World Applied Sciences Journal 3 (Supple 1): 127-135, 2008
ISSN 1818-4952

Integration of Multi-Source Spatial Datasets Via Web Services Semi-Automatically and Ontology Based

[1]E. Mohammadi, [2]W. deVries and [3]A. Mansourian

[1]Department of Geological Engineering, Middle East Technical University (METU), Ankara, Turkey
[2]Department of Urban and Regional Planning and Geo-information Management, ITC, Netherlands
[3]Department of GIS Faculty of Geodesy and Geomatics Engineering,
K.N. Toosi University of Technology, Tehran, Iran

Abstract: This paper focuses on integration of multi-source datasets which often occur in spatial planning process, land administration and management, management of resources and site selection in Iran. In many cases in Iran; land records are available as textual data (definition of parcels in title deeds), vector datasets, aerial photos and satellite images in different forms. The problem is that there isn't a template and framework for integration of datasets from different sources. Poor interoperability among organizations, data sharing and data exchanging becomes complicated, if one does not have a complete or proper understanding of how text in a land related document relates to (spatial) reality or how solve the incompatibility of different datasets. Using Web services as a bridge between end users and spatial data providers, is a prudently and judiciously way for spatial data integration, that helps in saving both time and costs. Accordingly this paper, describes two scenarios where textual data are available which contain at least some spatial reference, which can be used to link to other vector datasets. The process through which integration of text and vector dataset is made possible is through the use of ontology. As the result of this research, semi automatic integration of textual datasets and other spatial datasets via Web services helps the people and decision makers in their house or office have the capability of integrating textual data and any other datasets from different sources. Some recommendations are presented for other researches.

Key words: Multi-source spatial datasets . textual datasets . heterogeneity . interoperability . integration . web services . ontology

INTRODUCTION

Integrating Web services capabilities, multi-source datasets, Web services' structure, concepts and relationship of its components helps to the clients and users to obtain their interested datasets in short time and low cost. Every research needs to an updated and processed spatial datasets. Moreover this helps both users and decision makers to access the datasets online. The main objective of this research is structuring a template for integrating multi-source datasets semi-automatically via Web services in Iran. Textual datasets and spatial datasets are two heterogeneous datasets that are integrated semi-automatically in this research [1].

Two scenarios are solutions for semi automatic integration of textual data and spatial datasets in Iran considering nation wide policy, regional culture and interoperability of software, hardware and any multi-source spatial datasets. Reforming and reorganizing the spatial datasets is real beginning of sustainable development. I have some recommendations for interoperability among National Cartography Centre (NCC), National Geography Organization (NGO) and other organisations in Iran. These attempts can be a base for building SDI and Cadastre of Iran in a Knowledge base method. Reorganizing the spatial datasets in Iran as a developing country is a vital step for every planning and designing in infrastructural projects.

Scenario one: Integration of text and vector dataset: It is a typical case where integration of vector data and a textual data related to a parcel is required by a land practitioner, but where these two data sources are heterogeneous in structure and content. Often, however, in such cases the vector dataset and the text dataset are assumed to relate to each other, either based on experience, or based on specific elements in the text. As an example, the vector dataset could be a topography dataset or a cadastre dataset and the textual data source

Corresponding Author: Dr. E. Mohammadi, Department of Geological Engineering, Middle East Technical University (METU), Ankara, Turkey

