Transport and Telecommunication, 2013, Volume 14, No 2, 102–115
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia
DOI 10.2478/ttj-2013-0009

MODELLING AND ANALYSIS OF LOGISTICAL STATE DATA

***Annegret Brandau*[1], *Jurijs Tolujevs*[2]**

[1]*Otto von Guericke University Magdeburg*
2. Universitätsplatz 2, Magdeburg, 39106, Germany
E-mail: annegret.brandau@ovgu.de

[2]*Transport and Telecommunication Institute*
Lomonosova str. 1, LV-1019, Riga, Latvia
E-mail: juri.tolujew@iff.fraunhofer.de

To manage perfectly an efficient and effective supply chain of continuous and undisturbed flow of goods is needed. To achieve this identification, location and sensor technologies must be implemented to generate state data of the logistics objects. However, the amount of information overstrains the operational logistics planner and the information systems have to face enormous data streams. Data mining methods are useful to cope with such big data streams, and they are well developed in the literature. But these methods are not often applied to logistical state data. Without knowledge of the processes, the results of the algorithms cannot be understood. Therefore, the objective of this work is to introduce a general concept to model and to analyse logistical state data, in order to find irregularities and their causes and dependences. This work shows that it is possible to use data mining methods on logistical state data to filter irregularities and their causes.

Keywords: modelling, logistical state data, supply chain event management, data mining

1. Introduction

In the last 60 years the globalisation of markets led to an enormous increase in the worldwide trade of goods [1]. Distribution networks are bigger and more complex than before because of the increasing volume and diversity of goods and the longer and intermodal transport channels. Additionally, the customer demands that the right good must be delivered in the appropriate quantity at the right place and time (as fast as possible). It must be of the right quality with the right costs (as cheap as possible), environmentally compatible and accompanied by the right information [2, 3, and 4]. Thus, as we can read in the mentioned literature source – "a continuous and undisturbed flow of goods is a central issue for an efficient and effective supply chain" [5, p. 1]. Unfortunately, a lot of unforeseeable irregularities in the supply chain happen, e.g. traffic jams, accidents, human errors and natural disasters. For this reason, companies, especially logistics service providers, must monitor their processes in real-time so that the reaction to a critical situation could be as fast as possible.

To achieve a transparent flow of goods, technologies are required to identify and to locate logistic objects and to measure the quality of goods, e.g. temperature, humidity and shock sensors. An intelligent information system is needed to handle the voluminous data flow of the installed technologies, e.g. tracking & tracing or supply chain event management software. Unfortunately, the existing tracking & tracing software has a lot of deficits. Most of the tools are not proactive and have no decision support implemented. [6 and 7] These deficits induce an information overload and increase costs to find the right information [8]. Supply chain event management tools are more sophisticated than tracking & tracing software tools and have an extensive functionality. Figure 1 shows the functionalities of a supply chain event management system. Tracking & tracing software supports the functionality "Monitoring & Reporting" by collecting and reporting data. The functionality "Identification" is important to find critical situations from the large set of state data. When a critical situation is identified, the logistics planner must be notified and decide what should be done to prevent damage. The logistics planner is supported by simulation tools to evaluate different suitable solution scenarios. The last two functionalities handle the execution of the intervention and the subsequent controlling [9, 10, and 11].

The most important functionality of the supply chain event management system is the "Identification", because the large set of state data has to be filtered, aggregated and analysed by different methods from information science. Mostly, the systems only alert when the current state of an object differs from the target state. But to make a sufficient decision the logistics planner needs more information about the current state of the system, current and further events and the causes and dependencies of the irregularities in the processes [5]. However, there is a deficit of clear and generally accepted concepts for analysing and interpreting logistical state data [12].

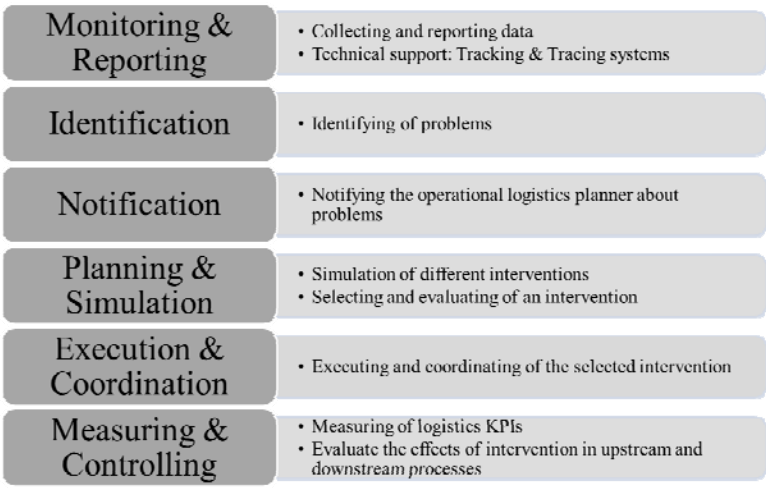

Figure 1. Functionalities of SCEM based on [9, 10, 11]

