CO-052

OPTIMIZATION OF GEOGRAPHIC DATA USING ENTROPY

BILGI S., IPBUKER C.
Istanbul technical university, ISTANBUL, TURKEY

ABSTRACT

Maps have been used as a medium for presenting the geographical information. There are many criteria for a "good" map; one of these criteria is information content. As the map is regarded as a communication tool, cartographers are interested in the effectiveness of map design and the information content of a map. Analyzing the intensity of topographic objects in a spatial information system is a considerable factor in gathering data during the installation progress and defining the optimal data. Optimal data is defined as the minimum necessary data for the aim and usage of the system which does not cause complexity on the screen. Thus, data excess and time-wasting are prevented by gathering optimal data in spatial information systems. Entropy theorem is used for measuring the statistical information. Although entropy is the second law of thermodynamics, it was introduced in the quantification of information. Entropy briefly means the measurement of indefiniteness of a system. Entropy has been adopted by Sukhov as a concept for cartographic communication but only the number of each type of symbol represented on a map is taken into account. Map entropy analysis can give a statistical evidence for determining the object resolution of a geographical information system during establishment and design. But comparing with the concept of scale on analog maps, the usefulness of such a measure for map content is still controversy. In this study, a map entropy analysis is processed on a spatial information system called FULBIS. An evaluation is also performed due to the object resolution on raster image maps. The statistical evaluation results are discussed in order to create a measuring tool for the optimization of geographic data behind a spatial information system.

KEYWORDS

Entropy, Map Entropy Analysis, Optimization, Object Resolution in Spatial Information Systems

INTRODUCTION

There are four components of Geographical Information Systems (GIS). These include data, hardware, software, and staff. Data is the most important and expensive component. Hardware and software constitute approximately 10% - 20% of the overall budget, while staff costs comprise 10%; however, the budget for data consists of nearly 70% - 80% of the total system budget (Ekincioğlu, 1998).

Object catalogue is the conceptual data model of a GIS and defines the frame of data which would be gathered for the system. The object is everything that can be monitored and has a historicity or a possibility of existence in space. An object with a spatial reference should thus satisfy a planimetric geometry, topological relations and attributes. Analyzing the intensity of topographic objects in a GIS is a considerable step in gathering data during the installation progress and defining the "optimal" data. Optimal data is defined as the minimum data needed for the aim and usage of the system which does not cause complexity on the screen (Bilgi et.al.2008).

Maps have been created and used for centuries as abstractions of the real world, and as such are capable of giving us a picture of our environment which distils its full complexity into an effective graphical rendering. By considering its general nature, the term 'complexity' can be defined as applicable to the description of cartographic representations (maps). Attempts have thus been made to measure complexity through practical testing (Fairbairn, 2006).

Interest in map information dates back to the late 1960s following the publication of work on quantitative measures of information by Shannon (1948). This work has come to be known as 'information theory,' and was also applied to communication theory. Pioneering work in quantitative measurement of map information was done by Sukhov (1967, 1970), who considered the statistics of different types of symbols represented on a map (Li and Huang, 2002). For designers, creating a conceptual data model is one of the challenging tasks of GIS. The main problem for such an automation task is how to quantify and formalize the properties of the process as an information system (Bjørke, 1996). Thus, quantifying the efficiency and information content of the base map used by the information system is of paramount importance. However, there remains no metric measure for quantifying the cartographic efficiency of a map (Knopfli 1983). As such, Shannon's theory based on entropy may be a useful measure by which to optimize a map's informational content. In this study, an evaluation of the FULBIS project object catalogue is processed and its information content analyzed using Shannon's theory.

FULYA INFORMATION SYSTEM (FULBIS)

FULBIS is an abbreviation of Fulya Bilgi Sistemi (Fulya Information System) project, developed by Cengizhan Ipbuker and Serdar Bilgi from Istanbul Technical University (ITU) Department of Geomatics Engineering.

