Semantic Geospatial Data Integration and Mining for National Security

Latifur Khan

Ashraful Alam

Ganesh Subbiah

Bhavani Thuraisingham

University of Texas at Dallas

(Funded by Raytheon Corporation)

Shashi Shekhar

University of Minnesota

Geospatial Data Integration: Motivating Scenario

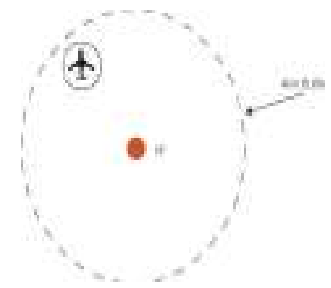
Query: "Find movie theaters within 30 miles of 75080"

within, near, overlap – Geospatial Operators

Theaters, Restaurants – Businesses (Non-Geospatial data)

Miles – Distance Unit

75080 , Richardson – Geo References



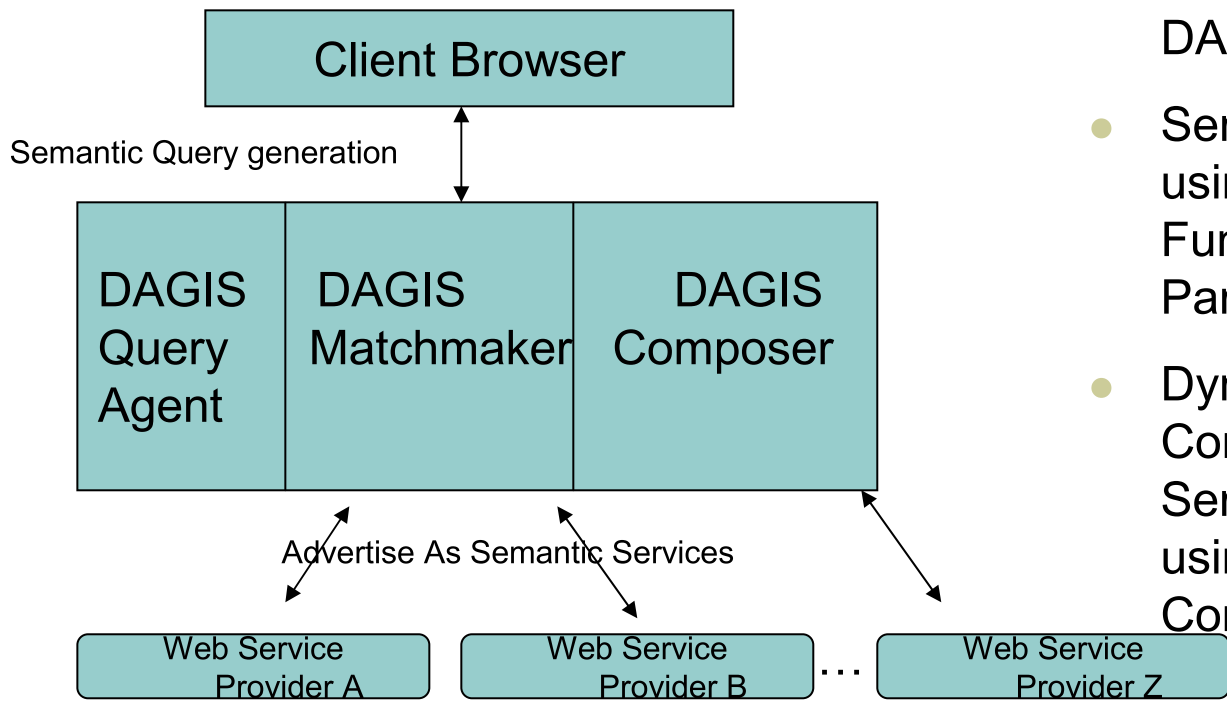
Cinemark Movies 10

Radisson Hotel Dallas North-Richardson

