

# CHOICE OF THE MODEL OF PRESENTING TEMPORAL DATA IN THE SYSTEM OF THE TRAIN TRAFFIC SCHEDULE

Eugene Kopytov\*, Vasilijš Demidovs, Natalia Petukhova  
Transport and Telecommunication Institute,  
State Joint-Stock Company “Latvian Railway”  
Riga, Latvia  
E-mail: [kopitov@tsi.lv](mailto:kopitov@tsi.lv), [dem@ldz.lv](mailto:dem@ldz.lv), [natalia@ldz.lv](mailto:natalia@ldz.lv)

## ABSTRACT

The authors have suggested three possible models of presenting temporal data in Information Systems (IS) of the train schedule, differing in the building principles. Choice of the optimal model of data presenting is performed with the account of different indicators characterizing the IS. For evaluating the efficiency of suggested models the Analytic Hierarchy Process (AHP) method is applied. There are developed 13 evaluating criteria, which are distributed in three groups: Hardware, Maintenance and Performance. The research is carried out for four different classes of IS. For user friendliness there have been used different tools of visualization which allowed visual presentation of the most important criteria and of the alternatives' preference.

Keywords: Train traffic schedule, model of temporal data, criteria, Analytic Hierarchy Process.

## 1. Introduction

The principal issues of the train traffic schedule storage in the databases of the railway IS are connected with the existence of the multitude of its versions conditioned by the seasonal cycles of schedule changes, the days of week, by the systematic and unplanned repair operations, the transfers of the working and festive days, and other factors. The basic problem is contained in the complexity of the data reliability providing and the immediacy of the schedule amending. Under the condition of the frequent alternations introducing these tasks become extremely urgent. Building a flexible schedule system, accounting the above limitations and rules and giving the employees of the railway companies effective mechanisms of controlling the schedule for a more successful meeting of the passenger transportations demands, is a challenging scientific-technical task of developing a model of presenting the train traffic schedule data and the method of interaction with the stored data, the method of searching an actual schedule version for a certain date in particular (Kopytov, Demidovs, and Petukhova, 2008).

The efficiency of manipulating with the schedule greatly depends on the organization of temporal data in the database. Here, various models of presenting temporal data of the train schedule are used. Choice of the best model in the terms of the given criteria efficiency is the task of the multi-criteria optimization. For solving this task, we can use various methods, among which the AHP method seems to be the best since it helps to find the solution of the problem in question (Saaty, 2001).

## 2. Models of presenting temporal data of the train traffic schedule

The authors have suggested three possible models of presenting temporal data of the train schedule: the model of the daily schedule presentation, the model based on logical rules and the model based on temporal elements, differing in the building principles, the capabilities, the requirements for the

---

\* Corresponding author

resources employed, etc. In suggested models two basic sets are introduced for the train schedule description: the set of all stations of the railways  $S = \{s_1, s_2, \dots, s_l\}$  and the set of the all trains  $N = \{n_1, n_2, \dots, n_m\}$ , where  $l$  and  $m$  are the capacities of the sets  $S$  and  $N$  correspondingly. Let us consider these models in details.

*Model of the daily schedule presentation (Duplicating model).* The given model is based on making a calendar of each train traffic for every day. Then  $i$ -th schedule version  $v_i^{(n)}$  for the train with number  $n \in N$  will be determined by the tuple:

$$v_i^{(n)} = \{\langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle, t_i\}, i = 1, 2, \dots, k, \quad (1)$$

where the pair  $\langle s_j^{(n)}, T_j \rangle$  determines the  $j$ -th stop of train with number  $n$ , the station  $s_j^{(n)} \in S$  exactly, and the train departure time  $T_j$ ;  $h$  is the number of the stops of the train running on the timetable  $v_i^{(n)}$ ;  $k$  is the number of days on which the schedule of  $n$ -th train is stored in database.

It should be noted that the number of the stored records, stating on what day, at what time, at what station every train stops, will be approximately equal to the product of multiplication of the number of trains by the average number stops and by the average number of days for which the schedule is stored. For example, with such approach for storing in the relation database the data of the Latvian Railway train traffic for a period of one year there will be needed about 2 million records, and for larger railway companies the volume of the corresponding database will increase ten times as much. And for storing the schedule for the period of several years, the number of records will be described in hundreds million. Though, the main problem of the above approach lies not in the volume of the stored data but in the complexity of providing the reliability of data and the immediacy of the schedule changes. Under the condition of frequent changes, these tasks become urgent.