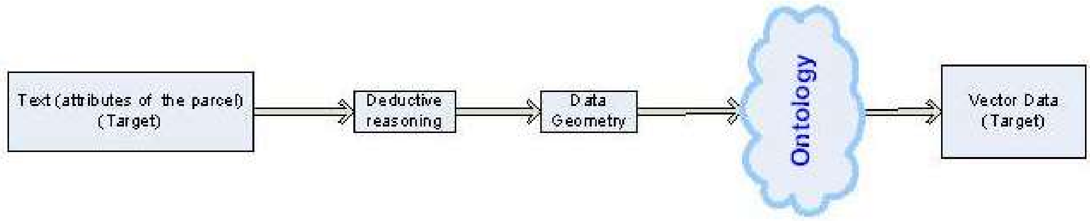


Fig. 1: Converting text dataset to vector dataset semantically

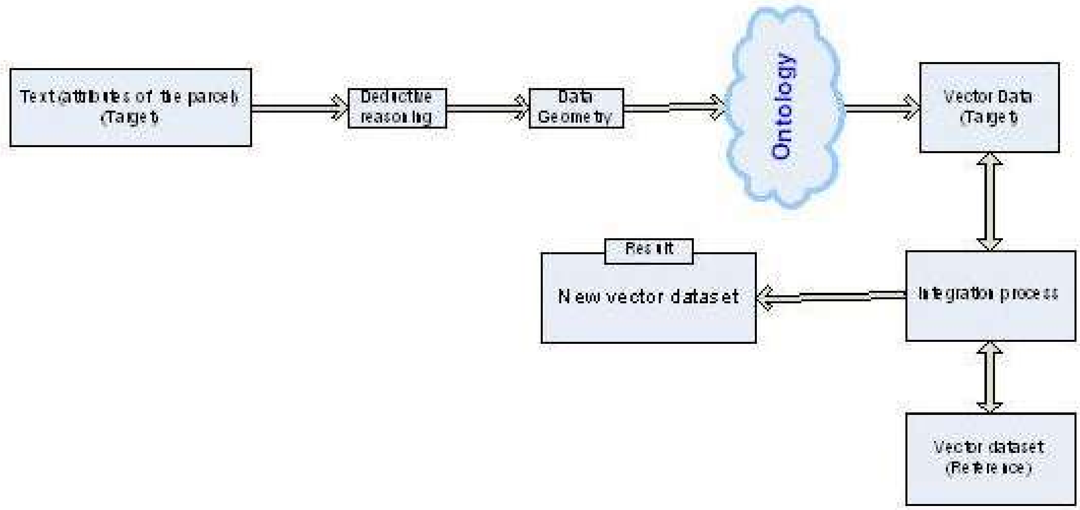


Fig. 2: Schema for integration of text and vector dataset

could be a text which is derived from a title deed. A title deed typically contains the rights of the ownership of the property, but also contains specific spatial information about the parcel, such as the dimension, area and address. First I will describe how this process can be done through human intervention and interaction; secondly I will aim an automating most of the procedure (Fig. 1).

General process of integrating text and existing vector dataset: To link the two, I base myself on the concepts of ontology. Ontology theory helps to present and dissect text in a structural way [1]. This can be done through mediators. Mediators are any information about a spatial entity, in this case the parcel, which can help link the geometry of the parcel to the surface of the ground in order to demarcate the parcel on the ground. The information about the dimension and the area can be extracted from the text to define a geometry concept of the property (data geometry). This text semantically addresses a unique property which is single in kind (a property has registration number, owner, address, dimension and area). Schematically this can be depicted as follows in Fig. 1.

Often title deeds contain text which relates directly or indirectly to parcels. An experienced user of such texts or someone who is acquainted with the area can deduct what sort of parcel is mentioned in the text. With some form of deductive reasoning about text one could draw basic conclusions with regards to a geometry shape. For example, certain references could be made with regards to the dimension and area of a parcel and object can be a parcel. Once this deductive reasoning is completed, the next step would be to translate the obtained data into data geometry. In schematic sequence, one could say:

Text of a deed contains dimension and area of a property. Dimension and area can be explained by data geometry. Text of a deed can be explained by data geometry. 1- a -> b2- b -> c3- a -> c

This whole process could be extended to a computer environment through the use of ontology. In both computer science and information science, ontology make use of data models that represent a domain and that use reason about the objects in that domain and the relations between them. Ontology is used in artificial intelligence Web, software engineering and information architecture as a form knowledge representation about the world or some part of it. Semantic data (text), deductive reasoning, ontology in computer language, data model, domain of objects and reasoning about existence of parcels helps to integrate a text with a vector dataset.

When a text describes a parcel, some spatial information comes to mind. For example if the text describes a property in Karaj, the reader does not think to a property in Taleghan. After extracting the kind of the parcel; house or farm and concerning dimension and area, the reader builds an image (semantic data) of the property in his or her mind. It is expected; using an intelligence mechanism and representing a domain, two previously mentioned heterogeneous datasets be integrated and harmonized. Figure 2 shows procedure of converting a text to data geometry and the process of the integration of these two heterogeneous datasets.

The intelligence mechanism can greatly benefit from the use of interoperable dimensions of the parcel, which are transformed to a geometry shape (data geometry). The Italic text below shows data geometry, which is derived from a deed title. To use this theory on a practical example, a typical text from a title deed is provided below. Most title deeds in Iran have prescribed words and elements in their descriptions of parcels. In translation, a part of the text, which describes dimensions and the area of the parcel in such a title deed are as follows:

- "Area is 45751.29 m2. The boundary's dimensions in clockwise direction are:
- North: from west to the east is 183.99m.
- East in five parts: first from north to south is 99.07m, second part from east to west is 22.95m, third part from north to south is 83.40m, fourth part from west to east 29.66m and fifth part from north to south is 89.76m.
- South: from east to the west is 163.28m.
- West: from south to north is 289.34m" (ends to the start point).

What are clear are the references to spatial features. The words or abbreviations "area", "north", "m", etc. refer to these elements. Moreover, it was agreed that all the numbers have a prescribed sequence, namely clockwise descriptions of the boundaries. Again, this pre-known information can be used to translate the text into spatial data. In Fig. 3, the data geometry is derived from descriptions of the parcel. A deductive reasoning (by a professional and/or someone familiar with the area) describes the data geometry to make a polygon.

The geometry that is shown in Fig. 3 is created based on human intelligence using deductive reasoning. Based on this manner to translate individual parcels into spatial objects, one could continue for the whole area of parcels in which one would be interested.

Figure 4 and 5 show the result of drawing the parcel automatically, then adjusting the polygon using written software.