Key Contributions

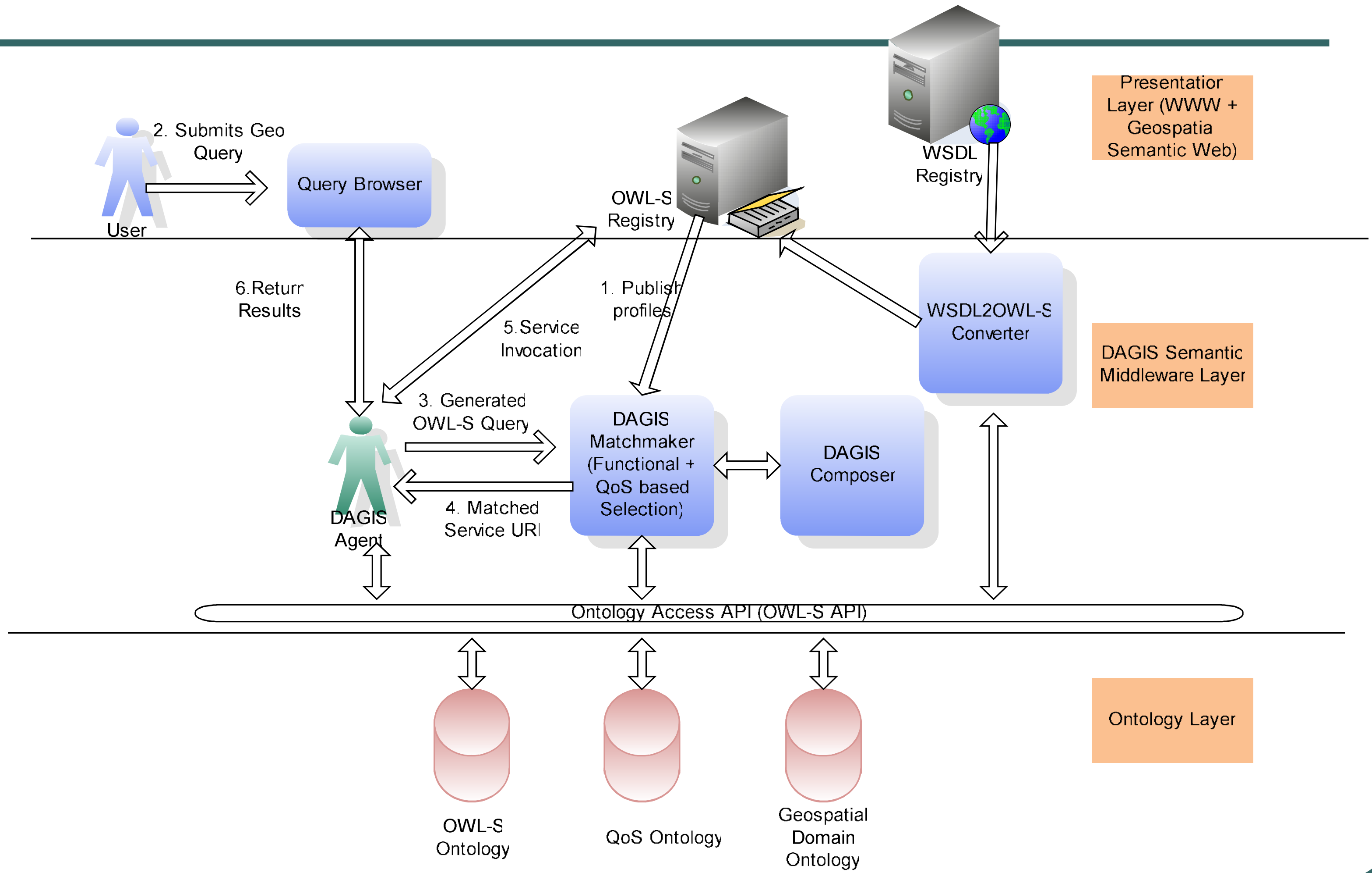
- Query can be handled by
  - Traditional search engine – Google, Yahoo –Not at Semantic level
  - Almost no search engine facilitates finding relevant web services (except Woogle – template matching for web services & no composition) and handle complex queries
- **DAGIS – Discover geospatial semantic web services using OWL-S Service ontology coupled with geospatial domain specific ontology for automatic discovery, dynamic composition and invocation**
  - Facilitates semantic matching of functional and non-functional services from various heterogeneous independent data sources.

