

# CONVERSION OF RANDOM DIGITAL FILES AND PRINTED DOCUMENTS INTO A STANDART DIGITAL FILE FORMATS WHICH IS SUITABLE FOR GIS

*Ismail Rakip KARAS<sup>1</sup> and Ibrahim BAZ<sup>1</sup>*

<sup>1</sup>Department of Geodetic and Photogrammetric Enginnering,  
Gebze Institute of Technology, Gebze, Turkey

*Management and organization of spatial data values form main topics of GIS. In Turkey and elsewhere, some computer programs have been used to carry out computational and graphical side of the mapping work without any idea of their contribution to a GIS, and outputs of these programs have been kept in digital files. On the other hand, some activities of municipalities and the government offices have been carried out with the use of computer programs, and the outcomes are kept in digital files or printed forms. That means, some of the data for GIS already exist in some way or another. Therefore, instead of trying to produce all GIS data from the beginning, converting the existing digital files which may be at random format into a standard format or recognizing characters of printed text files in digital forms will be a big support to establish GIS. This paper presents software development and application case studies.*

## 1. INTRODUCTION

It is not possible to mention about the existence of an information system which has no data at all. According to some researches, proportion of cost of the data gathering work to the total cost of an GIS is about 60% - 80%. On the other hand, it is also a fact that an effective GIS needs to be upgraded and renovated continuously. The results of some investigations indicate that proportion of a GIS's hardware cost to software cost and to data production cost is 1:10:100 respectively [Comert, 1996]. To make a GIS more economical and productive, if not necessary, GIS data, therefore, should not be obtained from the beginning stage, or should not be obtained from newly started projects. Instead, GIS data should be gathered from existing data resources as much as possible.

In this case, existing data formats may conflict with GIS's data formats. But this can be overcome by converting the formats of existing data in to a standard format which is suitable for GIS.

In Turkey, nationally well known CADD style computer programs such as EGHAS, KARTOCAD, NETCAD, etc. have been used to carry out automated mapping and zone application projects for the last ten years, an data belong to these projects has been archived either on computer storage facilities or in printed files. Therefore it may be considered that most of the data needed for establishment of a GIS exists in some way or another. This situation increases importance of converting the existing data formats to GIS data formats.

In this paper, software developments in which a number of data transformation processes carried out are introduced.

## **2. TRANSFORMATION OF GRAPHICAL DATA**

Effective usage and sharing of data appears to be an unavoidable fact in any situation where GIS applications are mentioned. In spatial data management, conventional form of data sharing is the data exchanging, or transferring data from different sources. In any data exchange process, there are two sides. The one which provides the data is named as “server”, the other one which requires the data is named as “applicant”.

In general, data transferring process is known as transformation of data from one format to another. Transferring process is being carried out with the use of two methods known as “direct data transferring” and “indirect data transferring” methods. In the first one, transferring is applied between two different data formats. Here, as transferring is carried out from one format to another directly, quality of data transferring is very high. Data transferring softwares developed for this purpose are named as “spatial data converters” [Comert, 1999].

In the second one, transferring between two data formats is being carried out through a interval format. Here, first of all, format to the server data is being converted to an interval format, and then the applicant transforms it into its own format [Comert, 1999].

## **3. TRANSFORMATION OF NON-GRAPHICAL DATA INTO A STANDARD FORMAT**

Most important peculiarity of GIS which distinguishes itself from other information systems is processing and management of attached graphical and non-graphical data. In the very first place, information of properties and then quantitative values of geographical features can be given as examples to the spatial non-graphical data. Non-graphical data examples can be seen as hand-writing or printed documents inside the land registry records, etc. or can be seen as either printed materials or computer text files of Word, Excel, Access programmes, etc.

Nowadays, it is possible to save or process hand-writing or printed files and documents as data files in a computer system through use of Optical Character Recognition (OCR) methods. For this, these printed files or documents are needed to be transferred into a computer system. For this purpose, raster images of these materials should be obtained with the help of a scanner. These raster images are then separated and converted into a text files with the use of OCR programs. Serious problems may occur during the separation process. Some conflicts can be faced while separating characters of the text whether or not they are letters or words. There are some methods which are used to ease or overcome this kind of problems. [Oztürk, 1998]

## 4. SOFTWARE INTRODUCTIONS

As stated above, to be able to make a GIS more economical and productive, it is necessary to benefit from the existing data at the very beginning stage of GIS establishment. For this purpose, some computer programs should be developed to make the existing data usable for GIS databases. In following texts, two computer programs dealing with graphic and non-graphic data transfers are introduced.

