NIKO JELUŠIĆ, Ph.D.
E-mail: niko.jelusic@fpz.hr
MARIO ANŽEK, Ph.D.
E-mail: mario.anzek@fpz.hr
BOŽIDAR IVANKOVIĆ, M.Sc.
E-mail: bozidar.ivankovic@fpz.hr
University of Zagreb,
Faculty of Transport and Traffic Sciences
Vukelićeva 4, HR-10000 Zagreb, Republic of Croatia

Intelligent Transport Systems (ITS)
Preliminary Communication
Accepted: July 16, 2009
Approved: Mar. 19, 2010

INFORMATION SOURCE QUALITY IN INTELLIGENT TRANSPORT SYSTEMS

ABSTRACT

Advanced automatic traffic control systems and various other ITS (Intelligent Transport Systems) applications and services rely on real-time information from the traffic system. This paper presents the overview and general functions of different information sources which provide real-time information that are used or could be used in ITS. The objective is to formally define the quality of information sources suitable for ITS based on formal models of the traffic system and information sources. The definition of quality encompasses these essential factors: traffic system information that exists or may be requested, user requirements and attributes that describe the information sources. This provides the framework and guidelines for the evaluation of information sources that accounts for relevant factors that influence their selection for specific ITS applications.

KEY WORDS

information source, information source quality, Intelligent Transport Systems (ITS), automatic traffic control

1. INTRODUCTION

Efficient operation of ITS depends on the quality of the information source. Here, the automatic traffic control (ATC) systems are highlighted as extremely important regarding their function, and they rely on the traffic flow sensors as sources of real-time information. In principle the sources of information are specialized, so that every type of information source collects data for the respective user. The availability and sharing of information from the traffic system with the application of information and communication technology can enhance the overall efficiency of ITS. Regarding the choice of the information source, the users should express their requirements regarding information quality. When defining the information source quality, the user or a group of users should be provided with a framework and guidelines for the evaluation of the information source that encompass all relevant factors.

There are a number of publications and papers on the application and evaluation of information sources, especially of the traffic flow sensors for ITS. Detailed presentation of ITS data requirements and applications of sensor technologies is provided in [1]. In [2] and [3] the detectors per single criteria are evaluated, and the procedure of selecting the detectors based on the elimination of the unsuitable ones is provided in [3]. There are also papers from other fields of science that deal in general with information quality and that provide the definition of information quality [4,5]. However, the information source quality for ITS application has not been satisfactory defined.

The paper presents the function of information source and gives a brief overview of the possible information sources. The formal models of the traffic space and information source have been set, and in the end, based on these models the quality of the information source has been defined.

2. SIGNIFICANCE AND TYPE OF INFORMATION SOURCE

The basic function of the information source is to collect data on traffic variables that describe the current condition in the traffic system. Based on the collected data in the traffic control system, a series of processes is performed to provide commands to executive control devices, and to also extract information useful to various ITS services. *Figure 1* presents the basic processes related to the processing and usage of information that form the information cycle [6]. The condition in the traffic system changes either by the action of the executive devices of the automatic control system or by informing the traffic participants