*Model based on logical rules (Model LR).* The main point of the considered model is calculation of an active schedule version on the logical rules basis, which takes into account all the specificities of the schedule: multi-versioning, caused by the operative changes, multi-variance in case of a periodical schedule, replacements and movements. For the specific schedule version identification for the train with number  $n \in N$  the tuple  $\langle n, C, t_f, t_s, t_e \rangle$  is employed. The property  $C$  presents the logical expression consisting of one or several elementary characteristics of the periodicity  $p_m$  connected by the logical operations signs  $\vee$ ,  $\wedge$  and  $\neg$ . Then  $i$ -th schedule version  $v_i^{(n)}$  for the train with number  $n \in N$  will be

$$v_i^{(n)} = \{\langle n, C, t_f, t_s, t_e \rangle, \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle\}, i = 1, 2, \dots, q, \quad (2)$$

where  $q$  is the number of  $i$ -th train schedule versions, it is significant that  $k \gg q$ .

The model works with the data of the schedule's operative temporal database and accounts all changes in the schedule as soon as they appear in the system. The central parameter of an access method, relatively to which all the algorithms' logical rules go off, is the day which we need to find an active train schedule version for. However, as it was mentioned above, the schedule system's calendar is subject to changes, which appear as exception declarations and replacement assignments.

*Model based on temporal elements (Model TE)* (Terenziani, 2003). The peculiarity of the suggested model of determining the active version of the temporal object lies in the preliminary reckoning the dates of activity for every schedule version and saving it in the temporal elements for the posterior employment. The timetable version  $v_i^{(n)}$  of the train with number  $n \in N$  can be determined with the temporal element:

$$v_i^{(n)} = \{\langle n, TE_i \rangle \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle\}, i = 1, 2, \dots, q, \quad (3)$$

where the temporal element  $TE_i$  determines all the dates when the version  $v_i^{(n)}$  is active. The temporal element represents the calculable set of dates and serves for the time determination when the exact version of the train schedule on the extent of its life cycle becomes active. The temporal element comprises all information on the periodicity of the timetable, taking into consideration all operative changes several versions overlapping and special calendar (Kopytov, Petukhova, and Demidovs, 2010). One of the most serious problems of the temporal elements employment is maintenance of the temporal elements in the consistent condition, because in the case of the schedule changes occurrence the versions temporal elements might become out-of-dated and they are to be recalculated.

The variety of the data presentation models is determined by a set of different systems using the train schedule and characterized by its functionality, the requirements to the resources and the traditions of development. The given paper gives consideration of four classes of such systems described below in the section 3.

### 3. Systems of the train traffic schedule

The train traffic schedule is used in different IS. According to the character of the tasks under solution and the specific features of functioning, we'll distinguish four classes of systems: Web-based schedule systems, Mobile schedule systems, Ticket sales systems and Rail traffic management systems. Let us consider these systems in details.

*Web-based schedule systems* serve to provide information about the trains' traffic schedule for the passengers using the Internet. Such systems must have sufficiently broad information in order to satisfy most of the users' possible enquiries, there included: information about convenient changes, travel time, schedule for the given day, week, month, etc. The user's Internet-browser is only a terminal for the interaction with the schedule system and the storage of data, and all the calculation processes take place at the specially designed servers.

*Mobile schedule systems* support the users of the mobile means, such as mobile phone, the portable terminal, etc. Such systems are characterized by the autonomy of functioning: in contrast to the previous class systems, all calculation processes and data storage are performed in the mobile device, but the schedule updating periodically is downloaded from a remote server. Due to the limited technical capabilities of most current mobile means, the efficiency of the available resources employment comes to the foreground. The systems are characterized by a limited set of the schedule enquiries, which are quite simple, for example: search of the trains' departure from one station to another, presentation of the train's route for the given section with the indication of the stops' time.

*Ticket sales systems* are designed for selling tickets. The systems are highly specialized; their main task is search of the actual train schedule for a given day. The most essential requirement for these class systems is on-line entry of changes into the schedule.