Because of the enormous data volume, data mining methods could be suitable to analyse logistical state data and to identify irregularities and their causes and dependencies. Unfortunately the application of these methods on real logistical state data has some disadvantages [13]:

- knowledge about the processes is needed to evaluate the results,
- the results depend on the cluster compositions and the choice of attributes,
- problems with nominal attributes,
- there are no general rules for choosing relevant attributes (irrelevant attributes induce susceptibility to discrepancies).

Therefore, the objective of this work is to introduce a general concept to model and analyse logistical state data, in order to find irregularities and their causes and dependences. The concept consists of four steps in which the first two steps consist of data modelling, the third step performs data preparation and the last step comprises data analysis and interpretation by data mining methods. The advantages of this holistic concept are that the data modelling for logistics systems is general and uses tools which are well known. The analysed data may have nominal and metric attributes; this is standard for logistical state data. Furthermore, general rules should be developed for the attribute choice to identify irregularities and their causes. In this work, the holistic concept and the first results of application of data mining methods on logistical state data are demonstrated.

In chapter 2, the notation, important definitions and basic methods are introduced. The generally concept is described in chapter 3. The application of the concept to an example is described with results in chapter 4. The last chapter includes the discussion, the conclusion and planned further research.

2. Notation, Definitions and Basic Methods

To understand the concept, a description of the terms “object type”, “object”, “state” and “event” must be given, because these terms have different meanings in the literature. Also, the data mining method which was used for the example is introduced shortly.

2.1. Object type, Object, State and Event

The term “logistical object” has various definitions [3, 14, 15, and 16]. Fleischmann defines a logistical object as a material good, e.g. materials and products for industry, persons or information [17]. In this work (logistical) objects of the systems are defined as objects, which are equipped with identification, location and sensor technologies so that it is possible to track their state and events; the only important objects are these that contribute to reaching the goal of the system, based on [15 and 16].

All objects of the system are classified in different object types. The first group are the physical object types which are divided in moving and stationary objects, e.g. goods, loading units, means of transport or transport channels and storage points. The physical objects are additionally classified in complex and not complex. An object type is complex if it can contain other object types. The classification is described in Table 1 [5, 12, 18, 19, and 20].

Table 1. Physical object types [5, 12, 18, 19, 20]

Physical objects types		
Moving object types	**Stationary object types**	**Complex object types**
- Goods - Loading unit - Means of transport - …	- Transport channel - Storage/transfer point - …	- Loading unit - Means of transport - Transport channel - Storage/transfer point - …

The other class of object types are abstract object types. These could be defined by humans or generated automatically. Furthermore, it is differentiated between business and information object types and groups of other object types, e.g. all goods from one supplier or a specified storage zone. The abstract object types are described in Table 2. [5, 12, 18, 19, and 20]

Table 2. Abstract object types [5, 12, 18, 19, 20]

Abstract object types		
Defined by humans		**Generated automatically or defined by humans**
Business object types	Groups/fragments	Information object types
- Contracts - Orders - Deliveries - …	- All goods from one supplier - Specified storage zone - …	- Objects that include data from physical or abstract objects - …

For all relevant objects of the system state data must be generated and be stored in a data base. The current state of an object is described by its attributes, which have been predefined by humans. One state is an n-dimensional attribute vector and has the form:

(Identification number of the object, time stamp, attribute 1, …, attribute n).

If at least one of the attribute changes an event happens and a new state must be created. Not all events are critical, because they are only alterations of the state data. The state data for all object types must be stored in a state protocol for analysing current and past state data [5, 12, 18, 19, and 20].