### 4.1 *Programme CFT For Coordinate and Parcel Data Format Transformation*

A computer program-CFT which carries out coordinate and parcel data transformation was developed in Visual Basic programming language. This programme accepts the coordinate and parcel data files and their formats as server format, and then transforms them to data files or formats of ACCESS programme.

### 4.2 *Standard Format Formation Programme – SFF*

There are two circumstances to mention about presence of non-graphical data. These data can be in the form of printed pages, hand or type-writer writings, or can be in the form of text files of EGHAS or NETCAD programmes.

At the first possible situation, there is a need for transferring of printed documents to a standard format. For this, raster images of these data should be obtained via a scanner scanning. Later on, with the use of OCR methods, the raster image firstly converted to a text file, and then some editing work is done to rearrange the text.

A computer programme-SFF which ables printed document or digital data usable for GIS was developed in Visual Basic programming language.

### 4.3 *Application Of CFT*

Flow chart of the execution of main steps of this programme is given in Figure 1. For the test of the CFT EGHAS's coordinate and parcel data files with ".ALN" extension and NETCAD's coordinate and parcel files with ".TXT" extension are used. Figures 2 and 3 shows formats of these data files. CFT programme reads these files and converts them to Access database which show propriety to GIS databases. Figure 4 shows format of Access database.

### 4.4 *Application of SFF*

Flowchart of execution of main steps of this programme is given in Figure 5. For the testing SFF, a text file produced by EGHAS programme (Fig. 9) and a document given in Figure 7 were chosen as the text file and printed document. Both the printed page and the text file consist of land registry information about parcels which were produced from a district zone application project in town Gebze.