The project is a GIS-based information system which is designed to provide the accurate and consistent data about Fulya quarter in Sisli county of Istanbul city. It is going to be used for planning and management of the quarter by governor. An object catalogue prepared for the FULBIS project which consists of object domains, object groups, object types, lump objects and attributes. Object domains are settlement, administrative regions, transportation structures and infrastructure, running and dead waters, plant cover, geodesic reference points, topography and cadastral domains. Object group is a concept covers same type objects. There can be a single object domain that the object group belongs. Object type is the information about topological relationships, attributes, definition and geometry of the objects. The amount of statistical information content of the object catalogue and topographic maps of the project are measured using entropy and the results are commented to determine the optimal data.

"Object piece" is defined as the subpart of an object, which has its own geometry and attributes. Attributes consist of qualitative or quantitative information about an object or an object piece. Attribute values of some object types are also given in the object catalogue. Data types for the attributes are number, real, integer, logical (and, or etc.), boolean (true or false), string, binary, array, date, and currency (Sahin et al., 2002). A sample of the FULBIS object catalogue is shown in table 1.

Table 1: FULBIS object catalogue for building object type

FUKBS Object Catalogue				
A. Topographic Object Domain: Catalogue		**Object Group:**		
Code:	**Name:**	**Code:**	**Name:**	
A.A.	**Settlement Area**	**A.A.4.**	**Buildings**	
Object Type:				
Code:	**Name:**			
A.A.4.01	**Building**			
Description: A structure for living and other vital activities. Every building will be modeled as an object with the base of a door number. **Spatial Reference Type: Point, Area** **Name:**				
Abbreviation	**Name (Data Type: string, Cardinality: 1:1)**			
NAME	**Spatial Name (Formal Name)**			
Attributes:				
Abbreviation	**Code**	**Name**	**Cardinality**	**Data Type**
ENK		**Usage Type**	**1:?**	**integer**
	701	**Administration**		
	506	**Theatre**		
	507	**Opera**		
	...	**...**		
ENB		**Usage Type for Basement**	**1:?**	**integer**
	101	**Residence**		
	102	**Commercial**		
	105	**Commercial+Residence**		
	...	**...**		
ENZ		**Usage Type for Ground Floor**	**1:?**	**integer**
	101	**Residence**		
	102	**Commercial**		
	105	**Commercial+Residence**		
	...	**...**		
ENU		**Usage Type for Flats**	**1:?**	**integer**
	101	**Residence**		
	102	**Commercial**		
	105	**Commercial + Residence**		
	...	**...**		
DKS		**Number of Filled Houses**	**1:1**	**integer**
BKS		**Number of Vacant Houses**	**1:1**	**integer**
DIS		**Number of Filled Offices**	**1:1**	**integer**
BIS		**Number of Vacant Offices**	**1:1**	**integer**
YGM		**Fire Escape**	**1:1**	**integer**
OAB		**Front/Back Yard**	**1:1**	**integer**
YCS		**Structure Type**	**1:1**	**integer**
KSO		**Number of Floor (Front Side)**	**1:1**	**integer**
KSA		**Number of Floor (Back Side)**	**1:1**	**integer**

STATISTICAL INFORMATION OF A MAP: ENTROPY ANALYSIS

Entropy is the second law of thermodynamics and, briefly defined, represents the measurement of the indefiniteness of a system (www.biltek.tubitak.gov.tr). This measurement does have an explicit link to information content and is associated with attempts to quantify information transfer within a communication system (Fairbairn, 2006).

Sukhov (1967, 1970) has adopted the entropy concept for cartographic communication, but only took into account the number of each type of symbol represented on a map. Thus, Let N be the total number of symbols on a map, M the number of symbol types and Ki the number of symbols for the each type. The probability for each type of symbol on the map is then as follows:

(5.1)

The probability for each type of symbol on the map is then as follows:
(5.2)
Pi is the probability for the ith symbol type, i=1, 2, . . . M.
The map entropy analysis requires the calculation of the following entropy quantities:
(5.3)

STUDY AREA AND DATA USED

The study was done in the Fulya quarter in Sisli county of Istanbul city in Turkey, as shown in Figure 1. Fulya is one of 28 quarters in Sisli County. There are;

- 431 buildings in the quarter. These buildings are used as business offices and shops (2443) and as houses (8146),
- 6 main and 42 streets,
- 10 hotels,
- 4 educational institutes (primary, high schools and university),
- 6 park lands,
- 2 mosques,
- 1 stadium

Population of Fulya quarter is 18.373. Its area is 830.000 m² (www.sisli.bel.tr).