2.2. Knowledge Discovery in Databases and Data Mining

"KDD (Knowledge discovery in databases) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [21]. In this case, data mining is a part of the KDD process and is understood as "the application of specific algorithms for extracting patterns from data" [22]. Data mining methods are structured in four different groups: time series analysis, association, classification and clustering [23]. This work focuses on clustering methods which have the goal to find different clusters in the data. To achieve this, the similarity between the objects in one cluster must be high and between objects from different clusters low [23]. At first, the problem must be defined, in order to apply clustering methods. Subsequently, the relevant objects and attributes must be chosen. As next step the proximity measure must be chosen on the basis of the attributes´ scale levels (nominal, ordinal or metric). Finally, the data mining method may be selected and applied to the data set [23].

In this work logistical state data with nominal attributes are analysed. As proximity measure the trivial Hamming-distance for nominal attributes is chosen. For arbitrary n-dimensional nominal attributes vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ (assuming the ranges of attributes are discrete and finite), the Hamming distance is defined as

$$H(x,y) = z(x_1, y_1) + \ldots + z(x_n, y_n),$$

whereas for all $1 \leq i, j \leq n$ $z(x_i, y_j) = 0$ if $x_i = y_j$ else $z(x_i, y_j) = 1$, [24].

The Hamming distance is suitable to cluster state data in cluster with the same attribute value and is a trivial distance for first analyses [25]. In this paper, one method of clustering is used, the *k*-Means clustering.

The k-Means clustering is a centroid-based and partitioning clustering algorithm. The algorithm is described in Table 3. The inputs are the data set, the number of clusters k, the maximal number of iterations and arbitrary initial prototypes, in this case the status data [26]. The composition of the clusters depends on the initial prototypes, the maximal cluster number and the maximal number of runs. To overcome the dependence on the initial prototypes, it is possible to apply several runs with different initial prototypes. The dependence on the maximal cluster number is not trivial. One condition is that small cluster numbers are preferred, but it depends on the individual problem that needs to be solved [27].

Table 3. *k*-Means Algorithm based on [26]

Input: Data Set D, |D|=n
Value k
Maximum number of iterations
Initial prototypes (state data) $s_1, \ldots, s_k$
Output: Optimized prototypes $s_1, \ldots, s_k$

Repeat
  For all $1 \leq j \leq n$, $x_j \in D$:
    Update $p_{i|j}$ according to $p_{i|j} = 1$ if $H(s_i, x_j)$ becomes minimal for i; ties are broken arbitrarily
      otherwise $p_{i|j} = 0$
  For all $1 \leq i \leq k$
    Update s_i according to $s_i = (p_{i|1}\, x_1 + \ldots + p_{i|n}\, x_n) / (p_{i|1} + \ldots + p_{i|n})$
Until Maximal number of iteration reached
  or $s_1, \ldots, s_k$ converged

3. Concept to Model and Analyse Logistical State Data

The goal of the concept is to model and analyse logistical state data for identifying irregularities and their causes and dependences. It is based on four consecutive steps, cp. [28]. The first step comprises understanding the logistical system, especially the goal of the system. The second step is the data modelling step in which the structure of the state data is created. The data preparation is needed to transfer the generated data from the identification, location and sensor technologies to the previously developed data model. In the last step, the data analysis step, the state data are analysed by data mining methods to find irregularities and their causes and dependences. Figure 2 shows the four steps and their results and methods. In the following, the four steps are briefly introduced and afterwards, in Chapter 4, illustrated in an example.