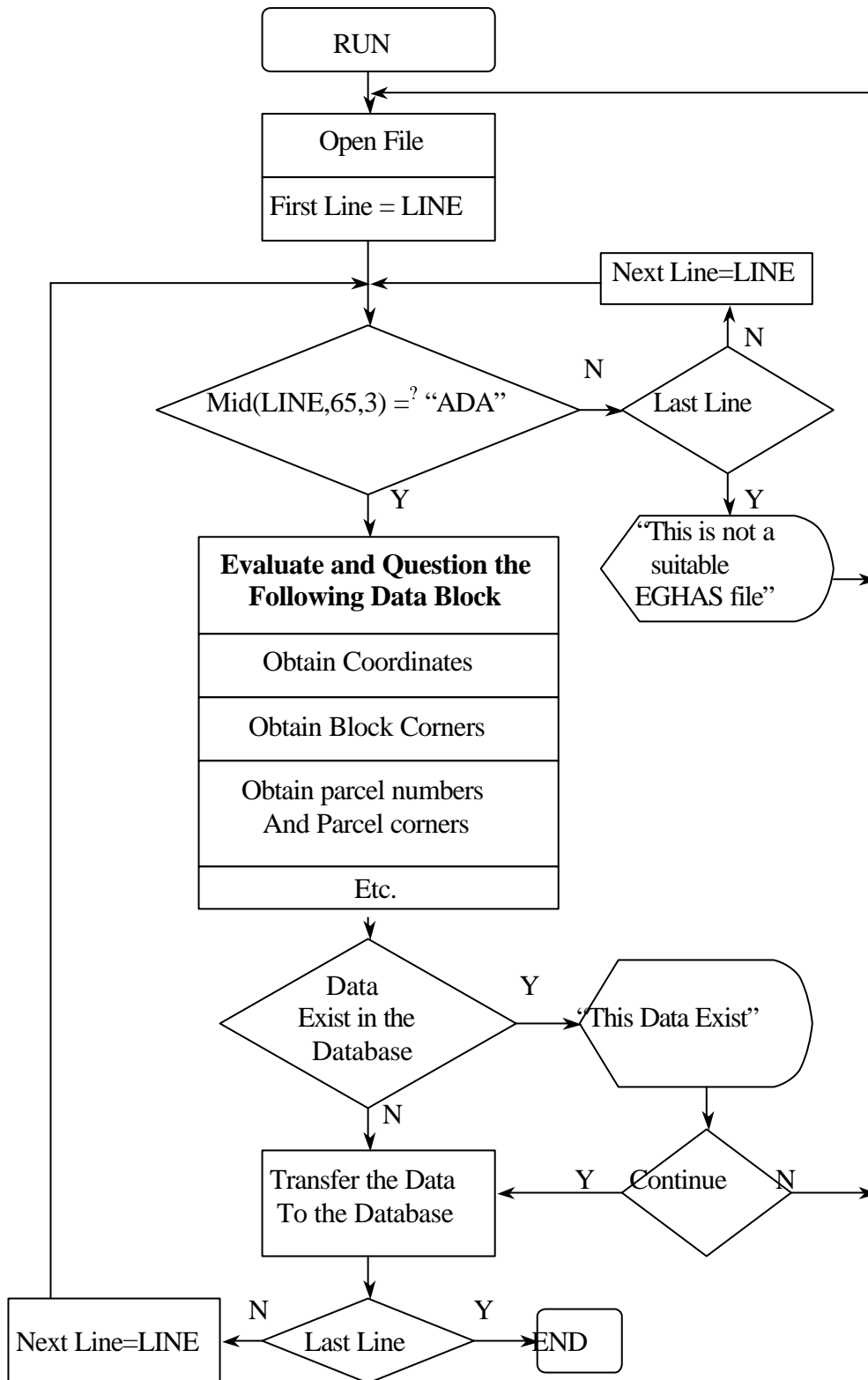


Figure 1: Flow Chart of the program CFT for EGHAS data

MS-DOS Komut İstemi - EDIT

Dosya Düzen Ara Görünüm Seçenekler Yardım

C:\WINDOWS\Desktop\bildiriler\GUZEL.ALN

İŞİN ADI : GÜZELLER ADA NO : 4155

KÖŞE KOORDİNATLARI

KÖŞE	Y	X	KÖŞE	Y	X
1	50576.04	52094.58	2	50590.51	52095.54
3	50591.46	52055.72	4	50574.81	52055.71
5	50575.18	52067.43	6	50591.18	52067.43
7	50575.66	52082.44	8	50590.82	52082.43

ADA KÖŞELERİ : 8,2,1,7,5,4,3,6,8

PARSEL ALANLARI

PARSEL NO	ALANI	KESİN	PARSELİ ÇEVRELEYEN KÖŞELER
1	187.02	187	1,2,8,7,1
2	233.78	234	7,8,6,5,7
3	191.25	191	5,6,3,4,5

F1=Yardım Satır:79 Sütun:68

Figure 2: Sample of server data (EGHAS's coordinate and parcel data)

MS-DOS Komut İstemi - EDIT

Dosya Düzen Ara Görünüm Seçenekler Yardım

C:\WINDOWS\Desktop\bildiriler\Neadkord.txt

Ada/Parse1	Alan	Nokta No	Y	X	Cephe
6339	2328.952	X1170	17499.829	-18184.046	67
		X1141	17522.468	-18247.103	33
		X1185	17489.407	-18254.640	66
		X1212	17466.260	-18192.197	31
		X1175	17495.894	-18181.494	4
6339/1	308.823	X1186	17489.021	-18202.771	17
		X1195	17481.560	-18186.671	16
		X1212	17466.260	-18192.197	18
		X1202	17472.754	-18209.716	17
6339/2	295.749	X1186	17489.021	-18202.771	16
		X1166	17504.352	-18196.643	13
		X1170	17499.829	-18184.046	4
		X1175	17495.894	-18181.494	15
		X1195	17481.560	-18186.671	17

Figure 3: Sample of server data (NETCAD's coordinate and parcel data)

Microsoft Access

Dosya Düzen Görünüm Ekle Biçim Kayıtlar Araçlar Pencere Yardım

4155 : Tablo

nno	y	x	adakose
1	50576,04	52094,58	-1
2	50590,51	52095,54	-1
3	50591,46	52055,72	-1
4	50574,81	52055,71	-1
5	50575,18	52067,43	-1
6	50591,18	52067,43	-1
7	50575,66	52082,44	-1
8	50590,82	52082,43	-1
*			0

parsel\_bilgileri : Tablo

pafta	adaparsel	kackose	1	2	3	4	5	6	7	8	9	10	11	12	nizam	kat	onbahce	yanbahce
G22b18c3b	6368/1	11	X73	X73	X73	X730	X	X	X						i	2	3	4
G22b18c3d	4200/2	5	79	80	81	82	7i								a	3	4	3
G22b18c4a	4155/1	4	1	2	8	7									a	4	5	4
G22b18c4a	4155/2	4	7	8	6	5									a	4	5	4
G22b18c4a	4155/3	4	5	6	3	4									a	4	5	4
G22b18c4a	6369/1	8	X71	X73	X74	X743	X	X	X					i	3	3	3	
G22b18c4a	6369/2	8	X70	X70	X66	X700	X	X	X					i	3	3	3	