The object catalogue and vector maps of the FULBIS Project are used in the study. The project uses a FULBIS object catalogue as seen in table 1. The maps are taken from the municipality of Istanbul in AutoCAD (.dxf) format. The dxf (Drawing Exchange Format) is a CAD (Computer Aided Design) data file format for enabling data interoperability between AutoCAD and other programs. Attribute data is gathered both from the municipality and the project staff. The data consists of five types of objects: building, cadastral land, road, quarter and helicopter runway objects. All the data are matched with the object catalogue and use its coding system in the attribute table.

The topographical data belonging to Fulya Quarter in vector format is shown in figure 1. The study data consists of eight types of objects, and 502 objects in total as seen in table 2.

Figure 1 : Fulya Quarter

A map is designed in the project for the billboards in Fulya for inquiring of an address inside the quarter boundaries (Figure 2). AutoCAD2010 and Macromedia Flash MX2004 software are used for this process. The vector data of the quarter is converted to .wmf file in AutoCAD software and Flash Macromedia MX2004 software is used for designing the end map.

Same design form is used in ArcGIS-ArcMap10 GIS software for Fulya Information System. All attributes are entered to the database of ArcMap software. A view of project on GIS software is given at figure 3.

Table 2: Object types and number of objects for each type

Object Type	Object Numbers
Building	431
Main street	6
Street	42
Hotel	10
Educational institute	4
Park land	6
Mosque	2
Stadium	1
Total Objects	**502**