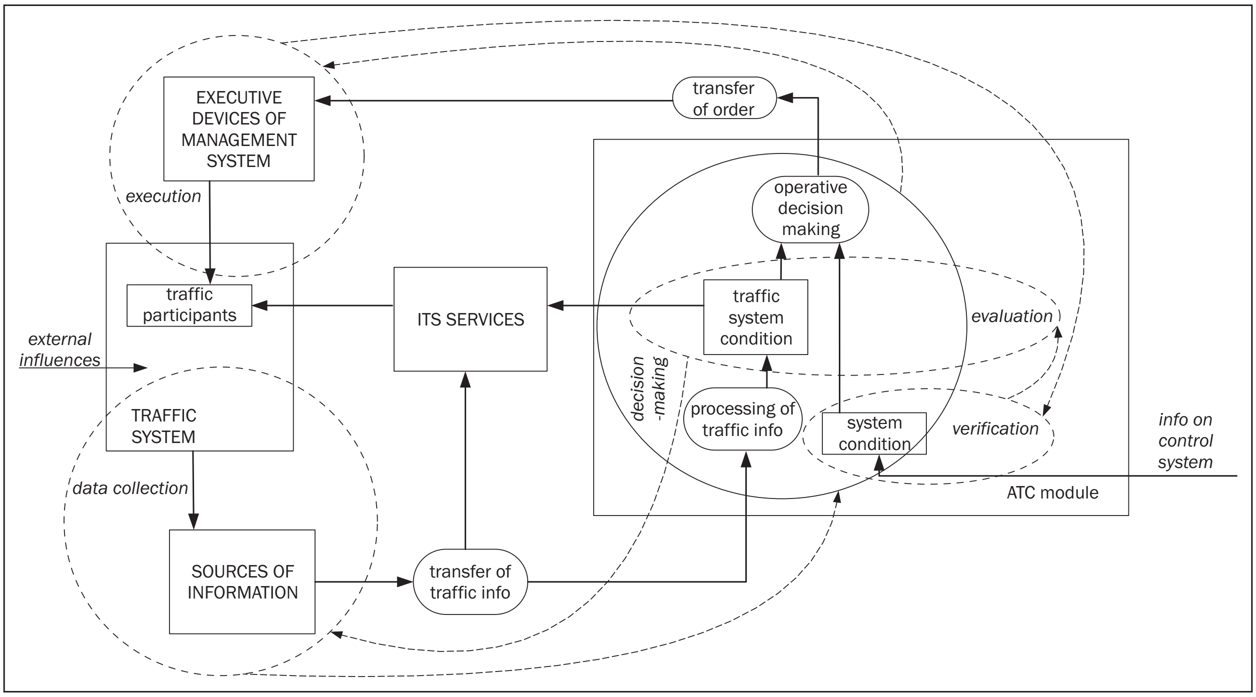

Figure 1 - Information cycle processes related to the collection, processing and usage of information from the traffic system

using ITS services. This makes the importance of high-quality information source obvious.

For the purpose of collecting information the traffic system has been divided into traffic flow, traffic infrastructure, control infrastructure and traffic infrastructure environment. *Table 1* lists these traffic system components and typical sources that contribute information concerning their status and reports.

The presented classification recognizes two different methods of data collection: automatic and reports. Automatic data collection encompasses the usage of traffic flow sensors, video surveillance, satellite and aircraft recording, sensors on traffic entities, positioning technologies and communication technologies, infrastructure sensors and automatic reports about the condition of the control infrastructure and meteorological sensors. For the needs of modern ITS, especially in the area of automatic traffic control, traffic flow sensors are of highest importance, contributing fundamental space and time-continuous data. The reports are pieces of information obtained from the perspective of the traffic participants such as reports on accidents, injuries, and requests for action of special services.

3. INFORMATION SPACE MODEL OF THE TRAFFIC SYSTEM

Information space model serves to define the quality of information source. Because all observations are maintained in the predetermined part of space and

Table 1 - Sources of information

Sources of information	Traffic flow	Traffic infrastructure	Control infrastructure	Environment
Traffic flow sensors	X			
Video surveillance	X	X	X	X
Satellite and aircraft recording	X	X		X
Sensors on traffic entities – mobile observers	X	X		X
Technologies of positioning and communication	X			
Reports of special services	X	X	X	X
Reports of traffic participants	X	X	X	X
Infrastructure sensors		X		
Automatic reports on the condition of control infrastructure			X	
Meteorological sensors				X

X – can collect information

period of time, the mathematical formalization is presented.

Traffic area $P \subset \mathbb{R}^3 \times \mathbb{R}$ consists of the points from environment $\mathbb{R}^3$ being observed in the moment $t \in \mathbb{R}$. Each traffic system in the traffic area *P* is given by its dynamic components, so that the parameters that describe its condition depend on the position in the traffic system and on the moment in time. The possibilities of the sources of information are limited, so that in order to describe the traffic system an ordered pair is introduced $(gp,t) \in P$, which represents space-time scope (space-time area) referred to by the variable values.