Figure 4: Sample of applicant data in Access database

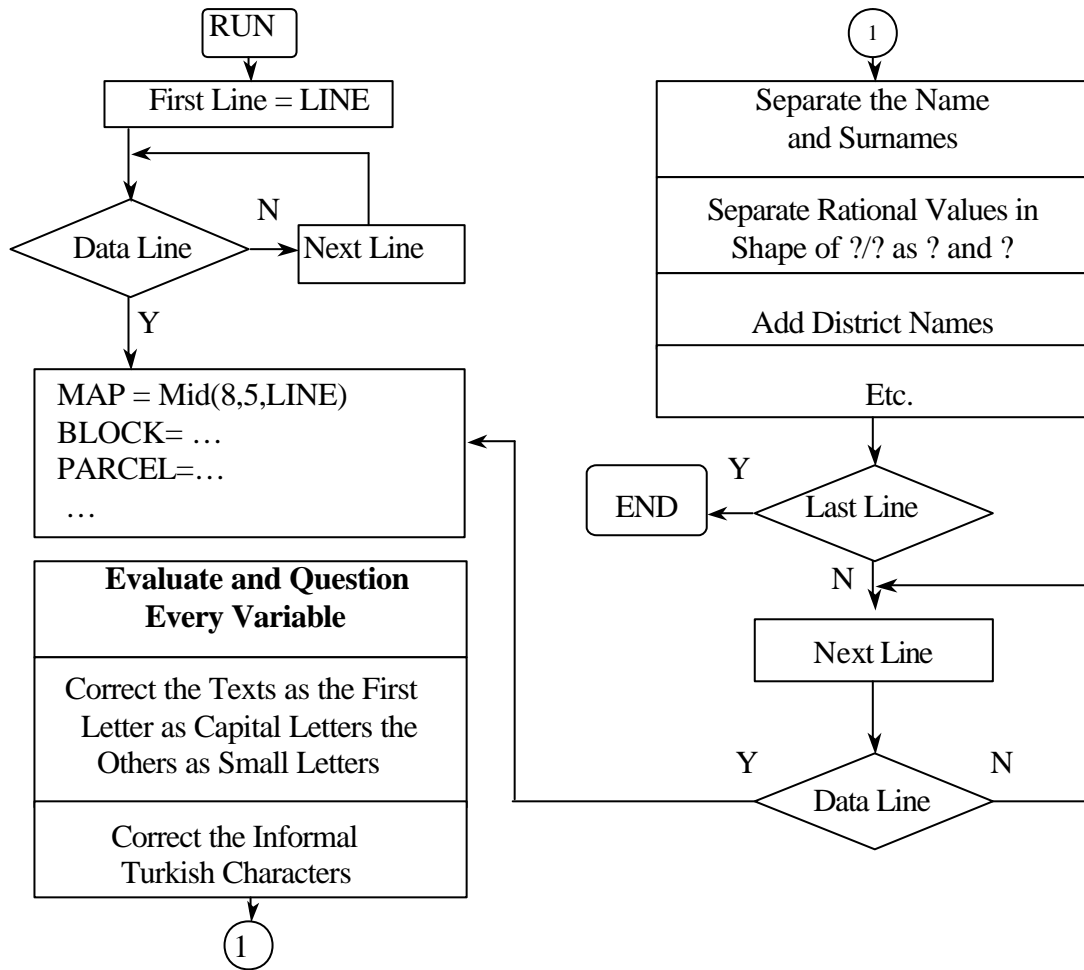


Figure 5: Flow chart of the program SFF for EGHAS text files

sıra	ypafta	yada	yparsel	yhispay	yhispayda	hisyuzolçui	yalan	adi	soyadi	babaadi
14348		4851	14	149	186	149	186	Adem	Tekir	İsmail
14349		4851	4	1	1	212	212	Ahmet	Karakaya	Hüseyin
14350		4940	1	75	3596	75	3596	Ahmet	Karakaya	Hüseyin
14351		4852	3	211	236	211	236	Ali	Durdu	Rasim
14352		4853	13	225	254	225	254	Ali	Şahin	İbrahim
14353		4853	14	223	304	223	304	Ali	Yiğit	Salih
14354		4851	6	163	436	163	436	Avni	Gürsoy	Müslüm
14355		4851	6	149	436	149	436	Avni	Gürsoy	Müslüm
14356		4853	12	218	219	218	219	Bekir	İlkun	Mehmet
14357		4851	13	1	1	182	182	Cemal	Malkoç	Hüseyin
14358		4988	1	54	1239	54	1239	Cemal	Malkoç	Hüseyin
14359		4851	2	143	213	143	213	Emine	Kılıç	Temel
14360		4851	8	1	1	326	326	Gazi	Şahin	Bayram
14361		4851	9	130	169	130	169	Gazi	Şahin	Bayram
14362		4851	11	109	181	109	181	Hamdi	Demir	Ali
14363		4851	7	228	387	228	387	Hasan	Tekir	İsmail
14364		4852	1	1	1	328	328	Hüseyin	Kaşık	Ali
14365		4862	10	63	158	63	158	Hüseyin	Kaşık	Ali
14366		4853	16	53	314	53	314	İsmail	Yılmaz	Hasan
14367		4853	15	284	537	284	537	İsmail	Yılmaz	Hasan
14368		4853	15	222	537	222	537	Mehmet	Toncu	Bilal

Figure 6: Sample of standard format in Access database

ŞUYULANDIRMA CETVELİ																					
DÜZENLEMEDEN ÖNCEKİ DURUM											DÜZENLEMEDEN SONRAKİ DURUM										
İLİ	SAYFA	CİNSİ	PAFTA	ADA	PARSEL	ALAN (M <sup>2</sup> )	HİSSE PAYI	HİSSE PAYIDA	TAPU BEDELI	D.Ö.P.	TAKSİT	ADİ	SOYAD	BABA ADI	PAFTA	ADA	PARSEL	ALAN (M <sup>2</sup> )	HİSSE PAYI	HİSSE PAYIDA	NOT/ÖZELLİKLER
36	3541	Arsa	18c1c	3908	1	314	TAM	314.00	0.00	314.00	Ahmet	Nas	Arif oğlu	G22B18c1c	103	1	314.00	314	5376		
36	3542	Arsa	18c1c	3908	2	241	TAM	241.00	0.00	241.00	Osman	Ayyıldız	Ahmet oğlu	G22B18c1c	103	1	5376.00	241	5376		
36	3543	Arsa	18c1c	3908	3	239	TAM	239.00	0.00	239.00	Osman	Ayyıldız	Ahmet oğlu	G22B18c1c	103	1	5376.00	239	5376		
36	3544	Arsa	18c1c	3908	4	239	TAM	239.00	0.00	239.00	Osman	Ayyıldız	Ahmet oğlu	G22B18c1c	103	1	5376.00	239	5376		
36	3545	Arsa	18c1c	3908	5	278	TAM	278.00	0.00	278.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00	278	5376		