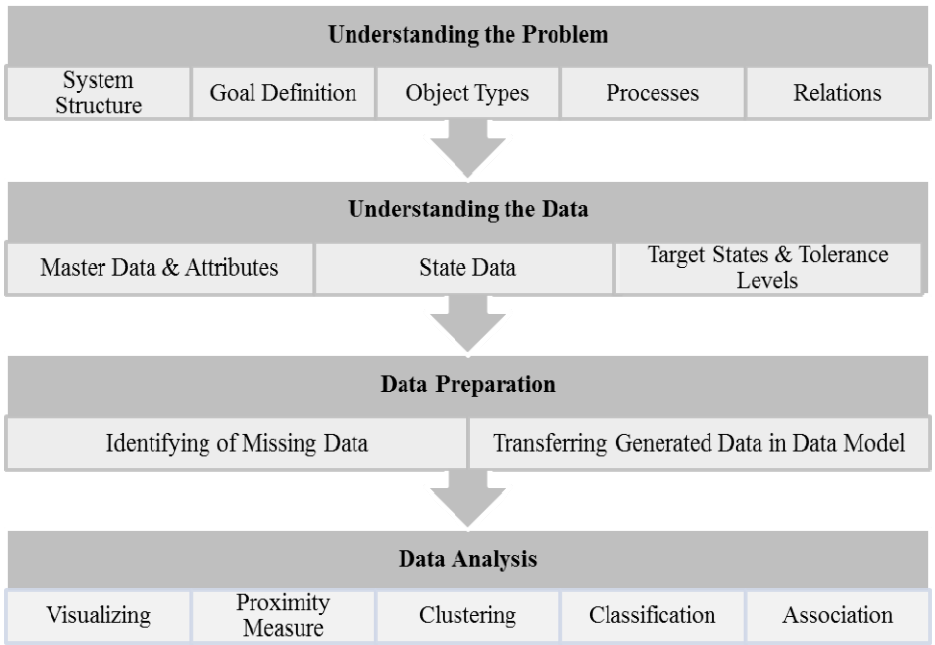

Figure 2. Concept to model and analyse logistical state data, based on [28]

3.1. Understanding the System

To get an overview over the *structure of the logistical system*, it is necessary to visualise it by e.g. a CAD design or a schematic drawing of the spatial arrangement. The *goal of the system* is defined by the 8 rights of logistics which means that the right object is in the right quantity, at the right place, at the right time, with the right costs, of the right quality, environmental-compatible, and with the right information [2, 3, and 4].

The objects of the logistics system must be classified based on an object-oriented view and the above-defined object types. Afterwards the *processes* of the moving object types and the *relations* between all object types must be defined.

3.2. Understanding the Data

For data understanding and data modelling, different data types for all object types must be defined. The temporally constant data of the object types, the *master data*, must be collected and stored in a data base.

The *state data*, in the form described in chapter 2.1, are n-dimensional attribute vectors with an ID, time stamp and predefined attributes. Therefore, for all objects of the object types a unique identification number must be chosen. The *attributes* are derived by the above-mentioned 8 rights of logistics, the processes and the relations. From these results a general data model can be deduced, which contains all relevant attributes to monitor the goal of the system. Also, the ranges of the attributes must be defined.

Finally, the *target states* and *tolerance levels* must be defined to compare the current and target states for identifying irregularities in the target processes.

3.3. Data Preparation

Generally, the implemented identification, location and sensor technologies do not generate all necessary data of the object types and mostly not in the right format. Therefore, data transformations in the previously defined data model must be applied. To fill up the missing attributes and state data, the means of the generated attributes could be helpful. In conclusion, the state data must be stored in state protocols which are constantly updated.

3.4. Data Analysis

The last and important step is data analysis. At first it is helpful to visualise the state data in three dimensions. This is only possible if the data volume is not too large. If a discrepancy between target and current state of an object happens the data analysis must be initiated. To achieve this, methods of clustering, classification and association are useful. It should be noted that without knowledge of the analysis methods and of the logistical system, it is very difficult to understand the results and, in turn, to deduce the causes and dependences of the detected irregularity.

4. Application Example and Results

The application example is a model of an airport where six aircrafts are handled. The observing time window is circa 3 hours from October 12, 2009 23:00:00 until October 13, 2009 02:12:30. The state data of all objects are generated by a discrete simulation model influenced by stochastic parameters. The logistical system includes 202 unique logistical objects which are classified in the above-mentioned object types. These objects generate altogether 4228 state data in the observed time window. This application example is based on a real freight airport but the arrival and departure times, the transhipped aircraft cargo and the stochastic default rates of the objects are fictitious.

4.1. Understanding the System for the Application Example

The airport has three stands for aircrafts and a warehouse for goods. The goods are ULDs (Unit Load Devices), which are standard containers and pallets for aircrafts. The transport of the ULDs on the airport ramp is organized by tugs and dollys. Tugs tow the trailers (dollys), which have the ULDs loaded. Special transport ways and areas are defined for the tugs, e.g. tug and dolly pool, garage and service station. On Figure 3 the spatial structure of the airport is visualised.

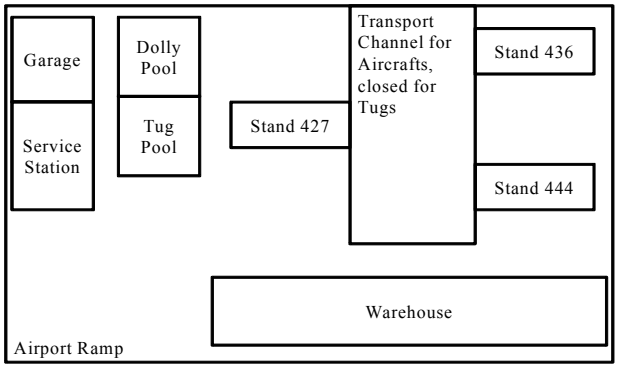

Figure 3. Spatial structure of the airport ramp

The goal of the system is that the right ULD is in the right aircraft at the right time in the right quality. The costs, the information on the ULD and the environmental-compatibility is not important for this example.