Key Contributions

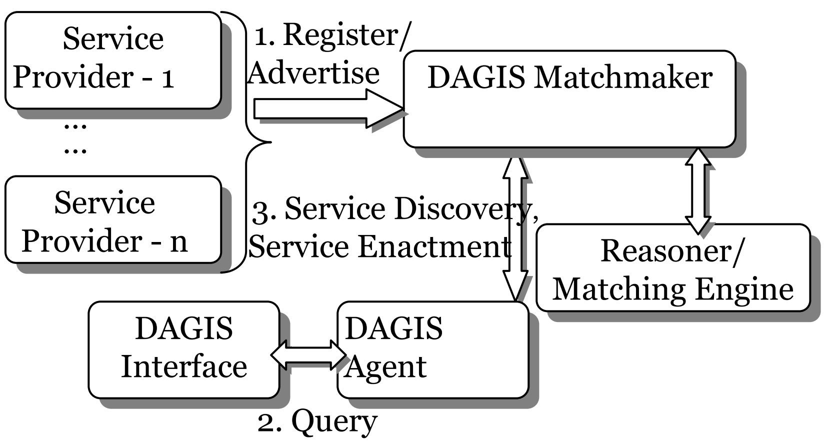


- Automatic Semantic Query Generation by DAGIS Query Agent
- Semantic Matching using Matchmaker for Functional and QoS Parameters
- Dynamic on the Fly Composition for Service orchestration using DAGIS Composer