36	3546	Arsa	18c1c	3908	6	259	118 259	118.00	0.00	118.00	Ali Haydar	Diyaroglu	Huseyin oğlu	G22B18c1c	103	1	5376.00	118	5376		
36	3546	Arsa	18c1c	3908	6	259	118 259	118.00	0.00	118.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00	118	5376		
36	3546	Arsa	18c1c	3908	6	259	23 259	23.00	0.00	23.00	Gebze Belediyesi			G22B18c1c	103	1	5376.00	23	5376		
36	3547	Arsa	18c1c	3908	7	262	TAM	262.00	0.00	262.00	Hanım	Usta	Mustafa kızı	G22B18c1c	103	1	5376.00	262	5376		
36	3548	Arsa	18c1c	3908	8	264	TAM	264.00	0.00	264.00	Mehmet	Usta	Mustafa oğlu	G22B18c1c	102	2	171.00	TAM			
36	3548	Arsa	18c1c	3908	8	264	TAM	264.00	0.00	264.00	Mehmet	Usta	Mustafa oğlu	G22B18c1c	103	1	5376.00	53	5376		
36	3549	Arsa	18c1c	3908	9	267	157 267	157.00	0.00	157.00	Abdurrahman	Gündüz		G22B18c1c	102	1	168.00	157	168		
36	3549	Arsa	18c1c	3908	9	267	110 267	110.00	0.00	110.00	Gebze Belediyesi			G22B18c1c	102	1	168.00	11	168		
36	3549	Arsa	18c1c	3908	9	267	110 267	110.00	0.00	110.00	Gebze Belediyesi			G22B18c1c	101	2	252.00	12	252		
36	3549	Arsa	18c1c	3908	9	267	110 267	110.00	0.00	110.00	Gebze Belediyesi			G22B18c1c	103	1	5376.00	97	5376		
36	3550	Arsa	18c1c	3908	10	269	TAM	269.00	0.00	269.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	102	3	269.00	TAM			
36	3551	Arsa	18c1c	3908	11	339	TAM	339.00	0.00	339.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	102	4	339.00	TAM			
36	3552	Arsa	18c1c	3909	1	346	TAM	346.00	0.00	346.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00	346	5376		
36	3553	Arsa	18c1c	3909	2	365	TAM	365.00	0.00	365.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00	365	5376		
36	3554	Arsa	18c1c	3909	3	313	243 313	243.00	0.00	243.00	Recep	Türkü		G22B18c1c	103	1	5376.00	243	5376		
36	3554	Arsa	18c1c	3909	3	313	70 313	70.00	0.00	70.00	Sami	Yilmaz	Sabri oğlu	G22B18c1c	103	1	5376.00	70	5376		
36	3555	Arsa	18c1c	3909	4	272	TAM	272.00	0.00	272.00	Nuran	Cömert		G22B18c1c	103	1	5376.00	272	5376		
36	3556	Arsa	18c1c	3909	5	259	TAM	259.00	0.00	259.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00	259	5376		
36	3557	Arsa	18c1c	3909	6	253	TAM	253.00	0.00	253.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00	253	5376		
36	3558	Arsa	18c1c	3909	7	246	22 246	22.00	0.00	22.00	Sami	Yilmaz	Sabri oğlu	G22B18c1c	103	1	5376.00	22	5376		