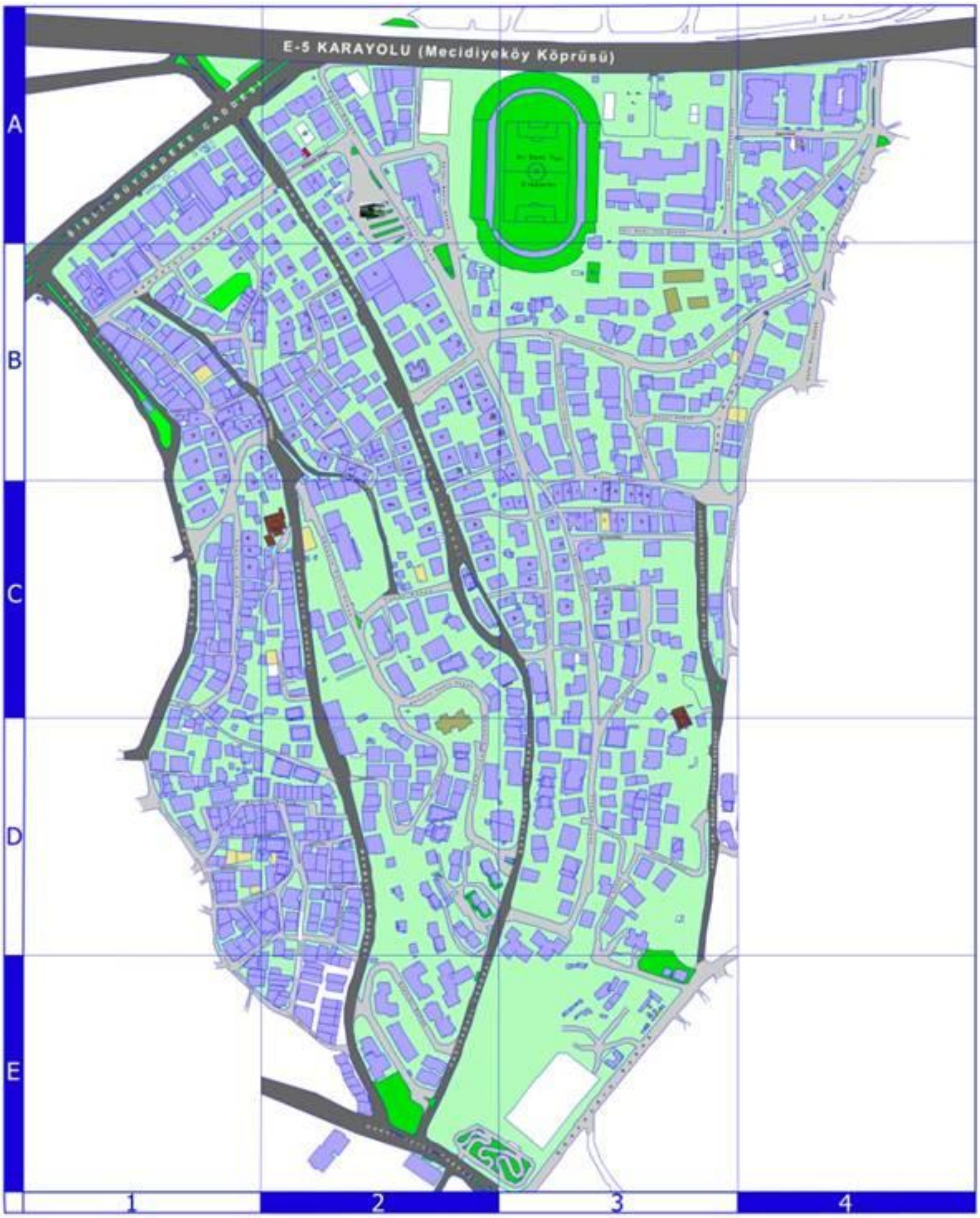

Figure 2 : Fulya Quarter map for billboards

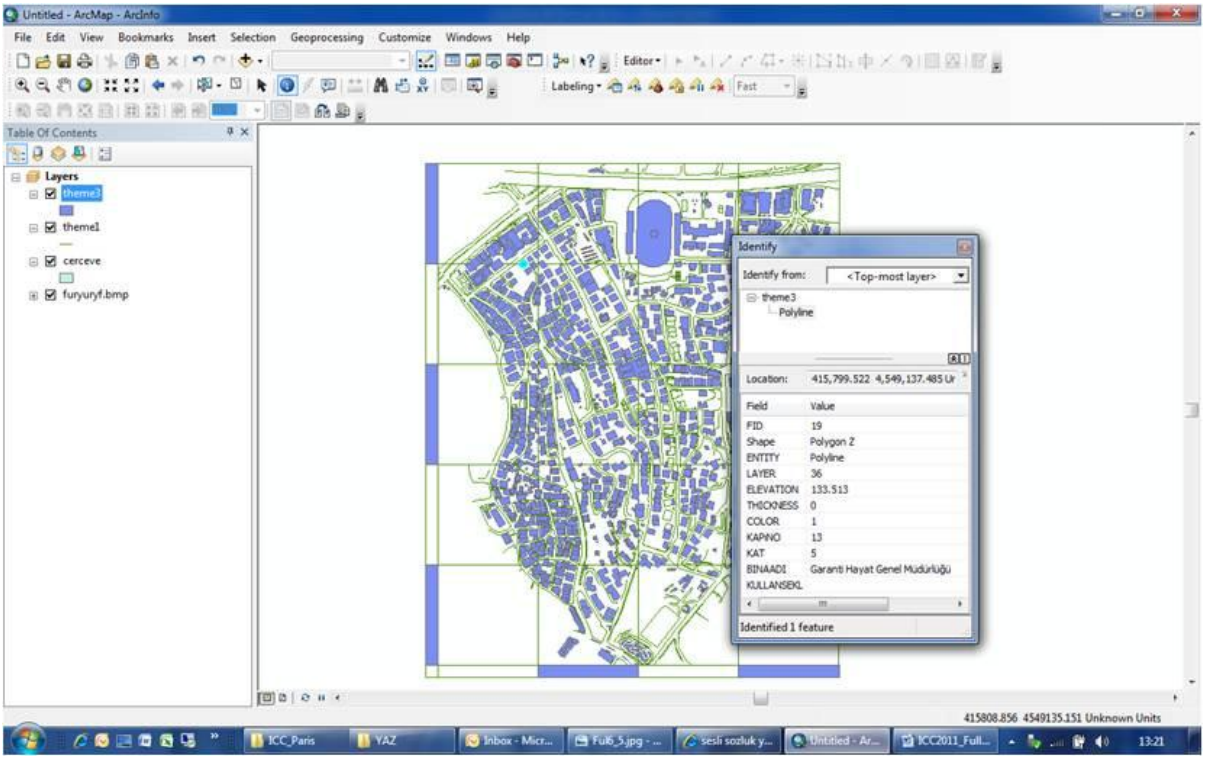

Figure 3 : FULBIS Project on ArcMAP GIS software

MAP ENTROPY ANALYSIS TYPES

Only the number of objects for each type represented on the map is taken into account when calculating the entropy. There are eight types of objects and 502 objects in total. Entropy results are calculated using some variations of the arbitrarily chosen object types, and using the equations (5.1), (5.2) and (5.3);