DAGIS System Architecture



DAGIS System Flow

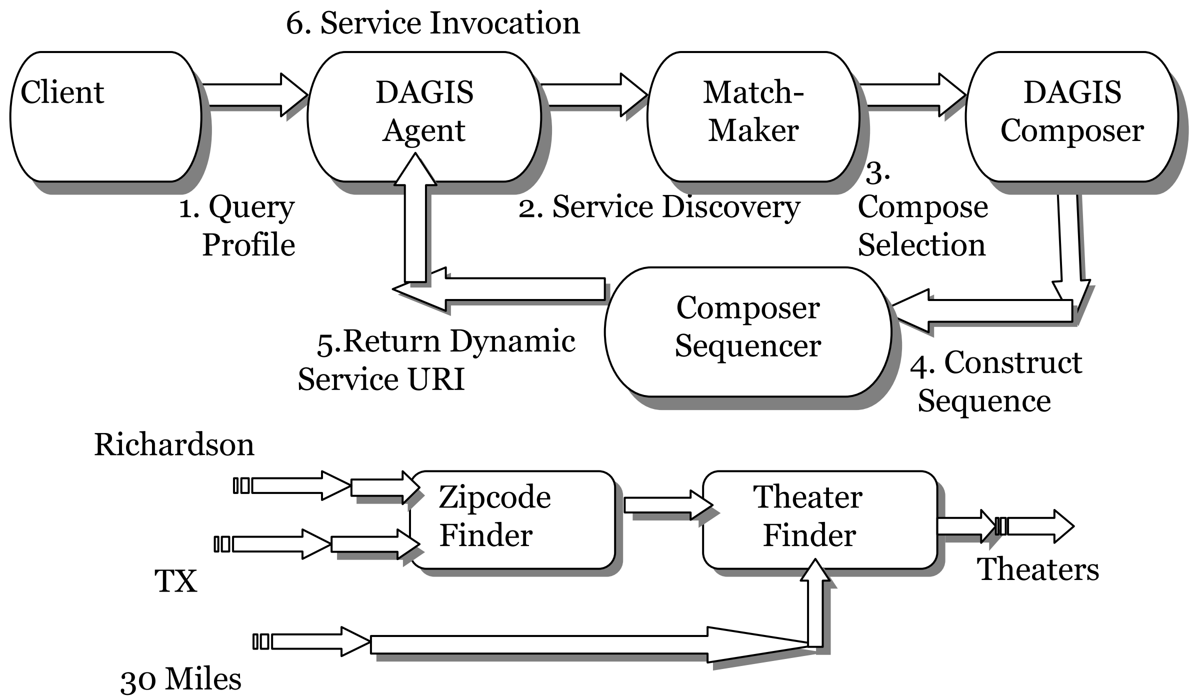


- DAGIS Query Interface
- OWL-S MatchMaker
- OWL-DL Reasoner for Matchmaker[1]
- Service Providers