*Rail traffic management systems* are necessary for planning the trains' traffic. The users of such systems are the railway specialists. The systems are characterized by broad functionality connected with the necessity of sorting out various planning tasks and providing their results in different forms. In such systems there are solved the issues of coordinating the railway traffic with the account of other transport means' schedule, the issues of allocating time-windows in the trains' traffic for repair works, etc. Due to the fact that a high level of the trains' traffic security must be provided, the above system requires special accuracy in trustworthiness and reliability of data. These systems also support planning of the locomotive crews' and conductors' performance, processing the locomotive-driver' itinerary, etc.

#### 4. Practical appliance of the AHP method to choose the model of presenting the schedule data

Search of an optimal model of data organization for each of the systems considered above must be performed with the account of various different criteria determining the efficiency of the schedule system on the whole. Concluding from the above, we can see that the priorities of these criteria may differ substantially depending on a particular schedule system. On the basis of different systems' requirements, there has been developed a set of criteria for evaluating the considered models of presenting data. The chosen criteria have been distributed in three groups: Hardware, Maintenance and Performance, which most aptly characterize the sphere of application of the models under analysis. Let us consider the given groups in detail. *Hardware* determines the importance of the effective use of the allocated calculation resources, there are included attributes CPU, HDD, RAM. *Maintenance* determines the costs of the system daily exploitation connected with entering a season schedule, entering random changes, additional data processing, integrity control and difficulty of adjusting to new rules. *Performance* is determined by the requirement to the efficiency of performing typical tasks of the schedule system, such as search of actual train traffic for a given day, making schedules for big time spans, making schedules for small time spans, search of activity intervals of a particular version of schedule, search of the nearest schedule change. The hierarchical structure of the employed criteria is shown in Figure 1.

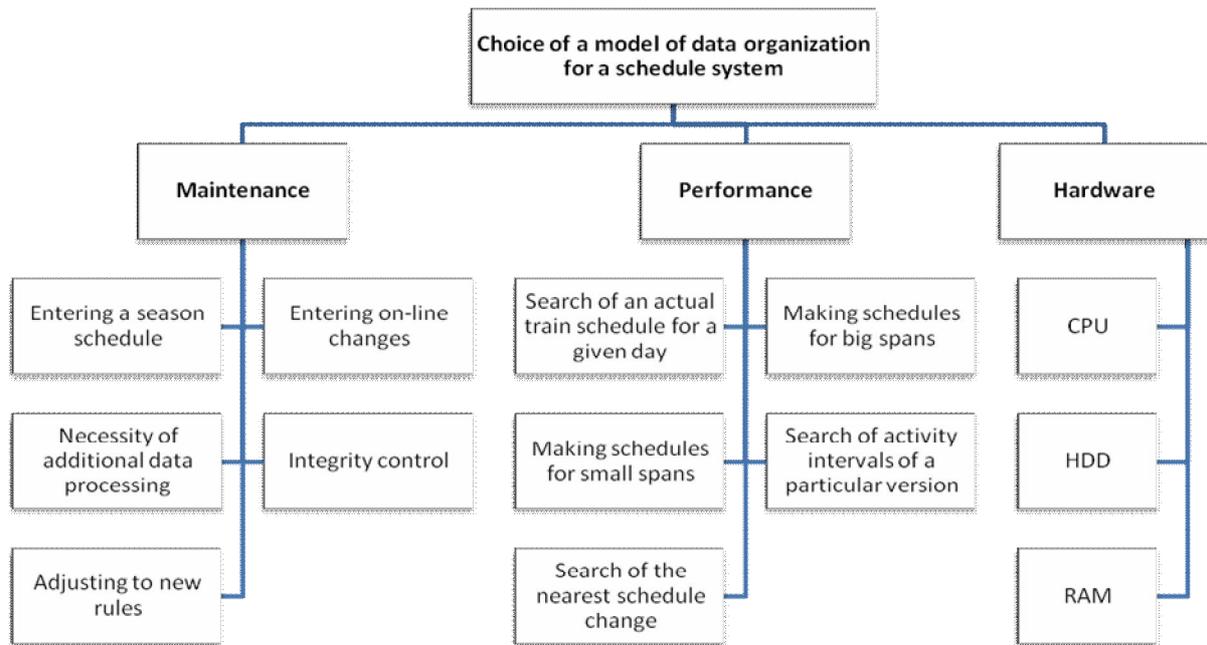


Figure 1. Hierarchy of the of the criteria for evaluating the considered models

To perform the calculations of criteria, the authors use standard algorithms of the AHP method with the commonly used pair wise comparison scale from 1 to 9 (Saaty, 2001). And the article gives only the evaluations of the pair wise comparisons of criteria and the summary results of the calculations.