Figure 7: Sample of scanned image data

36	3541	Arsa	18c1c	3908	1	314	TAM	314.00	0.00	314.00	Ahmet	Nas	Arif oğlu	G22B18c1c	103	1	5376.00			
36	3542	Arsa	18c1c	3908	2	241	TAM	241.00	0.00	241.00	Osman	Ayyıldız	Ahmet oğlu	G22B18c1c	103	1	5376.00			
36	3543	Arsa	18c1c	3908	3	239	TAM	239.00	0.00	239.00	Osman	Ayyıldız	Ahmetoglu	G22B18c1c	103	1	5376.00			
36	3544	Arsa	18c1c	3908	4	239	TAM	239.00	0.00	239.00	Osman	Ayyıldız	Ahmetoglu	G22B18c1c	103	1	5376.00			
36	3545	Arsa	18c1c	3908	5	278	TAM	278.00	0.00	278.00	Kemal	Erdogan	Mehmetoglu	G22B18c1c	103	1	5376.00			
36	3546	Arsa	18c1c	3908	6	259	118 259	118.00	0.00	118.00	Ali Haydar	Diyaroglu	Huseyin oğlu	G22B18c1c	103	1	5376.00			
36	3546	Arsa	18c1c	3908	6	259	118 259	118.00	0.00	118.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00			
36	3546	Arsa	18c1c	3908	6	259	23 259	23.00	0.00	23.00	Gebze Belediyesi			G22B18c1c	103	1	5376.00			
36	3547	Arsa	18c1c	3905	7	262	TAM	262.00	0.00	262.00	Hanım	Usta	Mustafa kızı	G22B18c1c	103	1	5376.00			
36	3548	Arsa	18c1c	3908	8	264	TAM	264.00	0.00	264.00	Mehmet	Usta	Mustafa oğlu	G22B18c1c	102	2	171.00			
36	3548	Arsa	18c1c	3908	8	264	TAM	264.00	0.00	264.00	Mehmet	Usta	Mustafaoglu	G22B18c1c	103	1	5376.00			
36	3549	Arsa	18c1c	3908	9	267	157 267	157.00	0.00	157.00	Abdurrahman	Gündüz		G22B18c1c	102	1	168.00			
36	3549	Arsa	18c1c	3908	9	267	110 267	110.00	0.00	110.00	Gebze Belediyesi			G22B18c1c	102	1	168.00			
36	3549	Arsa	18c1c	3908	9	267	110 267	110.00	0.00	110.00	Gebze Belediyesi			G22B18c1c	101	2	252.00			
36	3549	Arsa	18c1c	3908	9	267	110 267	110.00	0.00	110.00	Gebze Belediyesi			G22B18c1c	103	1	5376.00			
36	3550	Arsa	18c1c	3908	10	269	TAM	269.00	0.00	269.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	102	3	269.00			
36	3551	Arsa	18c1c	3908	11	339	TAM	339.00	0.00	339.00	Kemal	Erdogan	Mehmetoglu	G22B18c1c	102	4	339.00			
36	3552	Arsa	18c1c	3909	1	346	TAM	346.00	0.00	346.00	Kemal	Erdogan	Mehmetoglu	G22B18c1c	103	1	5376.00			
36	3553	Arsa	18c1c	3909	2	365	TAM	365.00	0.00	365.00	Kemal	Erdogan	Mehmetoglu	G22B18c1c	103	1	5376.00			
36	3554	Arsa	18c1c	3909	3	313	243 313	243.00	0.00	243.00	Recep	Türkü		G22B18c1c	103	1	5376.00			
36	3554	Arsa	18c1c	3909	3	313	70 313	70.00	0.00	70.00	Sami	Yilmaz	Sabri oğlu	G22B18c1c	103	1	5376.00			
36	3555	Arsa	18c1c	3909	4	272	TAM	272.00	0.00	272.00	Nuran	Cömert		G22B18c1c	103	1	5376.00			
36	3556	Arsa	18c1c	3909	5	259	TAM	259.00	0.00	259.00	Kemal	Erdogan	Mehmetoglu	G22B18c1c	103	1	5376.00			
36	3557	Arsa	18c1c	3909	6	253	TAM	253.00	0.00	253.00	Kemal	Erdogan	Mehmet oğlu	G22B18c1c	103	1	5376.00			
36	3558	Arsa	18c1c	3909	7	246	22 246	22.00	0.00	22.00	Sami	Yilmaz	Sabri oğlu	G22B18c1c	103	1	5376.00			