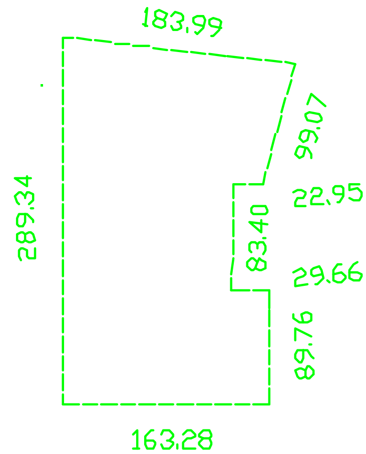


Fig. 3: Data geometry derived from attributes of the parcel

Now we are able match the target parcel to the reference map and then assign the attributes to the conjugate polygon in the reference map.

Scenario two: Integration of a text dataset and vector dataset, ontology based: The second scenario is integration of vector dataset and a text dataset (attributes of parcels) that are inherently two heterogeneous datasets. The difference with the previous scenario one is that there may not be obvious textual elements through which one can link the two data sources, or that the procedure in scenario one is still leading to inappropriate results. The mediation between the two data sources will then have to go in more than one step and by using a number of other data sources, such as aerial photo's, old paper maps, other datasets and additional field investigations. Also in this scenario ontology is used to be able to define a meaningful definition about elements in the text that can not be explicitly determined.

The obvious problem described here is that the parcel can't be matched with any topographic data. Figure 6 shows conceptually the text dataset (attributes of the property) and the up-date vector dataset. When there are some lost chains for integrating attributes and vector datasets, using the mediators and ontology help to find the lost chains.

Figure 6 shows the lost chains in integration of the text and spatial datasets. For integration we should find these lost chains using paper maps aerial photos, etc. The lost chains lead to refer to every available dataset and ontology to find link and relation between text dataset and the reference dataset. It is expected the attributes of the property which are perceived from the text, show a parcel or an area in the topography dataset.

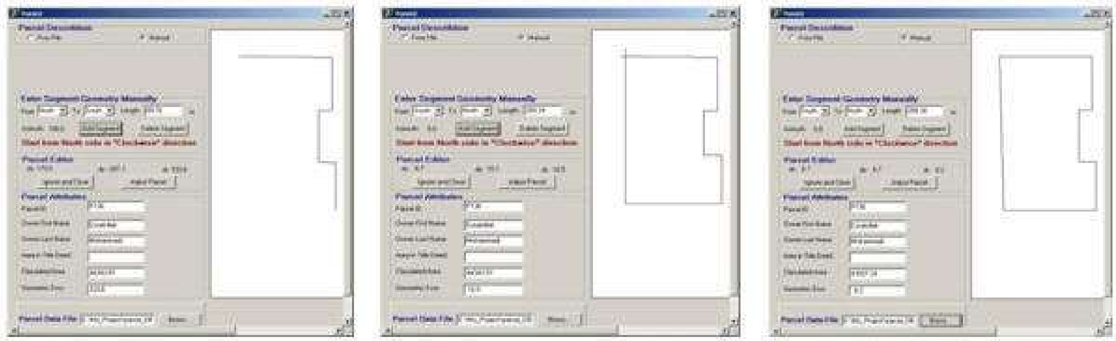


Fig. 4: Stages of completing the geometry of the parcel

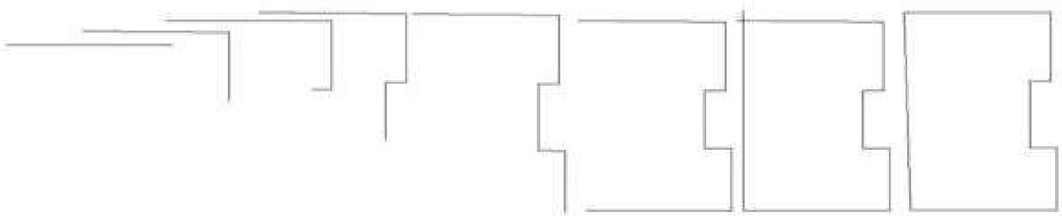


Fig. 5: The 8 stages of drawing the geometry of the parcel by the software

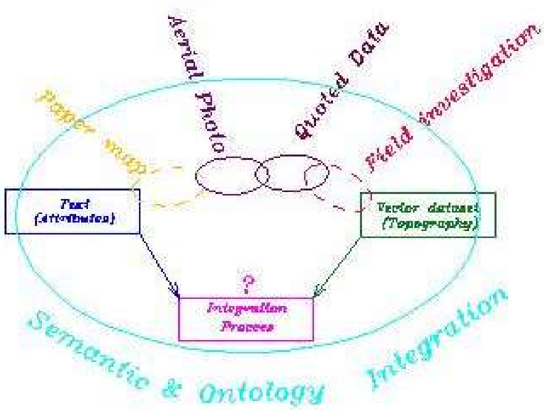


Fig. 6: The text dataset (attributes), topography dataset and the lost chains

However, many changes on the ground during years caused the trace of the boundary of the parcel on the ground be annihilated. These chains are mediators between attributes of the parcel and topography dataset. We link the text to the spatial datasets using mediators, relying on the ontology (Fig. 6). Ontology provides a way to share the semantics of concepts in some area of interest such as Lemmens [2] quoted data, field investigation, paper map and textual data.

Pragmatic way of integration (ontology based): These datasets are all or some of the lost chains. For a practitioner the best and most pragmatic way is by asking the relevant people for any information about the property (quoted data) and going to the field (for local investigation) as is shown in Fig. 7.