The summary data of the pair wise comparisons for the criteria of the first hierarchy level for each investigated system are presented in Table 1. From the received evaluations of the criteria priority vector, it is seen how the importance of the criteria is changed for different systems. Maintenance and Performance criteria with identical values 0,467 of priority vectors are more important for Web-based schedule system, Hardware criteria with value 0,761 – for Mobile schedule system, Performance criteria with value 0,559 – for Ticket sales system, but Maintenance criteria with value 0,709 – for Rail traffic management system.

Table 1. Paired comparison matrixes first hierarchy level with respect to the goal for different trains' traffic schedule systems

	Maintenance	Performance	Hardware	Priority Vectors
<i>Web-based schedule system</i>				
Maintenance	1	1	7	0,467
Performance	1	1	7	0,467
Hardware	1/7	1/7	1	0,067
<i>Mobile schedule system</i>				
Maintenance	1	1/2	1/9	0,082
Performance	2	1	1/5	0,158
Hardware	9	5	1	0,761
<i>Ticket sales system</i>				
Maintenance	1	1/2	7	0,371
Performance	2	1	6	0,559
Hardware	1/7	1/6	1	0,070
<i>Rail traffic management system</i>				
Maintenance	1	4	9	0,709
Performance	1/4	1	5	0,231
Hardware	1/9	1/5	1	0,060

Table 2 gives a normalised evaluation of the priority vector of the second hierarchy level for each system. The results of the evaluations show substantial differences of the considered criteria priorities in the systems of different classes.

Table 2. A normalised evaluation of the criteria priority vector for different schedule systems

Criteria		Systems			
		Web-based schedule system	Mobile schedule system	Ticket sales system	Rail traffic management system
Maintenance	Entering a season schedule	0,334	0,437	0,334	0,070
	Entering on-line changes	0,367	0,364	0,334	0,238
	Necessity of additional data processing	0,084	0,074	0,027	0,027
	Integrity control	0,166	0,077	0,110	0,450
	Adjusting to new rules	0,049	0,049	0,196	0,215
Performance	Search of an actual train schedule for a given day	0,484	0,574	0,692	0,556
	Making schedules for big spans	0,040	0,000	0,077	0,111
	Making schedules for small spans	0,198	0,183	0,077	0,111
	Search of activity intervals of a particular version	0,040	0,041	0,077	0,111
	Search of the nearest schedule change	0,239	0,164	0,077	0,111
Hardware	CPU	0,540	0,333	0,413	0,260
	HDD	0,297	0,333	0,260	0,413
	RAM	0,163	0,333	0,327	0,327

Proceeding from the received evaluations of the criteria priority vector of two levels of the hierarchy, we calculate the final matrix of the evaluations of the global priority vector for the suggested models of organising the schedule data (see Table 3), which may be used for choosing a model of presenting data in a particular schedule system.

The AHP method allows arranging the models in the order of efficiency of their employment in every system and showing their difference in the given set of criteria. Performing evaluation of the alternatives for selecting possible variants of their employment by the AHP method allows revealing their universe character. Thus, in the example under consideration there has been revealed the universe character of the TE model, which has advantage in three systems under consideration: over

the nearest models-competitors for a Web-based schedule system by 14.40%, for a Mobile schedule system by 9,90% and for a Rail traffic management system by 0.40%, and only for the Ticket sales system the TE system yields the Duplicating model by 1,00%.

Table 3. Evaluations of the vector of the global criteria priorities for different schedule systems

Alternatives	Web-based schedule system	Mobile schedule system	Ticket sales system	Rail traffic management system
<b>Model LR</b>	0.280	0.353	0.220	0.359
<b>Model TE</b>	0.432	0.452	0.385	0.363
<b>Duplicating</b>	0.288	0.195	0.395	0.279

**5. Visualization of the evaluations' results**

With a big number of criteria and various alternatives, the process of evaluation may be pretty long and boring. To sort out this problem, the authors suggest using the means of visualization (color and symbolic), which simplify the process of comprehending and verifying the results performed at every step of the evaluation. Verification of the numerical data, each of which consists of several numerals, allows the expert to get quickly aware. Also the same indication helps "to see" why this or that alternative has become leader in the selection.