Figure 8: Structure of edited text file of OCR processing

```

? UYULANDIRMA CETVEL - SAYFA: 1
UYGULAMAYA G'REN PARSEL N 3 KAYIT MAL - N N H'SSEYE S'ABET EDEN D'SZENLEME T E S C L E D L E N P A R S E L N 3 D S Z E N L E M E D E N D O L A Y I
SIRA PAFTA ADA PARSEL YSZ leSM H'SSE M'KTARI HESAP D'SZENLEME PAYI MAR MAR M'KTARI MAL N PARS PAYINA DS?EN (mi) FIY.TAKT TESCL ED LEN POTEK
NO NO NO NO MI ADI SOYADI BABA ADI ED LEN ORTAKLIK C'IKTIKTAN PAFTA ADA PAR H'SSES RAP.BR MI ALACAK BORE
629 49 11 56 10033 ADEM TEK R SMA L 687/30099 229 80 149 4851 14 186 149 186
633 49 11 56 10033 AHMET KARAKAYA H'SSEY N 1200/30099 400 139 261 4851 4 212 1/1
621 49 11 56 10033 AL DURDU RAS M 972/30099 324 113 211 4852 3 236 211/ 236
616 49 11 56 10033 AL PAH N BRAH M 1035/30099 345 120 225 4853 13 254 225 254
623 49 11 56 10033 AL Y T SAL H 1029/30099 343 120 223 4853 14 304 223 304
618 49 11 56 10033 AVN GRSOY M'SLEM 750/30099 250 87 163 4851 6 436 163 436
631 49 11 56 10033 AVN GRSOY M'SLEM 687/30099 229 80 149 4851 6 436 149 436
634 49 11 56 10033 BEK R LKUN MEHMET 1002/30099 334 116 218 4853 12 219 218 219
614 49 11 56 10033 CEMAL MALKOC H'SSEY N 1000/30099 333 116 217 4851 13 182 1/1
638 49 11 56 10033 EM NE KILIC TEMEL 657/30099 219 76 143 4851 2 213 143 213
626 49 11 56 10033 GAZ PAH N BAYRAM 2100/30099 700 244 456 4851 8 326 1/1
637 49 11 56 10033 HMD DEM R AL 500/30099 167 58 109 4851 11 181 109 181
617 49 11 56 10033 HAN TEK R SMA L 1050/30099 229 80 149 4851 7 287 229 287
622 49 11 56 10033 HSEY N KA?IK AL 1800/30099 600 209 391 4852 1 328 1/1
627 49 11 56 10033 SMA L YILMAZ HAN 1551/30099 517 180 337 4853 16 314 53 314
636 49 11 56 10033 MEHMET TOPEU B LAL 3000/30099 1000 349 651 4853 15 537 222 537
613 49 11 56 10033 MUSTAFA BSLBSL AL 3000/30099 1000 349 651 4903 4 250 1/1
615 49 11 56 10033 MUSTAFA KILIC ER F 300/30099 100 35 65 4904 10 239 65 239
620 49 11 56 10033 MUSTAFA KILIC ER F 357/30099 119 42 77 4904 10 239 77 239
639 49 11 56 10033 MER BIYIK MUSTAFA 725/30099 245 85 160 4851 9 170 160 170
624 49 11 56 10033 RECEP OKUR YUSUF 1320/30099 440 153 287 4853 11 190 119 190
YKLEN C KONTROL EDEN KONTROL EDEN TAD K
DENGE HAR TA LTD PT HAR TA EF BELED YE BA?KANI
Lyas SEKER Recep İverdi ?aban SARIAY Ahmet PENBEGSLLE
? UYULANDIRMA CETVEL - SAYFA: 2
UYGULAMAYA G'REN PARSEL N 3 KAYIT MAL - N N H'SSEYE S'ABET EDEN D'SZENLEME T E S C L E D L E N P A R S E L N 3 D S Z E N L E M E D E N D O L A Y I
SIRA PAFTA ADA PARSEL YSZ leSM H'SSE M'KTARI HESAP D'SZENLEME PAYI MAR MAR M'KTARI MAL N PARS PAYINA DS?EN (mi) FIY.TAKT TESCL ED LEN POTEK
NO NO NO NO MI ADI SOYADI BABA ADI ED LEN ORTAKLIK C'IKTIKTAN PAFTA ADA PAR H'SSES RAP.BR MI ALACAK BORE
625 49 11 56 10033 SADET DS MEO LU MEHMET 687/30099 229 80 149 4851 15 193 149 193
619 49 11 56 10033 SAL H ALAN MUHARREM 810/30099 270 94 176 4851 5 202 176 202
635 49 11 56 10033 SIDIK GELTEPE MUSTAFA 657/30099 219 76 143 4851 1 185 143 185
630 49 11 56 10033 ENEL EK NC MEHMET 687/30099 229 80 149 4851 12 190 149 190