The drawn (perceived) map from ontology reasoning in Fig. 7 is aimed to be integrated with the topography dataset semantically (right image in Fig. 8). These two dataset have not common co-ordinate system. The common lines and features are a Mosque in south and the sharp corners of the river, which are shown by straight lines. The integration operation is done using align command in Auto Cad. In this process the text dataset that is inherently a spatial dataset, is geo-referenced. The main goal in this research is to integrate text dataset and spatial datasets automatically as comes below:

Reading text to integrate with spatial data automatically: An attemp t is done to read the text in word format automatically in order to make the results accessible in client interface thru Web services. Why and how we do it is as follows:

Integration of text and spatial data is important in Iran, since millions of deed titles are in text format, without coordinate system. In this thesis software is developed to read the texts automatically, link them with the shape file of the region. The link is done based on area, perimeter, sides and segments of the parcels that is extracted from text (Fig. 9). In transferring the

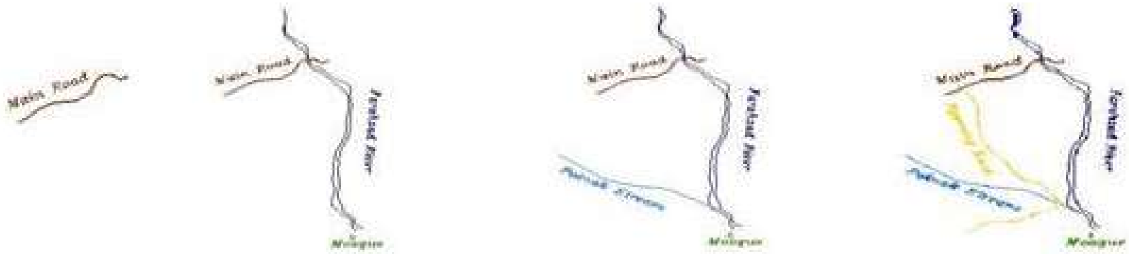


Fig. 7: Semantic perceiving of lost chains relying on quoted data, old maps, aerial photos and digital maps basing on ontology reasoning

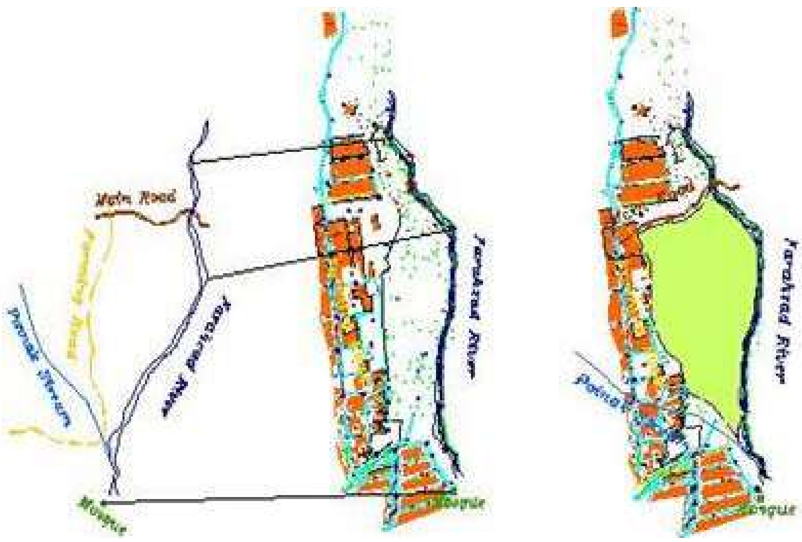


Fig. 8: Semantically link between integrated perceived map and up-date map

Parcel code is D123456, Owner is Mr. Ebrahim Omidi, Area is 250312.70 m2, perimeter is 2293.10m; the boundary's dimensions in clockwise direction are: North; from west to east is 224.20m. East; from north to south is 716.70m. South; from east to west is 880.10m. West; from south to north is 472.10m

Parcel code is M456, Owner is Mr. Kave Razmjoo, Area is 191079.02m2, perimeter is 1913.60m. The boundary's dimensions in clockwise direction are: North; from west to east is 645.30m; East; from north to south is 267.60m. South; from east to west is 695.90m; West; from south to north is 304.80m.

Parcel code is F658, Owner is Mr. Amir Radan, Area is 65320.26m2, perimeter is 997.52m. The boundary's dimensions in clockwise direction are: North; from west to east is 234.85m. East in four parts: 45.3m, 6.5m, 188m, and 36.20m. South; from east to west is 219.20m. West; from south to north is 267.62m.

Parcel code is F35680, Owner is Mr. Shervin Kamali, Area is 63915.55 m2, perimeter is 1025.87m. The boundary's dimensions in clockwise direction are: North, from west to east is 250.20m. East, from north to south is 309.22m. South; from east to west is 168.10m; West in two parts: 244.15m and 54.20m.

Fig. 9: Text of parcels that is read by software

attributes of the parcels, we do not use the name of the owner and parcel registration code, since in shape files normally there are not these attributes.