The physical object types are described in Table 4; there are no abstract object types in this example.

Table 4. Physical object types for the airport example

Physical objects types		
Moving object types	**Stationary object types**	**Complex object types**
- ULD - Dolly - Tug - Aircraft	- Stand - Transport channels (between Stands and Warehouse) - Warehouse - Tug Pool - Dolly Pool - Service Station - Garage	- Dolly - Tug - Aircraft - Stand - Transport channels (between Stands and Warehouse) - Warehouse - Tug Pool - Dolly Pool - Service Station - Garage

For process-understanding the processes of all moving objects must be described. Exemplary are the processes of the ULDs shown on Figure 4.

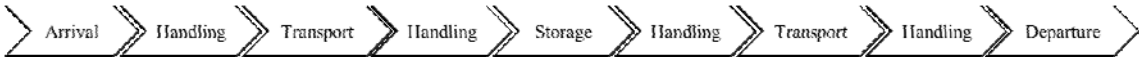

Figure 4. Processes of ULD

To identifying relationships between the object types all relations must be defined. The relation "contain" based on the object class "complex object types" is shown on Figure 5. The description follows the entity relationship model [14].

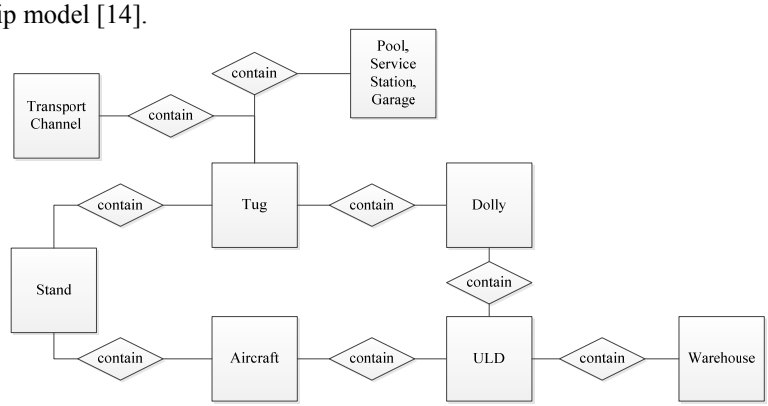

Figure 5. Relation "contain" for the application example

4.2. Understanding the Data for the Application Example

For seven different ULD types the master data are shown in Table 5. All ULD types are containers, type LD3_RKN is a refrigerated container. The content of the containers are fragile goods

Table 5. Master Data of the ULD Types

Typ	Function	Content
AAC	Container	Fragile Goods
AAK	Container	Fragile Goods
AAY	Container	Fragile Goods
AXY	Container	Fragile Goods
LD11_ALP	Container	Fragile Goods
LD3_AKE	Container	Fragile Goods
LD3_RKN	Refrigerated Container	Fragile Refrigerated Goods

The temporally constant data of the aircrafts are the total content, the number of different ULDs and the minimal transit time, shown in Table 6.

Table 6. Master Data of the Aircrafts

ID	Total Content [Unit ULD]	AAC	AAK	AAY	AXY	LD11_ALP	LD3_AKE	LD3_RKN	Min. Transit Time [Min]
FLZ1	17	0	0	14	1	0	0	2	37
FLZ2	28	20	4	0	0	4	0	0	56
FLZ3	20	0	10	0	0	0	5	5	39
FLZ4	24	4	4	4	4	4	2	2	48
FLZ5	17	0	0	14	1	0	0	2	37
FLZ6	28	10	4	0	0	4	5	5	56

For all complex object types the capacity which is temporally constant is described as master data in Table 7.