The first component $gp \in \mathbb{R}^3$ of the ordered pair (gp,t) presents the space scope of the traffic area *P*. The scope can be defined by the coordinates, and it also may refer to the cross-section, section, area or any geometrical form for which the variables are defined which then present the condition of a certain traffic layer. The second component $t \in \mathbb{R}$ (time scope) is a time coordinate. Time coordinate can be an assigned time interval $\Delta t \subset \mathbb{R}$ within which the variables of interest are studied. Depending on the given initial moment, the time component can contain negative values, as well.

Traffic system is completely set by the traffic variables p_i. The variables can be directly measurable values or values that can be calculated based on the measured data. The task of traffic variables is the presentation of the traffic area condition. The variables p_i describe the condition of the traffic area *P* by real values

$$p_i : P \rightarrow \mathbb{R}, i = 1,2, \ldots, k. \tag{3.1}$$

The set of all variables determining the traffic system is denoted by *PP*. The values of variables, apart from the numerical values may be also logical or descriptive qualitative values. The definition and the number of variables are based on the theoretical and scientific knowledge, so that the set of variables *PP* can be increased over time. However, it is assumed that *PP* is finite

$$P = \{p_i, i = 1, \ldots, k\}. \tag{3.2}$$

Accordingly, one defines Eq. (3.3) for the space-time scope (gp,t) which is attached by a set of variables *PP* with a finite number of elements p_i, whose actual values define the condition of the traffic space in $(gp,t) \in P$. Thus,

$$\forall (gp,t) \in P, \ \ \exists P \quad (gp,t) = \{p_i(gp,t) \mid i = 1,2,\ldots,k\} \tag{3.3}$$

In the spatial domain, from the aspect of the variables defined on a single element $(gp,t) \in P$ and the possibilities of collecting and using information, vertical stratification can be done. It divides the set of observed variables *PP* into the following layers: environment, traffic infrastructure, control infrastructure, and traffic flow.

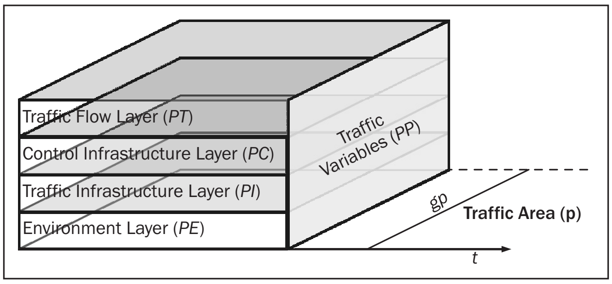

Figure 2 - Information space model layers of the traffic system

The layer of the traffic flow contains a set of traffic entities that change position in time and space. Therefore, the traffic flow is determined by the spatial and time coordinates of the traffic entities, and the traffic flow condition is described by a series of traffic flow variables. The environment layer encompasses the characteristics of space that set requests on the traffic infrastructure and on the influences of the weather conditions (humidity, temperature, rain, fog, ice). The traffic infrastructure layer determines the spatial and time macro-limits of the traffic entities coordinates. The control infrastructure layer manages the spatial and time micro-limits of the coordinates of the traffic entities. The dynamic traffic variables thus describe the condition in individual layers depending on each other.

The condition of every layer of the traffic space model presented in *Figure 2* is described by the respective set of variables. The traffic flow has been described by the set of variables PT, the traffic control infrastructure by the set PC, the traffic infrastructure by the set PI, and the traffic environment by the set PE. The union of all the sets of layer variables is the set PP, and their cross-section is an empty set. Accordingly,

$$PP = PT \cup PC \cup PI \cup PE \tag{3.3}$$

$$PT \cap PC \cap PI \cap PE = \varnothing \tag{3.4}$$

Expressions (3.2) and (3.3) hold for every layer. For instance, for the traffic flow layer expression (3.6) holds: within the space-time scope defined by the ordered pair (gp,t) there is a set of variables *PT* with finite number of elements pt_i, which define the condition of the traffic flow

$$\forall (gp,t) \subset P \quad \exists PT = \{pt_i(gp,t) \mid i = 1,2,\ldots,n\} \tag{3.6}$$

where pt_i is the traffic flow variable.

