A GML-Based Open Architecture for Building
A Geographical Information Search Engine
Over the Internet

Jianting Zhang
Le Gruenwald
School of Computer Science
The University of Oklahoma
U.S.A.
ggruenwald@ou.edu
Introduction

Objectives
Existing GRI Searching Engines: problem and propose
Geographical Markup Language: a Simple Illustration

Proposed Architecture and A Prototype Implementation

System Architecture
Motivation of the Prototype
Software Agents for Collecting GRI over the Internet
GML Wrappers for RDBMS and ORDBMS
Spatial-Enabled Mediator
Client Query and Visualization

Conclusion and Future work direction
Research Objectives

- Propose an open architecture for building a geographical information search engine over the Internet
- Develop a proof-of-concept prototype using University of Oklahoma’s student contact information list
Related Work: Map Servers
Related Work: Google Search

Search the web for 200 Felgar Street, Norman, OK 73019.

Category: Science > Environment > Environmental Impact Assessment > Cumulative Impacts

Show map of 200 Felgar Street Norman OK 73019 on Yahoo! Maps - MapBlitz

Purchasing Office Forms Gallery

... Mail the completed form, with signature, to: Kathy Gilley, Procurement Services. 200 Felgar Street, Engineering Lab, Room 201. Norman, OK 73019 ADOBE. ...
www.ou.edu/purchasing/gallery.htm - 13k - Cached - Similar pages

University of Oklahoma Department of Procurement Services RFP ...

... Department of Procurement Services. 200 Felgar Street. Engineering Lab, Room 201. Norman, OK. 73019. Phone: (405) 325-2811. Fax: (405) 325-5068. ...
www.ou.edu/purchasing/rfp/ - 15k - Cached - Similar pages
[ More results from www.ou.edu ]

PDF rk.gsfc.nasa.gov/richcontent/MAPLDCon00/Abstracts/Muehring_A.PDF

... and John K. Antonio School of Computer Science University of Oklahoma 200 Felgar Street Norman, OK 73019-6151 Phone: 405-325-7859 Fax: 405-325-4044 antonio@ou ... Text version - Similar pages


... Sudarshan K. Dhall School of Computer Science University of Oklahoma 200 Felgar Street Norman, OK 73019 Phone: (405) 325-7859 antonio@ou.edu Abstract. The work ... Text version - Similar pages
Related Work: Google Search

Searched the web for 200 Felgar St., Norman, OK 73019

Show map of 200 Felgar St Norman OK 73019
on Yahoo! Maps - MapBlast

INFORMS/CORS Montreal 1998 Cluster: Metaheuristics
... Michel Toulouse --- Univ. of OK, Sch. of Comp. Sci., 200 Felgar St., Rm. 114, Norman, OK 73019, (toulouse@cs.ou.edu) We provide a new metaheuristic for graph ...
www.informs.org/Conf/Montreal98/TALKS/C8.html - 29k - Cached - Similar pages

Improving Decision Trees Using Tabu Search
... Teodor Gabriel Crainic --- Univ. of OK, Sch. of Comp. Sci., 200 Felgar St., Rm. 114, Norman, OK 73019, (toulouse@cs.ou.edu) Self-organization is a spontaneous ...
www.informs.org/Conf/Montreal98/TALKS/TB18.html - 7k - Cached - Similar pages
[ More results from www.informs.org ]

... 114 Norman, OK, 73019 lbrown@cs.ou.edu Tel: (405) 946-7381 Fax: (405) 325-4044 Le Gruenwald The University of Oklahoma School of Computer Science 200 Felgar St ...
Text version - Similar pages

Academics: Graduate Program: Doctorate Program
... addressed to Graduate Liaison, School of Computer Science, 200 Felgar St, Room 116, Norman, OK 73019-6151. Inquiries concerning graduate applications should be ...
www.cs.ou.edu/academics/grad_program/phdprogram.shtml - 11k - Cached - Similar pages
[ More results from www.cs.ou.edu ]
Related Work: Alexandria Digital Library Gazetteer at UCSB

ADL Gazetteer Help Information (07/03/2001)

This version of the ADL Gazetteer is based on the ADL Gazetteer Content Standard. Currently it contains data from the NIMA Gazetteer, a set of counties and U.S. cities, set of U.S. topographic map quadrangle footprints, set of volcanoes, and set of earthquake epicenters. The CRIS data from the U.S. Geological Survey has been partly added to the collection. As of today, there are approximately 4.4 million entries. For the full description, see the ADL Gazetteer Collection metadata.

- This interface has been developed and tested using Internet Explorer 5. The interface can also be used with Netscape Communicator 4.6.1 or higher. We noticed that one minor feature for search result display is not supported by the current Netscape Communicator 4.6.1 release. For complete functions, please use the Microsoft Internet Explorer.
- If you have any problems using this interface, please send email. To help us to identify the problem, we would like to have the description of the problem, name and version of your Web browser, machine and operating system.

Last modification: February 21, 2000
Problems with Current Solutions

• Full text search does not work well for Geo-Referenced Information (GRI)

• Proprietary data and software introduce integration and interoperability problems

• Centralized GRI data management does not match its distributed nature and application requirements
Proposed Solution

GRI Transformation

-- Geocoding

Interoperability:

-- GML

Distributed and Heterogeneous Data Management:

-- Wrapper/Mediator
What is Geographical Markup Language (GML)?