N is total object number for all object types and calculated using equation (5.1)

N= 431+6+42+10+4+6+2+1=502

The probability of each type of objects on the map (Pi) is calculated using equation (5.2)

P1(building)= P2(Main Street) = P3(Street)= P4(Hotel)=

P5(Educational Ins.)= P6(Park Land)= P7(Mosque)= P8(Stadium)=

The map entropy is calculated using equation (5.3) and the result is ≈0.60 (Table 3).

Table 3 : Entropy for All Object Types

Object Type	Object Number	Probability		LN(Pi)	Entropy
Building	431	431/502	0.85856574	-0.15249203	-0.130924432
Main street	6	6/502	0.01195219	-4.42684065	-0.052910446
Street	42	42/502	0.08366534	-2.4809305	-0.207567891
Hotel	10	10/502	0.01992032	-3.91601503	-0.078008267
Educational institute	4	4/502	0.00796813	-4.83230576	-0.038504428
Park land	6	6/502	0.01195219	-4.42684065	-0.052910446
Mosque	2	2/502	0.00398406	-5.52545294	-0.022013757
Stadium	1	1/502	0.00199203	-6.21860012	-0.01238765
TOTAL	502	1	1	-31.9794777	-0.595227317

Entropy for Building, Main Street, Street Object Types

The probability of each type of objects on the map (Pi) is calculated using equation (5.2)

P1(building)= P2(Main Street) = P3(Street)=

The map entropy is calculated using equation (5.3) and the result is ≈0.36 (Table 4).

Table 4 : Entropy for Building, Main Street, Street Object Types

Object Type	Object Number	Probability		LN(Pi)	Entropy
Building	431	431/479	0.899791232	-0.105592507	-0.095011212
Main street	6	6/479	0.012526096	-4.379941128	-0.054863563
Street	42	42/479	0.087682672	-2.434030979	-0.213422341
TOTAL	479	1	1	-6.919564615	-0.363297116

Entropy for Hotel, Educational Institute, Park Land, Mosque and Stadium Object Types
The probability of each type of objects on the map (Pi) is calculated using equation (5.2)
P1(Hotel)= P2(Educational Ins.)= P3(Park Land)= P4(Mosque)= P5(Stadium)=
The map entropy is calculated using equation (5.3) and the result is ≈1.37 (Table 5).
Table 5 : Entropy for Hotel, Educational Institute, Park Land, Mosque and Stadium Object Types

Object Type	Object Number	Probability		LN(Pi)	Entropy
Hotel	10	10/23	0.434782609	-0.832909123	-0.362134401
Educational institute	4	4/23	0.173913043	-1.749199855	-0.30420867
Park land	6	6/23	0.260869565	-1.343734747	-0.350539499
Mosque	2	2/23	0.086956522	-2.442347035	-0.212378003
Stadium	1	1/23	0.043478261	-3.135494216	-0.136325835
TOTAL	23	1	1	-9.503684976	-1.365586409

Entropy for Building, Main Street, Educational Institute and Stadium Object Types
The probability of each type of objects on the map (Pi) is calculated using equation (5.2)
P1(building)= P2(Main Street) = P3(Educational Ins.)= P4(Stadium)=
The map entropy is calculated using equation (5.3) and the result is ≈0.14 (Table 6).
Table 6 : Entropy for All Object Types

Object Type	Object Number	Probability		LN(Pi)	Entropy
Building	431	431/442	0.975113122	-0.025201792	-0.024574598
Main street	6	6/442	0.013574661	-4.299550413	-0.058364938
Educational Institute	4	4/442	0.009049774	-4.705015521	-0.042579326
Stadium	1	1/442	0.002262443	-6.091309882	-0.013781244
TOTAL	442	1	1	-15.12107761	-0.139300106