When software is executed; a window appears (Fig. 10) with two main browse buttons. In upper browse button, the shape file address is interred and in

Fig. 10: Our local Web page is ready for interring the texts of the parcels

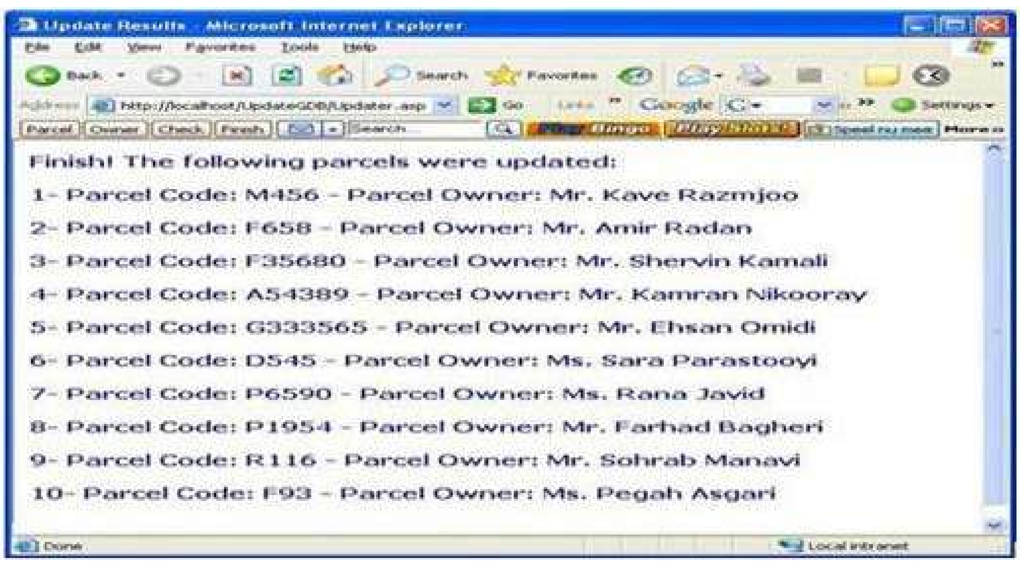
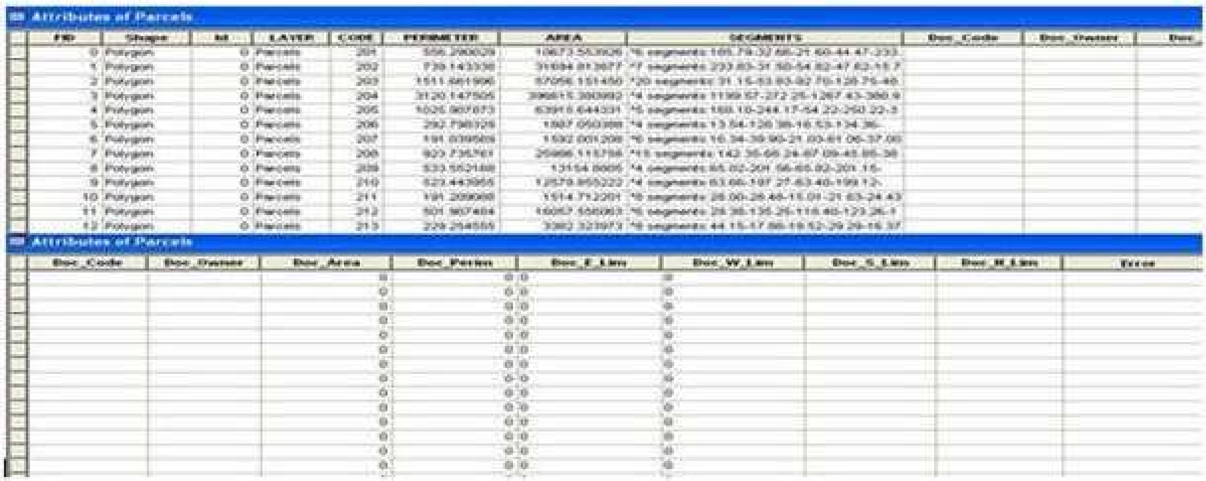


Fig. 11: The result of sending texts to geo-database and updating

lower browse button, the text address is interred. The linear and area tolerance buttons are to compare the parcels in the texts and the shape file according to predefined threshold for their area and perimeter. The software has the capability for reading one or more text chronologically and links their attributes with attribute table of the shape file.

Making integrated text and spatial data accessible via Web services (Intranet) for users: For this scope, we built Intranet in order to follow the procedure as is in the Internet. In fact, after being sure of working software successfully in VB6 and Arc GIS environment in, in the second step we implemented the procedure in our Web service (Intranet) using my laptop as a server. Programming language is Visual Basic 6 (VB6) and Web language programming to do link between algorithms and geo-database is Application Service Provider (ASP).