Table 7. Master Data of complex object types

Object Type	Object ID	Capacity
Dolly	Dolly1, …, Dolly40	1 ULD
Tug	Tug1, …, Tug10	4 Dolly
Stand	St427, St436, St444	1 FLZ
Transport Channel	Way St427-W, Way W-St427, Way St436-W, Way W-St436, Way St444-W, Way W-St444	4
Warehouse	W	250 ULD
Pool	TugPool, DollyPool	40, 10
Service Station	GS	3 Tug
Garage	G	3 Tug

To deduce attributes for the moving objects, the meaning of the 8 rights of logistics with regard to each moving object type must be determined, cp. Table 8. The attribute "ID" and "Date Time" are the first two components a priori of the state vector. The right quantity of all objects is constant one because the objects are not aggregated. To monitor that all objects are on the right place it is necessary to collect the attribute "Location". The quality of the objects in this example is only important for the ULD and the Tug. Hence the state data of the ULD have the attributes "Refrigeration" and "Shock", cp. Table 5, and the state data of the Tug "Wear" and "Filling level". All other rights are not important for this example.

Table 8. Attributes of moving object types deduced by 8 rights of logistics

8 Rights of Logistics	Attribute (ULD)	Attribute (Tug)	Attribute (Dolly)	Attribute (Aircraft)
Right Object	ID	ID	ID	ID
Right Quantity	one	one	one	one
Right Place	Location	Location	Location	Location
Right Time	Date Time	Date Time	Date Time	Date Time
Right Costs	-	-	-	-
Right Quality	Refrigeration, Shock	Wear, Filling Level	-	-
Environmental-compatible	-	-	-	-
Right Information	-	-	-	-

The next step is to deduce attributes from the relations. For this example, the relation "contain" is important. Therefore all complex object types must have the attribute "content", cp. Table 4.

The values of the important attribute "status", which describes the physical state of the object or what happens to the object, can be derived from the above-defined processes. For the ULD, the following status values are deduced, cp. Figure 4:

- Process: "Transport", status value: "From Aircraft";
- Processes: "Storage", "Handling", status value: "To Warehouse";
- Processes: "Transport", "Handling", status value: "From Warehouse";
- Process: "Departure", status value: "To Aircraft".

Additional status values can be deduced from the master data and relations. For example, when there is a defect in the ULD refrigeration, the status value is "Mistake". For the complex object types, status values can be derived from the relation and the master data "capacity", e.g. for stands, there are the following status values: "FLZ in", "FLZ out", "Tug in", "Tug out", "Dolly in" and "Dolly out".

Therefore, it is important to include the above-defined object types, processes, relations and master data to determine all important attributes and their ranges. In this example the ranges of the most attributes are nominal, except of "Time", "Wear" and "Filling Level".

Finally, the structure of the state data of all objects is determined, for the application example the structure is shown in Table 9. It is obvious that for the object type "ULD" the attribute "Content" is missing. This is because ULDs are not complex object types. Also for the stationary object types the attribute "Location" is not necessary.

Table 9. Structure of the state data off all object types

	ULD	FLZ	Tug	Dolly	Stand	Transport Channel	Warehouse	Pool	Service Station	Garage
ID	X	X	X	X	X	X	X	X	X	X
Time	X	X	X	X	X	X	X	X	X	X
Location	X	X	X	X						
Status	X	X	X	X	X	X	X	X	X	X
Content		X	X	X	X	X	X	X	X	X
Refrigeration	X									
Shock	X									
Wear			X							
Filling Level			X							

The last part of this step is to define the target states and tolerance levels. In this example, there are targets for the arrival time of the aircrafts. Furthermore, the departure time, as shown in Table 10, is deduced from the minimal transit time, cp. Table 6, and the availability of stands. Also, the stands are given for each aircraft. For the ULDs, it is predetermined in which aircraft the objects must arrive and depart, a short extract from the list is shown in Table 11.

Table 10. Target states of the aircrafts

ID	**Arrival Time [Date Time]**	**Departure Time [Date Time]**	**Stand**
FLZ1	12.10.2009 23:30:00	13.10.2009 00:07:00	St427
FLZ2	12.10.2009 23:40:00	13.10.2009 00:36:00	St436
FLZ3	13.10.2009 00:05:00	13.10.2009 00:44:00	St444
FLZ4	13.10.2009 00:18:00	13.10.2009 01:06:00	St427
FLZ5	13.10.2009 00:32:00	13.10.2009 01:14:00	St436
FLZ6	13.10.2009 00:40:00	13.10.2009 01:41:00	St444

Table 11. Extract from the list of target states of ULDs

ID	**Typ**	**Arrival Aircraft**	**Departure Aircraft**
ULD1	AAY	FLZ1	FLZ1
ULD3	AAY	FLZ1	FLZ1
…	…	…	…
ULD265	LD3_RKN	FLZ6	FLZ6
ULD267	LD3_RKN	FLZ6	FLZ6

For the attributes "Refrigeration", "Shock", "Wear" and "Filling level" are defined tolerance levels. These tolerance levels are stochastic parameters in the simulation model.