- GML is an extension to XML
- GML defines XML representations of geographical features, such as Point, Polyline and Polygon.
- GML is standardized by OGC (Open GIS Consortium)
- Current version is 2.0 released on Feb. 20th, 2001
A Simple GML Document

<Feature fid="142" featureType="school" Description="A middle school">
  <Polygon name="extent" srsName="epsg:27354">
    <LineString name="extent" srsName="epsg:27354">
      <CData>
        491888.999999459,5458045.99963358
        491904.999999458,5458044.99963358
        491908.999999462,5458064.99963358
        491924.999999461,5458064.99963358
        491925.999999462,5458079.99963359
        491977.999999466,5458120.9996336
        491953.999999466,5458017.99963357
      </CData>
    </LineString>
  </Polygon>
</Feature>
Proposed Architecture

Category 1 Websites
Website 1
Website m

Category k Websites
Website 1
Website m

Distributed server 1
Agent
DB
Wrapper

Distributed server m
Agent
DB
Wrapper

HTML
HTML

Negotiation
Control

GML
GML

Mediator
MetaDB
Replica
Query Engine

Client: GUI for Visualization and Spatial Query
Advantages of Proposed Architecture

• **Distributed**: it allows small and specialized software agents to search for region/domain specific GRI over the Internet and makes the search result reusable.

• **Open**: any system supports GML can make use of the result from search agents and integrate it into its own application to provide new value-added services.
Motivation of the prototype

The current OU search engine cannot answer searches likes:

• Search all the freshmen that are less than 5 miles away from me

• List all the graduate students who live around the Parkview apartments
Prototyping Based on the Architecture

We need to materialize the four components in the proposed architecture:

- Software agents to collect GRI over the Web pages
- GML Wrapper for Flat File, RDBMS and ORDBMS
- Spatial-Enabled Mediator
- Client Query and visualization GUI
Prototype Design

- Two distributed servers.
  - SQL Server 2000/Retrieve the information of students whose surname begin with a letter ‘A’ (198)
  - Oracle 8i /Retrieve the information of students whose surname begin with a letter ‘B’ (519)
- One Mediator Server using Oracle 8i
- One Web Server hosting Client Visualization Java Classes

- Two types of Base Map Servers (Wrappers)
  - ESRI ShapeFile
  - Oracle 8i
Experiment Environment

Operating System: Windows NT 4.0
Web Server: Java Server Web Development Kit (JSWDK) 1.0
Client Browser: Internet Explore 5.5
Java Compiler and Runtime: SUN JDK SE 1.3

Application Servers by using Virtual Directories:

• Distributed Server 1 (Wrapper for SQL Server 2000)
• Distributed Server 2 (Wrapper Oracle 8i)
• Base Map Servers (Wrapper for Flat File/DBMS)
Building the Agents using Java Swing HTML Parser

1. Open an URLConnection for a given URL
2. Open an InputStreamReader to read the HTML page
3. Build an HTMLEditorKit to process the HTML page
4. Create a Document from the kit to represent the page
5. Bind InputStreamReader and Document using HTMLEditorKit
6. Go through the hierarchy of HTML document tree
7. Get tag name, tag values and textual data
Building the Agents using Java Swing HTML Parser (Cont.)

Parse the textual data

Name, email, status, college

Street address, city, state, zip

Phone

Address Info

Geocoding

longitude/latitude

Validation

File/Database update

End
Geocoding: An Example

Tanya Babarykina
tatyana@ou.edu
Senior
College of Business
402 Wadsack Dr E
Norman, OK 73072
(405) 325-9282

(-97.443067, 35.194425)
Building GML Wrappers

• Flat File (ESRI ShapeFile): data Storage and No Query Support at all – Wrapping the whole file.

• RDBMS (Microsoft SQL Server): support Text/Numeric query but not Spatial Query –Wrapping Text/Numeric query result along with its associated geometric information

• ORDBMS (ORACLE 7.0+): has predefined Spatial query operators –Wrapping spatial/non-spatial query result
Building GML Wrapper (Cont)

Create New GML Document

Loop Through GRI records in File/DBMS

Switch type:

Point

Polyline

Polygon

CreateElement

CreateAttribute

CreateElement (Clist for Geometric Data)

Append Node to GML Document

Output the Generated GML Document
Building GML Wrapper: An Example from Oracle ORDBMS
Building the Mediator

Client Request

Break down the query into queries for individual distributed servers involved

For each distributed server

Spatial Query Support?

Yes → Send sub-request to distributed server

No → Check last update time

Is update needed?

No → Send request to distributed server

Yes
Building the Mediator (Cont.)

1. Transfer the returned GML document into XML nodes

2. Send request to distributed server
   - Update Mirror Database in Mediator with returned GML document
   - Query on mirrored database in the mediator
   - Transfer query result into XML nodes

3. Combine all the XML nodes and associated metadata information

4. Output the Generated XML Document
   - End
Client Side Spatial Query and Visualization

GeoTools: Open Source Java Package From University of Leeds, UK

Base Map: Spatial Query Result From Oracle/Flat File

Thematic Search Result From Mediator
Client Side Spatial Query and Visualization: National Scale
Client Side Spatial Query and Visualization: State Scale
Client Side Spatial Query and Visualization: County Scale
Client Side Spatial Query and Visualization: City Scale
Summaries

- We have proposed an open architecture for building GRI search Engine over the Internet

- We have built GML wrappers for Flat File, Pure Relational DBMS and Object-Relational DBMS

- We have built a mediator server with full spatial query capability using Oracle Spatial functionalities

- We have built client visualization functions for query results using the open source package GeoTools
Future Directions

- Parsing more generic HTML pages and searching GRI in them
- Efficient and Scalable Geocoding methods
- Spatial and temporal representation of geographical information
- Better Graphic User Interface (GUI)
Thanks!

Questions?