In connection with the fact that all calculations of the given research have been performed in the MS Excel 2007, for more clear presentation of the results there has been used the Conditional Formatting means which allowed, with the help of gradient colors, data bars, and icon sets, to emphasize visually the most important criteria and to show the alternatives' preference. In using the indication of gradient colors, two variants appear useful: differentiation in the saturation of one color and differentiation in the color gamma. But the color visualization is not always acceptable due to different reasons, for example, black-white print or special feeling of the colors by the user. Symbolic (sign) indication is more universal and it is it that is demonstrated in the given article (see tables 4 and 5). Note that with the evaluations' change, the indication changes automatically.

In the given article, for the symbolic indication of the criteria influence, there have been used the following icons: , indicating correspondingly the degree of influence from the maximum value to the minimum one; for the indication of the global priorities' alternatives, there have been used three icons:  - for the alternatives having high priority,  - for the alternatives with middle priority,  - for the alternatives with low priority. To demonstrate the application of the given icons, Table 4 gives an example of the results of calculating the priorities of the second level criteria "Maintenance" for selecting an optimal model of presenting data in the schedule Web-system. The evaluations of vector of the global alternatives' priorities shown above in Table 3 and supplemented with symbolic indication are presented in Table 5.

Table 4. Matrix of evaluations of the vector of the criteria priorities of the "Maintenance" group for the Web-based schedule system supplemented with symbolic indication

Alternatives	Group priority: 0.467								Priorities in group "Maintenance"
	"Maintenance" group's criteria								
	Entering a season schedule	Entering on-line changes	Necessity of additional data processing	Integrity control	Adjusting to new rules				
	Priority vector								
	 <b>0.334</b>	 <b>0.367</b>	 <b>0.084</b>	 <b>0.166</b>	 <b>0.049</b>				
Model LR	 0.444	 0.429	 0.444	 0.615	 0.087	 0.449			
Model TE	 0.444	 0.429	 0.111	 0.319	 0.162	 0.376			
Duplicating model	 0.111	 0.143	 0.444	 0.066	 0.751	 0.175			

Table 5. Evaluations of the vector of the global priorities' alternatives for different schedule systems supplemented with symbolic indication

Alternatives	Web-based schedule system	Mobile schedule system	Ticket sales system	Rail Traffic Management System
Model LR	✘ 0.280	🏆 0.353	✘ 0.220	🏆 0.359
Model TE	✔ 0.432	✔ 0.452	🏆 0.385	✔ 0.363
Duplicating model	✘ 0.288	✘ 0.195	✔ 0.395	✘ 0.279

## 5. Conclusions

With the use of the AHP method the given research fulfils the evaluation of the efficiency of application of the three models of presenting data for the train traffic schedule in different information systems with the account of the specifics of their application. For determining an optimal method, there has been developed a two-level hierarchy system of criteria with ranging expert evaluations according to the importance and preference for all models on each system under consideration. For verifying the correctness of judgments in the criteria evaluation, there has been used the consistency ratio which, for different criteria groups, has not exceeded 5,77% for the Web-based schedule systems, 8,45% for the Mobile schedule systems, 6,92% for Ticket sales systems and 6,14% for Rail traffic management systems.

The AHP method has allowed ranging the models in the order of efficiency of their employment for each of the systems and has shown their differences in the given set of criteria and has allowed revealing a universal model – Model TE.

The AHP method has been realized in the MS Excel 2007. For more visual presentation of the results, there has been used the tool of visualization Conditional Formatting which allowed visual presentation of the most important criteria and of the alternatives' preference.

## REFERENCES

- Kopytov E., Demidovs V., Petukhova N. (2008). Modelling of Railway Schedule in Temporal Databases. In: *Proceedings of the International Conference "Modelling of Business, Industrial and Transport System" (MBITS'08)*. Riga, Latvia, Transport and Telecommunication Institute, 107-116.
- Saaty T.L. (2001) *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World*, New Edition 2001 (Analytic Hierarchy Process Series, Vol. 2), RWS Publications.
- Terenziani P. (2003). Symbolic User-defined Periodicity in Temporal Relational Databases. *IEEE Transactions on Knowledge and Data Engineering*. 15 (2), 489-509.
- Kopytov E., Petukhova N., Demidovs V. (2010). Methods for Railway Schedule Periodicity Support in Temporal Databases. In: *Proceedings of the 9<sup>th</sup> Joint Conference on "Knowledge-based Software Engineering" (JCKBSE'10)*. Kaunas, Lithuania, Technologija, 178-191.