4.3. Data Preparation for the Application Example

Data preparation is important for the next analysis step. In many cases, the real state protocols are improperly formatted and must be transferred. In addition, some attributes are not generated by the implemented information, location and sensor technologies and must be filled up by means of the generated attributes.

In Table 12, an extract from the state protocol of ULDs is given. For a better understanding of the results it needs to be mentioned that the ID of the ULDs are only odd numbers. This is a technical detail of the simulation model and has no content relevance.

Table 12. Extract from the state protocol of ULDs

No	ID	Time	Place	Status	Refrigeration	Shock
1	ULD1	12.10.2009 23:31:00	Dolly1	From Aircraft	Refrigeration ok	None
2	ULD3	12.10.2009 23:32:00	Dolly2	From Aircraft	Refrigeration ok	None
3	ULD5	12.10.2009 23:33:00	Dolly3	From Aircraft	Refrigeration ok	None
	…	…	…	…	…	…
45	ULD31	12.10.2009 23:46:00	Dolly16	Mistake	Refrigeration Defect	None
46	ULD37	12.10.2009 23:46:03	W	To Warehouse	Refrigeration ok	None
47	ULD21	12.10.2009 23:46:28	Dolly11	From Warehouse	Refrigeration ok	None
48	ULD39	12.10.2009 23:46:33	W	To Warehouse	Refrigeration ok	None
49	ULD23	12.10.2009 23:46:58	Dolly12	From Warehouse	Refrigeration ok	None
50	ULD47	12.10.2009 23:47:00	Dolly24	From Aircraft	Refrigeration ok	None
51	ULD33	12.10.2009 23:47:00	Dolly17	From Aircraft	Refrigeration ok	None
55	ULD33	12.10.2009 23:47:00	Dolly17	Mistake	Refrigeration Defect	None
56	ULD41	12.10.2009 23:47:03	W	To Warehouse	Refrigeration ok	None
57	ULD35	12.10.2009 23:47:33	Dolly18	From Warehouse	Refrigeration ok	None
	…	…	…	…	…	…

4.4. Data Analysis and Results

The last step is to analyse the state protocols to find irregularities and their causes. Therefore, it is helpful to visualize the alteration of the attribute "status" (events) for all aircrafts over the time; this is shown on Figure 6. On the x-axis is the time and on the y-axis the status values. Every time an event happens to an aircraft, a point is drawn in the figure. The colour of the point describes the ID of the aircraft. The attributes "Content" and "Location" are not shown in the figure because there are only three dimensions.

The aircrafts FLZ1, …, FLZ6 arrive at the airport at target time, cp. Table 10 and Figure 6. After the arrival the aircrafts have repeated the status values "Unloaded" and then "Loaded". This means at each point that one ULD is unloaded or loaded. By comparing the target departure times, Table 10, with the real departure times, Figure 6, irregularities are identified. In addition the aircrafts FLZ1, FLZ2, FLZ3, FLZ4 and FLZ6 have reached the status value "Finished" in the observed time window. The aircraft FLZ5 is unfinished at the end of the time window. Now it is necessary to identify the causes of the delays.

To identify the causes of the delay, the analysis of the state protocol of ULDs is necessary. Visualization of the state protocol as above is not advisable because of the larger volume of state data. Therefore, the application of a clustering method could be helpful. In order to understand the resulting clusters, it is important to understand the Hamming distance between two state data of the ULDs. Due to the fact that the time stamp is not nominal but metric and that each row of the state protocol is uniquely defined by the row number, cp. Table 12, the time stamp will not be included in the distance calculation

and cluster determination. On Figure 7, the Hamming distances between all state data are shown, without the time stamp. On the x- and y-axis, the state data are labelled by the row number. The z-axis and the colours mark the Hamming distance between the respective state data.