The same holds also for other layers, so that further in the text only formalized records are given: expression (3.7) for the control infrastructure layer, expression (3.8) for the traffic infrastructure layer and expression (3.9) for the traffic environment layer.

$$\forall (gp,t) \subset P \quad \exists PC = \{pc_i, i = 1,2,\ldots,m\} \tag{3.7}$$

$$\forall (gp,t) \subset P \quad \exists PI = \{pi_i, i = 1,2,\ldots,k\} \tag{3.8}$$

$$\forall (gp,t) \subset P \quad \exists PE = \{pe_i, i = 1,2,\ldots,h\} \tag{3.9}$$

where:

pc_i – traffic control infrastructure variable,

pi_i – traffic infrastructure variable,

pe_i – traffic environment variable,

$m,k,h \in N$ – natural numbers that suggest the finite sets.

Additional assumption for this model is restriction of the traffic area and non-continuous time scope that could be observed by some kind of sensors. Thus, for the complete presentation of the entire traffic area P by variables, either for a certain time coordinate or within time interval t, a finite spatial gp is necessary, so P is given by

$$P = \bigcup_{i=1}^{n}(gp_i, t_i) \text{ or } P = \bigcup_{i=1}^{n}(gp_i, \Delta t_i). \tag{3.10}$$

Consequently, the traffic area $P \subseteq \mathbb{R}^3 \times \mathbb{R}$ is a subset of the Cartesian product of 3-dimensional space and time axis. There is an information source function defined as the vector mapping by expression (3.11):

$$D: \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^n \quad \text{given by}$$
$$D(gp,t) = (p_1(gp,t), \ldots, p_n(gp,t)). \tag{3.11}$$

4. INFORMATION SOURCE MODEL OF THE TRAFFIC SYSTEM

The actual sources of information do not present the completely actual traffic condition, so there is a transfer function of the source of information $T(p)$. Because of various factors and disturbances, the transfer function changes the actual, unknown condition of reality described in *Figure 3*. The transfer function $T(p)$ assigns each traffic variable values from the set of real numbers, $T: PP \to R$. The transfer function domain is naturally the set of traffic variables $PP = \{p_i,\ i = 1, \ldots, n\} = PT \cup PC \cup PI \cup PE$.

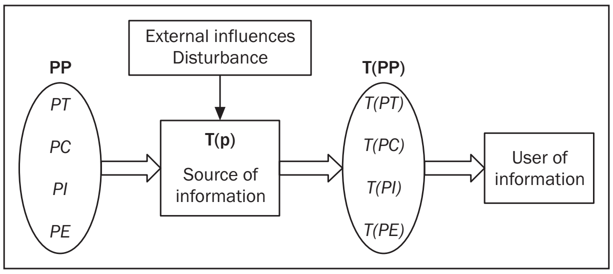

Figure 3 - Information source model

The transfer function of a converter (e.g. sensor) is considered as the functional dependence of inputs (stimulation, incentive) and outputs (electrical signal). The dependence of inputs and outputs is described by the mathematical functions, diagrams or other indicators. The transfer function $T(p)$ is considered in the sense of limiting the set of available variables and degeneration of the original values of traffic variables. Apart from the characteristics of the information sources that affect the restriction of possibilities and deformation of original parameter values, the sources are also affected by external influences – disturbances that additionally deteriorate the quality of output information. Therefore, if the sources are considered in the actual environment, with large and unforeseeable external influences, expressing the transfer function with mathematical exactitude is not useful for determining the quality of information obtained by the user. Therefore, $T(p)$ is described in terms of quality and the possibility of the information source to collect data that describe the traffic variables. Consequently, each element of the set of variables that is provided by the observed information source is assigned respective attributes that describe the transfer function.

