w3.org/TR/REC-DOM-Level-1/.

[19] W3C Recommendation, Document Object Model (DOM) Level 2 Core Specification, Version 1.0, *November*, 2000, http://www.w3.org/TR/DOM-Level-2-Core/.

[20] Yi S. Z., Huang B., Chan W T., 2005, XML application schema matching using similarity measure and relaxation labelling. *Information Science: an International Journal*, 169(1-2):27–46.