- My address in the implemented Intranet is:
- http://localhost/UpdateGDB/index.htm



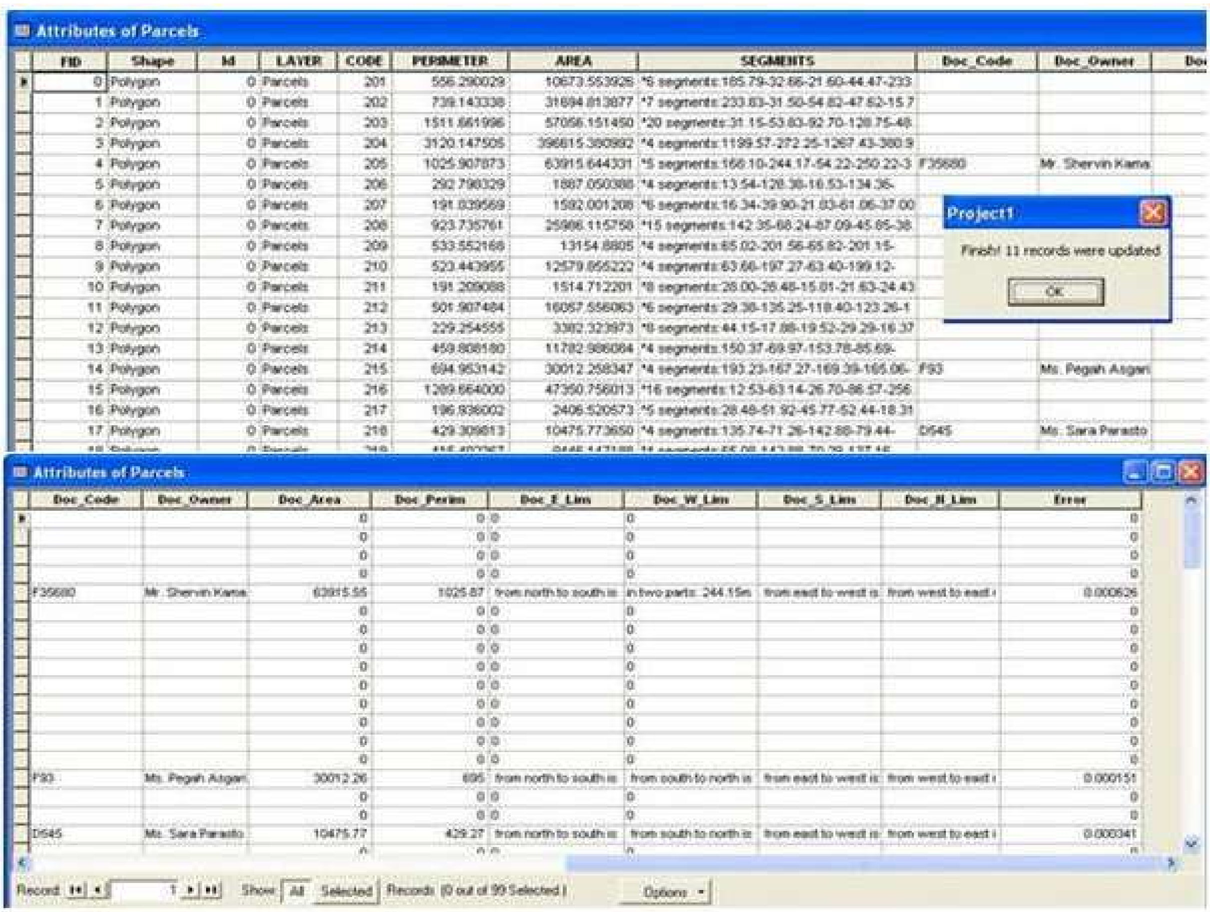
Attributes of Parcels

FID	Shape	Id	LAYER	CODE	PERIMETER	AREA	SEGMENTS	Doc_Code	Doc_Owner
0	Polygon	0	Parcels	201	556.290029	10673.553926	*6 segments:185.79-32.66-21.60-44.47-233		
1	Polygon	0	Parcels	202	739.143338	31694.813677	*7 segments:233.83-31.50-54.82-47.62-15.7		
2	Polygon	0	Parcels	203	1511.661996	57056.151450	*20 segments:31.15-53.83-92.70-128.75-48		
3	Polygon	0	Parcels	204	3120.147505	396615.380992	*4 segments:1199.57-272.25-1267.43-380.9		
4	Polygon	0	Parcels	205	1025.907873	63915.644331	*5 segments:168.10-244.17-54.22-250.22-3		
5	Polygon	0	Parcels	206	292.798329	1887.050388	*4 segments:13.54-128.38-16.53-134.36-		
6	Polygon	0	Parcels	207	191.039569	1592.001208	*6 segments:16.34-39.90-21.03-61.06-37.00		
7	Polygon	0	Parcels	208	923.735761	25986.115758	*15 segments:142.35-68.24-87.09-45.85-38		
8	Polygon	0	Parcels	209	533.552168	13154.8805	*4 segments:65.02-201.56-65.82-201.15-		
9	Polygon	0	Parcels	210	523.443955	12579.855222	*4 segments:63.66-197.27-63.40-199.12-		
10	Polygon	0	Parcels	211	191.209088	1514.712201	*8 segments:28.00-26.48-15.01-21.63-24.43		
11	Polygon	0	Parcels	212	501.907484	16057.556063	*6 segments:29.38-135.25-118.40-123.26-1		
12	Polygon	0	Parcels	213	229.254555	3382.323973	*8 segments:44.15-17.88-19.52-29.29-16.37		