Attributes are considered to evaluate a source of information. Formally, they are functions $\alpha_j: PP \to R,\ j = 1, \ldots, r$ defined on the set of traffic variables $PP = \{p_i,\ i = 1, \ldots, k\}$ that describe the quality of the information source. Typical elements of a finite set of attributes $A = \{\alpha_j,\ j = 1, \ldots, r\}$ are given in *Table 2*. The attributes should provide the quantitative description of the quality of the procurable set of parameters and the level of data fidelity in relation to the actual phenomena. Each element of the set of variables that is provided by the information source is assigned with a set of attributes. Thus, the output of the source is a set of variables $T(PP)$ that result from the action of the transfer function $T(p)$ and is formalized by expressions (4.10) and (4.11).

The transfer function $T(p)$ for the information source constitutes the set $T(PP) = \{T(p_i),\ i = 1, \ldots, k\}$ defined on the set of variables PP in the limited spatial-temporal traffic area P as a domain described by the ordered pair (gp,t) as shown in Eq. (4.10). The set $T(PP)$ is a set of all values $y \in \mathbb{R}$, which can be obtained by the information source with transfer function $T(p)$, and represents the traffic variables within the spatial-temporal scope (gp,t). The quality of mapping from the set PP into set $T(PP)$ is described by the finite set of attributes $A = \{\alpha_j,\ j = 1,2, \ldots, r\}$, which is assigned to every single element of the set PP. Therefore, the quality of the transfer function $T: PP \to R$ is described by matrix $TI(PP) = (f_{ij})$, where $f_{ij}: PP \times A \to R$ are functions that assign a real number to each pair $(p_i, \alpha_j) \in PP \times A$: value of attribute α_j for the traffic variable p_i. Thus,

$$T(PP) = \{y = R,\ \exists (gp,t) \subset \mathbb{R}^3 \times \mathbb{R},$$
$$\exists p_i \in PP,\ T(p_i(gp,t)) = y\} \tag{4.10}$$

and

$$TI(PP) = (f_{ij}),$$
$$f_{ij}: \{p_i, i = 1,2, \ldots, m\} \times \{\alpha_j, j = 1,2, \ldots, r\} \to R \tag{4.11}$$

where:

$T(p_i)$ – value of variable obtained from the information source,

α_j – attributes,

Matrix *TI*(*PP*) is required because the related attributes describe the matching of the original condition and the information obtained from the sources, which raises the problem of data verification, i.e., comparison with an actual and accurate condition. It is therefore necessary to select a reference source that is known to provide accurate data. Such a source represents the selected accuracy. Statistical analysis of the experiment results, based on the comparison of the outputs of the information sources, yields the elements f_{ij} of matrix *LI*(*PP*).

Expressions (4.10) and (4.11) hold for every layer with appropriate symbols. Therefore, the formal records for the traffic flow layer are given by expressions (4.12) and (4.13) as

$$T(PT) = \{y = R,\ \exists(gp,t) \subset \mathbb{R}^3 \times \mathbb{R},\ \exists pt_i \in PT,\ T(pt_i(gp,t)) = y\} \quad (4.12)$$

$$TI(PT) = (f_{ij}),\quad f_{ij}: \{pt_i, i = 1,2,\ldots,n\} \times \{\alpha_j, j = 1,2,\ldots,r\} \to R \quad (4.13)$$

5. QUALITY OF INFORMATION SOURCES

The information quality required by a user depends on the actual ITS application. Therefore, quality requirements from application to application can differ. The problem of defining and determining the information quality has not been solved. Universal quantitative and qualitative values that would give unambiguous evaluation of the quality of the information source have not been set nor adopted, although several attempts at this have been made. Strong, Lee and Wang [4,5] define information quality in general as the adequacy of usage for the information consumer. U.S. state transportation agencies, such as the California Department of Transportation, have attempted to define data quality in terms of types of information and the required accuracy for several ITS applications [7].

Consequently, the definition of information source quality can be stated as the level of correspondence of the set of variables and the related attributes that describe the information source and the set of variables and the related attributes requested by the information user. That is, in compliance with expressions (4.10) and (4.11), the quality of the information source is the level of correspondence of matrix *TI*(*PP*)*II* that describes the source and matrix *LI*(*PP*)*IZK* that is created by the information users based on their requests.