[1] Pellet is an open source, OWL DL reasoner: http://pellet.owldl.com/

DAGIS for Complex Queries

Find Movie Theaters within 30 Miles from Richardson, TX



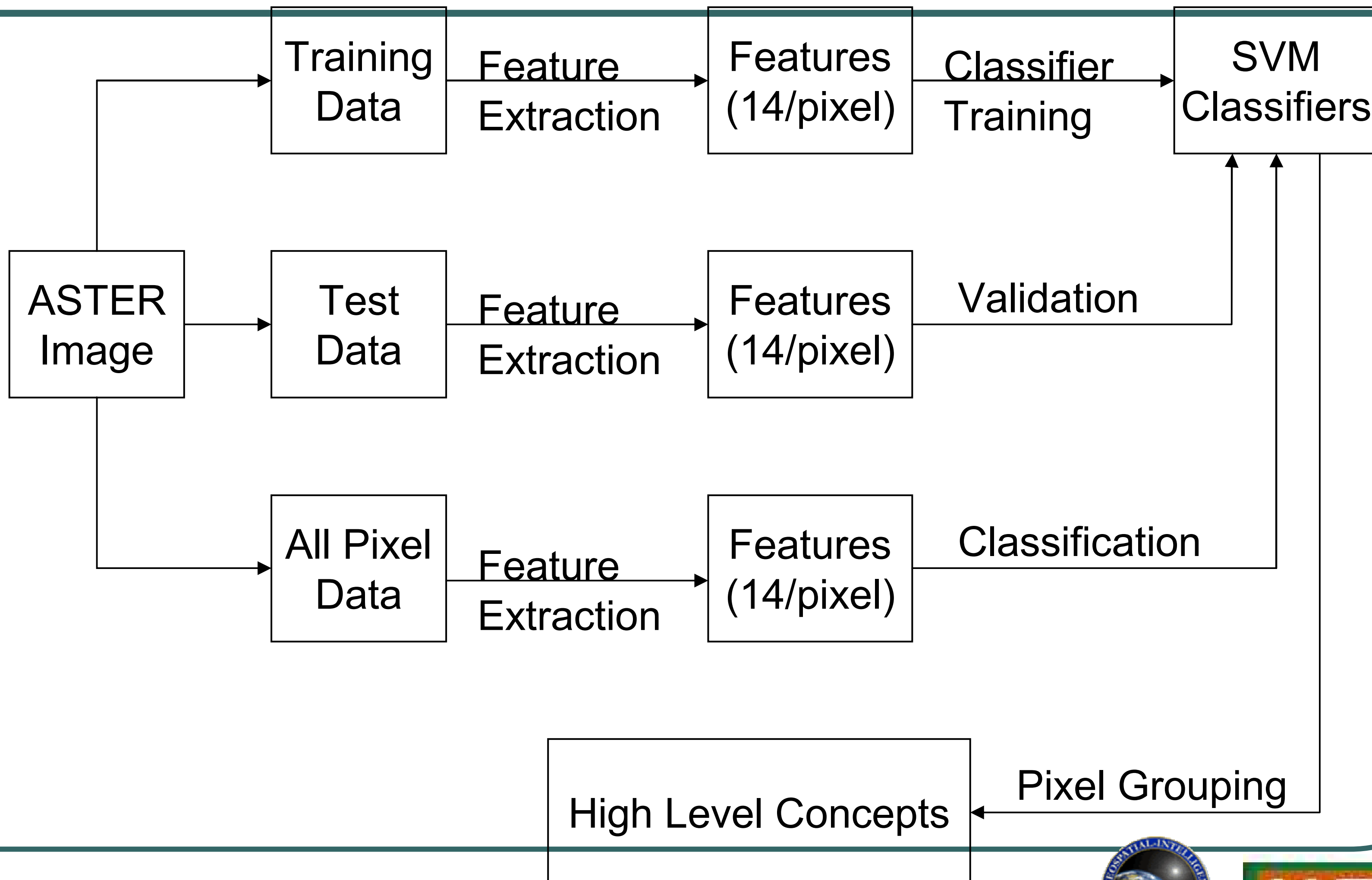
Geospatial Data Mining: Case Study: Dataset

- ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer)
  - To obtain detailed maps of land surface temperature, reflectivity and elevation.
- ASTER obtains high-resolution (15 to 90 square meters per pixel) images of the Earth in 14 different wavelengths of the electromagnetic spectrum, ranging from visible to thermal infrared light.
- ASTER data is used to create detailed maps of land surface temperature, emissivity, reflectivity, and elevation.

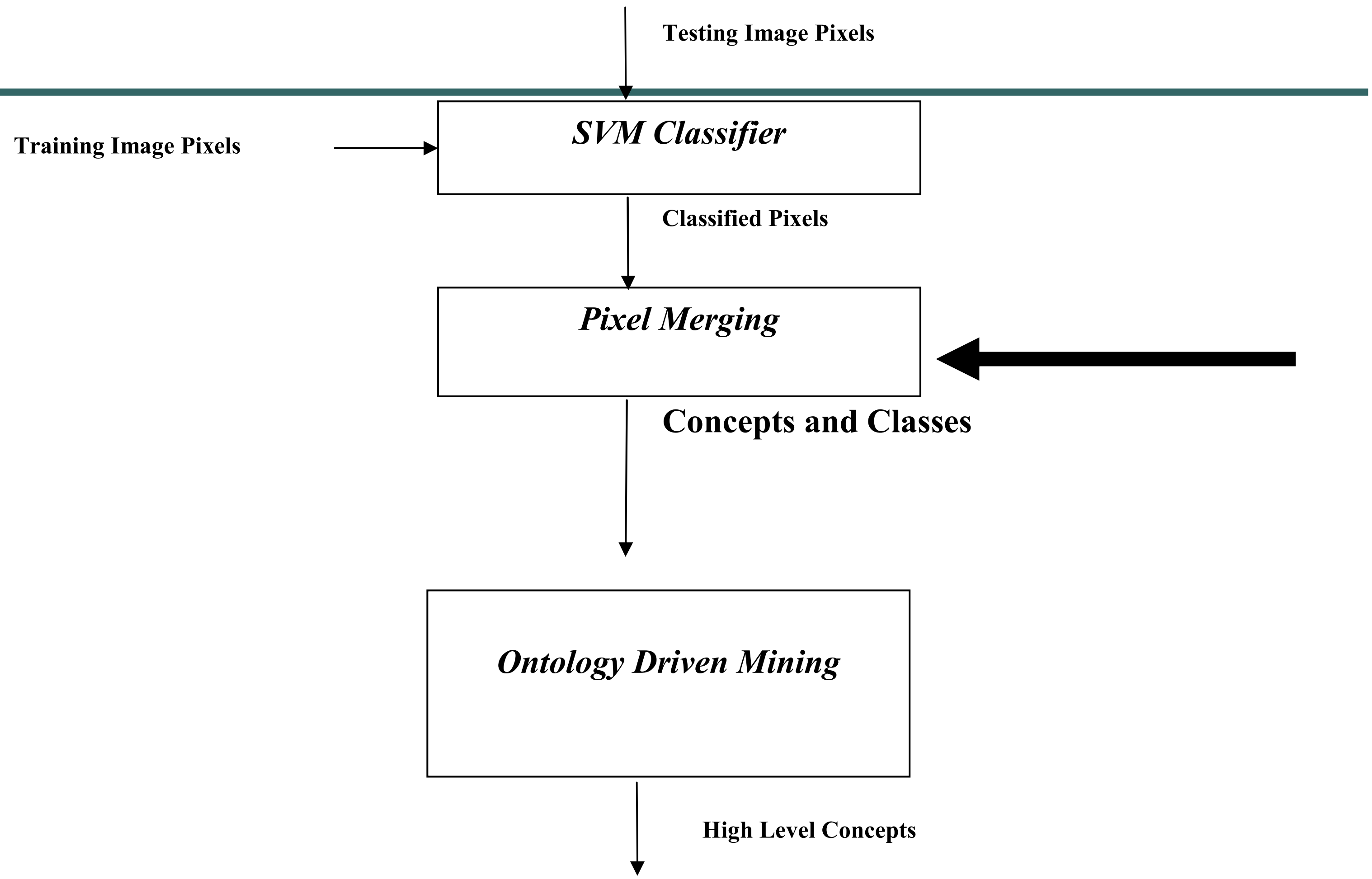
ASTER Dataset: Technical Challenges

- Testing will be done based on pixels
- Goal: Region-based classification and identify high level concepts
- Solution
  - Grouping adjacent pixels that belong to same class
  - Identify high level concepts using ontology-based mining

Sketches: Process of Our Approach



Process of Our Approach



Ontology-Driven Mining

- We have developed domain-dependent ontologies
  - Provide for specification of fine grained concepts
  - Concept, “Residential Area” can be further categorized into concepts, “House”, “Grass” and “Tree” etc.
- Generic ontologies provide concepts in coarser grain

Challenges

- Region growing
  - Find out regions of the same class
  - Find out neighboring regions
  - Merge neighboring regions
  - Not scalable
    - Irregular regions
    - Of different sizes
    - Hard to track boundaries or neighboring regions
- Pixel merging
  - Only neighboring pixels considered
  - Pixels are converted into Concepts
  - Linear

Output:

Table 1. Training and Test Pixels and Accuracy

	Water	Barren Land	Grass	Tree	Building	Road	House	Total	Accuracy(%)
Training	1175	1005	952	887	1041	435	1584	7079	99.8
Test	1898	1617	1331	1479	768	648	1364	9105	89.25

Table 2. Confusion Matrix for Independent Test Data

Class	Predicted						
	Water	Barren Land	Grass	Tree	Building	Road	House
Water	1898	0	0	0	0	0	0
Barren Land	0	1225	216	0	143	33	0
Grass	0	15	1175	54	69	0	18
Tree	0	0	0	1454	0	0	25
Building	0	1	0	0	578	189	0
Road	0	0	0	0	143	500	5
House	0	0	0	0	9	59	1296
Accuracy	100.00	75.76	88.28	98.31	75.26	77.16	95.01

Security and Privacy Challenges

- Security
  - Policy (context, association, event, time-based)
  - Access control, accountability
  - Policy integration
- Privacy
  - What does it mean to ensure privacy for geospatial data?
  - Protect the location of an individual?
  - If your residence can be captured by Google maps, then how can you protect it?