Entropy for Building, Main Street and Stadium Object Types
The probability of each type of objects on the map (Pi) is calculated using equation (5.2)
P1(building)= P2(Main Street) = P3(Stadium)=
The map entropy is calculated using equation (5.3) and the result is ≈0.09 (Table 7).
Table 7 : Entropy for Building, Main Street and Stadium Object Types

Object Type	Object Number	Probability		LN(Pi)	Entropy
Building	431	431/438	0.984018265	-0.01611082	-0.015853341
Main street	6	6/438	0.01369863	-4.290459441	-0.058773417
Stadium	1	1/438	0.002283105	-6.08221891	-0.013886345
TOTAL	438	1	1	-10.38878917	-0.088513103

The results of the applications using different object types are summarized on table 8.
Table 8: Number of the objects for each type and entropy of maps

Build.	P	Main Str.	P	Str.	P	Hotel	P	Edu. Ins.	P	Park Land	P	Mosq.	P	Stad.	P	Total Obj.	Entropy
431	0.86	6	0.01	42	0.08	10	0.02	4	0.01	6	0.01	2	0.00	1	0.00	502	**0.60**
431	0.90	6	0.01	42	0.09											479	**0.36**
						10	0.43	4	0.17	6	0.26	2	0.09	1	0.04	23	**1.37**
431	0.98	6	0.01					4	0.01					1	0.00	442	**0.14**
431	0.98	6	0.01											1	0.00	438	**0.09**

CONCLUSION

Entropy computed in this way only takes into account the number of symbols for each type. The result is relevant with existence probability for each type of object through the total number of objects. When the value of X (P1, P2, …) is certain, Pi=1, then H(X)=0. H(X) is at its maximum when all probabilities for each type of object have equal values (Li and Huang, 2002). As seen in table 8, entropy is high when the probability of the objects for each type is near to each other.

Map entropy analyses using the conceptual data model can produce different quantitative measures for interpreting information efficiency during the map design process in information systems. These measures may give some statistical evidence for determining the object resolution of digital cartographic products.

REFERENCES

Bilgi,S.,Ipbuker,C.,Ucar,D.and M.Sahin (2008), "Map Entropy Analysis of Topographic Data Used in Disaster Information Systems", Journal of Earthquake Engineering, 12(2),13-36

Bjørke, J. T., (1996), "Framework for entropy-based map evaluation," Cartography and Geographical Information Systems, 23, 78–95.

Ekincioğlu, İ. (1998), Konumsal Bilgi Sistemlerinin Kurumlar Arası Kullanılması, (PhD Thesis, YTU Institute of Science and Technology).

Fairbairn D., (2006), "Measuring Map Complexity," The Cartographic Journal, 43(3), 224-238.

Knopfli, R., (1983), Communication theory and generalization. In Graphic Communication and Design in Contemporary Cartography, edited by D. R. F. Taylor (New York, Chichester: John Wiley & Sons Ltd), pp. 177–218.

Li, Z. and Huang, P., (2002), "Quantitative measures for spatial information of maps," International Journal of Geographical Information Science, 16(7), 699 – 709.

Sahin, M., Uçar, D., Baykal, O., Türkoğlu, H., Tarı, E., İpbüker, C., Musaoğlu, N., Göksel, Ç., Coşkun, M.Z., Kaya, Ş., Yiğiter, R., Erden, T., Karaman, H., Yavaşoğlu, H., Bilgi, S., Üstün, B. (2002), "FULBIS: Development of a National Database Using Geographical Information Systems and Remote Sensing System and Standards for a Disaster Management Decision Support System Project," I.T.U., Istanbul.

Shannon, C. E., (1948), "A mathematical theory of communication," The Bell System Technical Journal, 27, 379–423 & 623–656.

Sukhov, V. I., (1967), "Information capacity of a map entropy," Geodesy and Aerophotography, X, 212–215.

Sukhov, V. I., (1970), "Application of information theory in generalization of map contents," International Yearbook of Cartography, X, 41–47.

http://www..sisli.bel.tr (Visited in February 2011).