Unambiguous and universal defining of the information source quality in the traffic system uniquely defines the problem of determining and defining the attributes that would be universal for all the sources of information in the traffic system. The quantification of the attributes and setting of the quality level thresholds is the next step in determining quality. Since traffic demand for data depends on the application, i.e., the information user, high-quality data for one application can prove low-quality for another. Thus, the required set of attributes and their thresholds may differ depending on the application. Therefore, it is impossible to set a universal measurement of quality for all sources of information and all ITS applications. However, it is possible to form a set of attributes as the basis for comparison and evaluation of the information source. In that case the user or several users set their requirements according to the relevant attributes.

Table 2 gives the initial proposal of attributes that represent the information quality criteria for the information source at a given time. The finite set of attributes is finalized after the study of the characteristics of the concrete type of information source and after defining the requests of the information users.

Availability is the attribute which can achieve a binary value. If variable pi is measurable with a source of information, then $\alpha_1(p_i) = 1$. Otherwise, $\alpha_1(p_i) = 0$.

Accuracy is the attribute which can achieves a value between 0 and 1. Each source of information has value for accuracy of each variable which is observed:

$$0 \leq \alpha_2(p_i) \leq 100\%.$$

Resolution is the attribute which expresses the minimal range on the measure scale:

$$\alpha_3(p_i) = \min\{\Delta p, T(p_i + \Delta p) \neq T(p_i)\}.$$

Reliability is the attribute which depends only on type of source of information. If the variable p_i is observed with the distinct type of source then

Table 2 - Initial set of attributes for the information quality of the information source

Information Quality Attributes for the Information Source	Description of Attribute
Availability	Possibility of the source to obtain a certain type of data (parameter)
Accuracy	Level of agreement between actual, i.e., true, values and values obtained from the information source
Resolution	Level of division up to which the value of the parameter of interest can be presented
Reliability	Ability of the information source to send correct data
Timeliness	Time coverage by data provided by the source. Can the data be obtained from the information source precisely at the time it is required.
Coverage	Spatial coverage of the traffic space, i.e., spatial domain from which the information source can collect information.

$\alpha_4(p_i) = n(T)/N_0$, where $n(T)$ is number of sources of information which send correct data for observed time period T and N_0 is total number of sources. If some other variable p_k is observed with the same type of source as p_i is observed, then $\alpha_4(p_k) = \alpha_4(p_i)$. Obviously $\alpha_4(p_i)$is in the range [0,1].

Timeliness is the attribute depending on available time of using a source of information. Defined as $\alpha_5(p_i) = \Delta T/T$, where T is the duration of period the information are collected in, and ΔT is in general a part of T, the source of information for p_i is available. If $\Delta T \geq T$ then $\alpha_5(p_i) = 1$.

Coverage is the attribute primarily taken in values of square meters (m^2), rarely in cubic meters (m^3) or length (m). If traffic area is given in advance with its size $S(m^2)$, then coverage can be expressed in ratio: $\alpha_6(p_i) = \Delta S/S$. The value ΔS represents the proportion of the traffic area that could be observed with particular source. In this case $\alpha_6(p_i)$is in the range [0,1].

For the description of the information source quality in a real environment and practical application, it is necessary to expand the set of attributes by exploitation, economics and other additional criteria. The exploitation criteria are related to the time period from the moment of releasing the information source until the end of its lifetime. The exploitation criteria certainly include: compatibility with other parts of the automatic traffic control (ATC), starting its operative functioning, durability and maintenance. The economic criteria refer to the costs related to the purchase and exploitation of the information source. Additional criteria are all the others that may influence the assessment of sensors, and do not belong to any of the previous groups such as resistance to malevolent damaging, violation of privacy, impact on humans and aesthetic criterion.

6. CONCLUSION

Efficient operation in ITS applications and services, especially automatic traffic control systems require high-quality information sources in order to provide high-quality outputs, i.e. commands to executive devices of the control system. Regarding the diversity of the information source and the different requirements regarding the information quality, the assumption has been made that no universal measurement of the quality of the information source can be formulated. However, formal definition of the quality is necessary in order to include the quality factors, determine the sets of the possible requested information from the traffic system and to set the principle methodology for the evaluation of the information source, either for one user or for a group of users. Thus, formal definition of the quality becomes the initial step in the evaluation of the information source.