```

Figure 9: A EGHAS text file consists of land registry information.



As mentioned earlier, there is a need for converting characters of printed pages to a text file. Therefore, application of OCR methods should be applied at first. Due to this necessity, a raster image of this page was obtained by scanning it, and then a text file of the page was obtained using a OCR programme- "Recognita Plus". Figure 8 shows structure of a text file output of OCR processing.

The programme SFF can be applied to these two text files separately, and converts the data into standard format in Access's database as seen in Figure 6.

Followings are some of the editing work carried out by SFF.

- a) Letter editing: Here first characters and the others corrected as capital and small letters respectively.
- b) Word separation: Data in the form of "Name Surname" is separated as "Name" and "Surname"
- c) Correction of informal Turkish characters.
- d) Addition of external data: Data such as district names are added to appropriate data lines, etc.

## **5. CONSLUSION**

With development of CFT, SFF or similar programmes, it was shown that graphical and non-graphical data obtained from text files of CADD applications, and obtained from printed documents, can be transformed to standard data formats which are suitable for GIS.

It is also shown that there is no need for reproduction of existing data to be used in GIS. Thus, data which is the most important factor for defining the production cost of GIS can be drawn back to low levels.

## **REFERENCES**

Cömert, Ç., Ulusal Konumsal Veri Altyapisi İçin Veri Değişim Standartinin Belirlenmesi, p.3, Doctorate Thesis, KTÜ, Trabzon (1996)

Cömert, Ç., Geleneksel Konumsal Veri Değişiminin Sorunları ve FME yazılımının Sunduğu Yeni Olanaklar, Harita Dergisi, No 122, p.11 (1999)

Öztürk, A., Osmanlıca Karakterlerin Bilgisayar Destekli Tanınması, p. 42, Master Thesis, GYTE, Gebze (1998)