Attributes of Parcels

Doc_Code	Doc_Owner	Doc_Area	Doc_Perim	Doc_E_Lim	Doc_W_Lim	Doc_S_Lim	Doc_N_Lim	Error
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			
		0	0	0	0			

Fig. 12: Attribute table in Arc GIS before link



Attributes of Parcels

FID	Shape	Id	LAYER	CODE	PERIMETER	AREA	SEGMENTS	Doc_Code	Doc_Owner
0	Polygon	0	Parcels	201	556.290029	10673.553926	*6 segments:185.79-32.66-21.60-44.47-233		
1	Polygon	0	Parcels	202	739.143338	31694.813877	*7 segments:233.83-31.50-54.82-47.62-15.7		
2	Polygon	0	Parcels	203	1511.661996	57056.151450	*20 segments:31.15-53.83-92.70-128.75-48		
3	Polygon	0	Parcels	204	3120.147505	396615.380992	*4 segments:1199.57-272.25-1267.43-380.9		
4	Polygon	0	Parcels	205	1025.907873	63915.644331	*5 segments:168.10-244.17-54.22-250.22-3	F35680	Mr. Shervin Kama
5	Polygon	0	Parcels	206	292.798329	1887.050388	*4 segments:13.54-128.38-16.53-134.36-		
6	Polygon	0	Parcels	207	191.039569	1592.001208	*6 segments:16.34-39.90-21.03-61.06-37.00		
7	Polygon	0	Parcels	208	923.735761	25986.115758	*15 segments:142.35-68.24-87.09-45.85-38		
8	Polygon	0	Parcels	209	533.552168	13154.8805	*4 segments:65.02-201.56-65.82-201.15-		
9	Polygon	0	Parcels	210	523.443955	12579.855222	*4 segments:63.66-197.27-63.40-199.12-		
10	Polygon	0	Parcels	211	191.209088	1514.712201	*8 segments:28.00-26.48-15.01-21.63-24.43		
11	Polygon	0	Parcels	212	501.907484	16057.556063	*6 segments:29.38-135.25-118.40-123.26-1		
12	Polygon	0	Parcels	213	229.254555	3382.323973	*8 segments:44.15-17.88-19.52-29.29-16.37		
13	Polygon	0	Parcels	214	459.808180	11782.986084	*4 segments:150.37-69.97-153.78-85.69-		
14	Polygon	0	Parcels	215	694.963142	30012.258347	*4 segments:193.23-167.27-169.39-165.06-	F93	Ms. Pegah Asgari
15	Polygon	0	Parcels	216	1289.664000	47350.756013	*16 segments:12.53-63.14-26.70-86.57-256		
16	Polygon	0	Parcels	217	196.936002	2406.520573	*5 segments:28.48-51.92-45.77-52.44-18.31		
17	Polygon	0	Parcels	218	429.309813	10475.773650	*4 segments:135.74-71.26-142.88-79.44-	D545	Ms. Sara Parasto

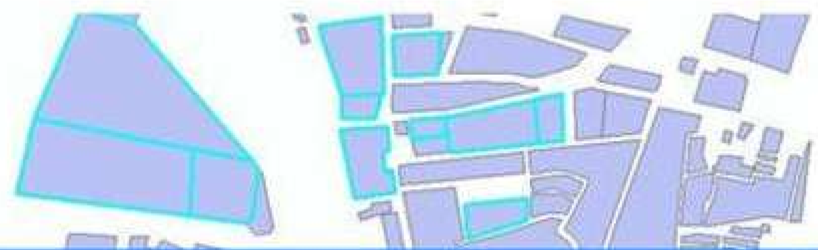
Attributes of Parcels

Doc_Code	Doc_Owner	Doc_Area	Doc_Perim	Doc_E_Lim	Doc_W_Lim	Doc_S_Lim	Doc_N_Lim	Error
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
F35680	Mr. Shervin Kama	63915.55	1025.87	from north to south is	in two parts: 244.15m	from east to west is	from west to east i	0.000626
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
		0	0	0	0			0
F93	Ms. Pegah Asgari	30012.26	695	from north to south is	from south to north is	from east to west is	from west to east i	0.000151
		0	0	0	0			0
		0	0	0	0			0
D545	Ms. Sara Parasto	10475.77	429.27	from north to south is	from south to north is	from east to west is	from west to east i	0.000341

Fig. 13: Attribute table in Arc GIS after link with attributes in the texts

- When we open "Launch Internet Explorer Browser" and inter our local address it is active (Fig. 9).When we push "Go" button Fig. 9 appears:

To test our local Web page (Fig. 10) we copy and paste the attributes of the parcels into the empty box "Enter your document text " in Fig. 11. Figure 12 is