Determination of the set of variables that describe the condition in a single layer is of benefit for potential information users. If the information users define their requirements on the presented set of traffic variables, and if the potential of different information sources (existing or planned) are known regarding the collection of traffic variables, it is possible to make the collected data available to a wider circle of users by upgrading the information source or organization of information distribution.

The evaluation of the information source, in the sense of assigning also a quantitative grade that will reflect the quality to the source, is a complex procedure that includes several interconnected parts, namely the study of the characteristics of single sources of information, selection of the set of attributes and evaluation per single attributes, determination and quantification of the significance of single attributes and quality criteria, and the synthesis into a unique grade which reflects quality.

Dr. sc. **NIKO JELUŠIĆ**
E-mail: niko.jelusic@fpz.hr
Dr. sc. **MARIO ANŽEK**
E-mail: mario.anzek@fpz.hr
Mr. sc. **BOŽIDAR IVANKOVIĆ**
E-mail: bozidar.ivankovic@fpz.hr
Sveučilište u Zagrebu, Fakultet prometnih znanosti
Vukelićeva 4, 10000 Zagreb, Republika Hrvatska

SAŽETAK

KVALITETA IZVORA INFORMACIJA U INTELIGENTNIM TRANSPORTNIM SUSTAVIMA

Suvremeni sustavi automatskog upravljanja prometom, razne druge ITS (Inteligentni transportni sustavi) aplikacije i servisi oslanjaju se na stvarno-vremenske informacije iz prometnog sustava. U radu je prikazana funkcija i pregled vrsta izvora stvarno-vremenskih informacija koji se koriste ili bi se mogli koristiti u ITS-u. Cilj rada je formalno definirati kvalitetu izvora informacija za ITS. Postavljeni su formalni model prometnog sustava i model izvora informacija kao temelji definiranja kvalitete izvora informacija. Definicija kvalitete obuhvaća bitne čimbenike: informacije o prometnom prostoru koje postoje ili mogu biti tražene, zahtjeve korisnika te atribute koji opisuju izvore informacija. Time se daju okviri i smjernice za evaluaciju izvora informacije za ITS, koja treba uzeti u obzir sve relevantne čimbenike koji mogu utjecati na izbor izvora informacija za konkretnu primjenu.

KLJUČNE RIJEČI

izvori informacija, kvaliteta izvora informacija, Inteligentni transportni sustavi (ITS), automatsko upravljanje prometom

LITERATURE

[1] **Klein, L.A.**: *Sensor Technologies and Data Requirements for ITS*, Artech House, Boston, London, 2001.

[2] **Klein L. A.**, **Mills M. K.**, **Gibson D. R.P**: *Traffic Detector Handbook*, Third EditionVol.1; Vol. 2, FHWA, October 2006.

[3] **Martin, P.T.**, **Feng, Y.**, **Wang, X.**: *Detector Technology Evaluation*, MPC Report No. 03-154, University of Utah Traffic Lab, 2003.

[4] **Strong, D.M.**, **Lee, Y.W.**, **Wang, R.Y.**: *10 Potholes in the Road to Information Quality*, Computer, IEEE Computer Society Press, Volume: 30, Issue: 8, 1997.

[5] **Strong, D.M.**, **Lee, Y.W.**, **Wang, R.Y.**: *Data Quality in Context*, Communications of the ACM, ACM Press, New York, Vol. 40, No. 5, 1997.

[6] **Jelušić, N.**: *The Evaluation of Sensor Technology in Automatic Road Traffic Control System*, Ph Thesis, Faculty of Traffic Science University of Zagreb, 2008. (in Croatian)

[7] *Statewide Detection Plan - Data Quality Evaluation*. Technical memorandum prepared for California Department of Transportation by Cambridge Systematics, Oakland, CA with Lawrence A. Klein. July 2, 2008.