Selected Attributes of Parcels

Doc_Owner	Doc_Area	Doc_Perim	Doc_E_Lim	Doc_W_Lim	Doc_S_Lim	Doc_N_Lim	Error
Mr. Kave Razmjoo	191079.02	1913.6	from north to south is 267.60m	from south to north is 304.80m	from east to west is 695.90m	from west to east is 645.30m	0.000043
Mr. Ebrahim Omidi	250312.7	2293.1	from north to south is 716.70m	from south to north is 472.10m	from east to west is 880.10m	from west to east is 224.20m	0.000245
Mr. Amir Radan	65320.26	997.52	in four parts: 45.3m, 6.5m, 188m,	from south to north is 267.62m	from east to west is 219.20m	from west to east is 234.85m	0.000258
Mr. Sohrab Manavi	19287.95	576.25	from north to south is 187.00m	from south to north is 173.42m	from east to west is 103.05m	from west to east is 112.10m	0.000678
Ms. Rana Javid	27560.15	719.95	from north to south is 90.95m	from south to north is 136.00m	from east to west is 246.90m	from west to east is 246.10m	0.000653
Mr. Farhad Bagheri	55908.2	1020.95	from north to south is 173.50m	from south to north is 141.60m	from east to west is 352.20m	from west to east is 361.65m	0.000566
Ms. Sara Parastooyi	10475.77	429.27	from north to south is 71.25	from south to north is 79.40m	from east to west is 142.90m	from west to east is 135.7m	0.000341
Ms. Pegah Asgari	30012.26	695	from north to south is 167.30m	from south to north is 165.10m	from east to west is 169.45m	from west to east is 193.20m	0.000151
Mr. Shervin Kamali	63915.55	1025.87	from north to south is 309.22m	in two parts: 244.15m and 54.20m	from east to west is 168.10m	from west to east is 250.20m	0.000626
Mr. Kamran Nikooray	13784.55	483.95	from north to south is 100.50m	from south to north is 94.50m	from east to west is 131.65m	from west to east is 157.30m	0.000868
Mr. Ehsan Omidi	44710.3	949.12	in five parts: first from north to s	from south to north is 283.05m	from east to the west is 163	from west to the east is 183.60	0.000106

Fig. 14: Result of using Intranet and Arc GIS to integrate text and spatial data automatically

attribute table of parcels in Arc GIS before link. Figure 13 is the attribute table in Arc GIS after linking with attributes in the texts.

When we submit the program, it links these texts with the spatial data in our geo-database and fills the columns of the owner, area, perimeter, four sides (north, east, south, west) and error automatically. The result can be seen thru Arc GIS as it is shown in Fig. 14 that eleven parcels are enhanced.

We assume that this geo-database and this sever belongs to Iran's SDI, every user from their PC interface can access and send their texts and integrate with spatial datasets. In this way, all the text of the parcels can be update with spatial datasets. The update parcels using Arc GIS can be found in this address:

c:\updatingSite\website\GDB\Parcels_GDB

In above address, C is the source drive in my laptop that now acts as server and the other parts of the address are the path to find the update parcels in the server. Update parcels are integrated texts and spatial data that can be used in cadastre, NSDI, municipalities and every application cases by users.

Figure 14 is the result of integrating the texts (containing attributes of 11 parcels) and shape file automatically using Visual Basic (VB) scripts and Arc GIS.

CONCLUSION

In Iran, there are a number of organizations that deliver data. The problem is however that in Iran data are still in different formats and those perceptions of the data are heterogeneous. This requires a further understanding of how to integrate Iranian data in a "regular PC environment" and also a better understanding and better mechanisms of how to integrate in a web-based environment.

It was found that integration of data in Iran must be placed within its context and the user needs. Like in other countries many data are distributed and maintained by different providers and users. A specific element in Iran is that while for many data sources NCC sets the accuracy standards, their data products are different from those of the private providers who have followed the other standards. One particular problem is also that many data are available in textual format only.

Integration of textual data and spatial data is important in Iran, since millions of deed titles are in the text format, without coordinate system, besides thousands of maps are available without any links among them. Heterogeneity of data sets strongly wastes time and cost. To reduce such problems; we tried to find a way for integration of textual and spatial datasets automatically.

We implemented integration of text and spatial (two heterogeneous datasets) semi-automatically and made it accessible via Web services for end users.

In this research software is developed to read the texts automatically, link them with the shape file of the region. The link is done based on area, perimeter, sides and segments of the parcels that is extracted from text (Fig. 14).

RECOMMENDATION

Integrating text and multi source spatial datasets thru our PC browser without using any software is recommended for another research. We attempted to integrate satellite images with different resolution in a meaningful way, but the result was not satisfactory. Hence, this scope can be the topic for other researches [3, 4].

REFERENCES

1. Mohammadi, E., W. de Vries, J. Morales and A. Mansourian, 2007. Addressing Spatial Data Integration Problems in the Context of SDI (MSc thesis). Department of Geoinformation Processing, ITC, Enschede, the Netherlands.
2. Lemmens, R., 2006. Semantic Interoperability of distributed Geo-Services, Published by: International Institute for Geo-Information Science and Earth Observation (ITC).
3. Chauhan, K.A., N.C. Shah and R. Venkata Rao, 2008. The Analytic Hierarchy Process as a Decision-Support System in the Housing Sector: A Case Study. World Appl. Sci. J., 3 (4): 609-613.
4. Mansourian, A. and M.J. Valadan-Zoje, 2008. Expanding SDI Hierarchy for Countries with Non-Federated System: A Case Study of Iran. World Appl. Sci. J., 3 (4): 597-602.