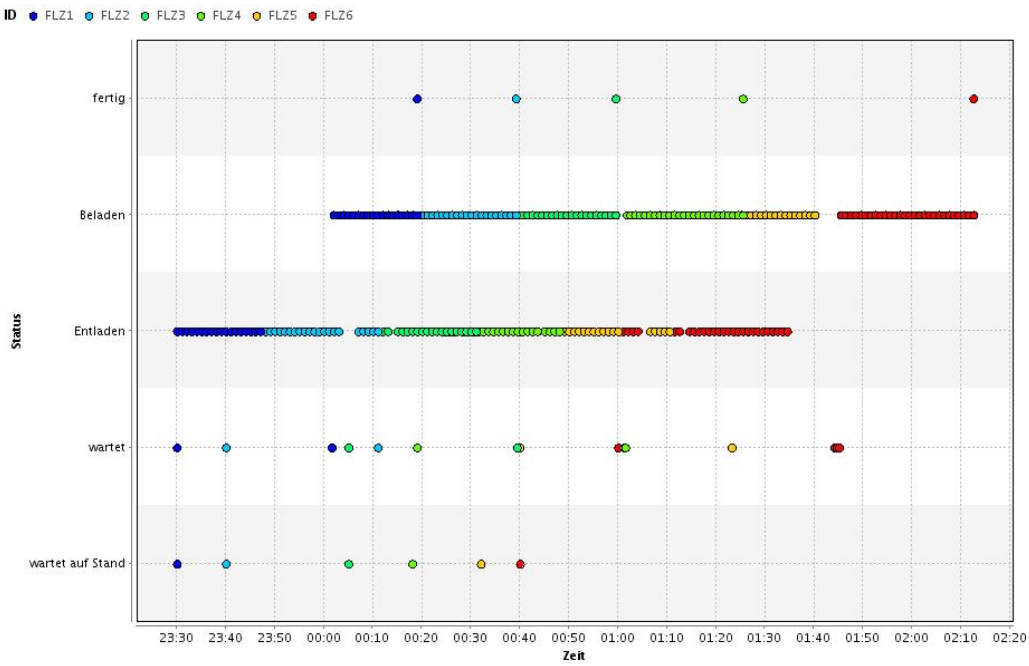

Figure 6. Events of the aircraft over time; only the attributes "Time", "Status" and "ID" are shown (Translation: Zeit=Time, fertig=Finished, Beladen=Loaded, Entladen=Unloaded, wartet=Waiting, wartet auf Stand=Waiting for Stand)

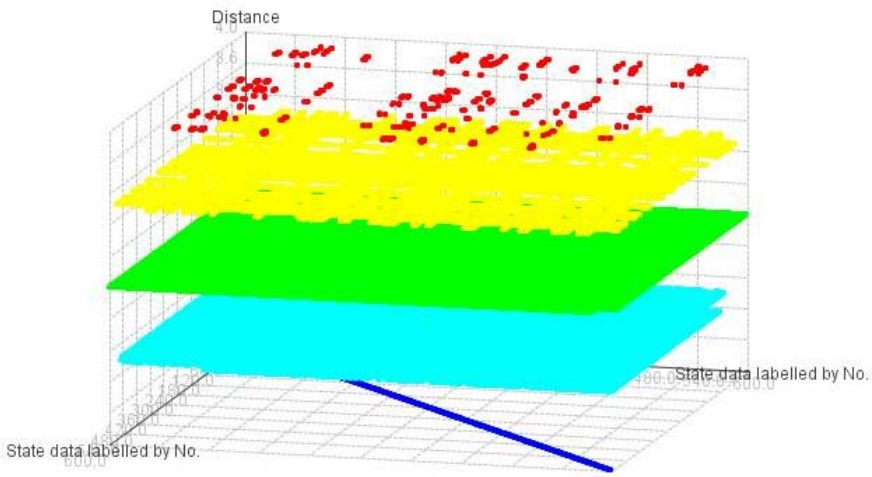

Figure 7. Hamming distances between all state data of the ULDs; on the *x*- and *y*-axis are the state data labelled by the row number shown; the z-axis and the colours mark the Hamming distance between the respective state data

Distance: 0.000 4.000

What do the differences in Hamming distances mean? To understand the distances it is important to look at the range of the attributes as shown in Table 13. If the state data have the same ID value, the state data refer to the same ULD. If the place value is the same, the ULDs are at the same place. The status values "From Aircraft", "To Warehouse", "From Warehouse" and "To Aircraft" should be passed once only by each ULD. The status values "Mistake" and "Mistake patched" could be passed more than once. The "Refrigeration" has only three values and most of the time, the ULDs should have the value "Refrigeration ok", if not there is a mistake or the mistake is patched. The same is true for "Shock", there are only two values and the goal is that the ULDs always have the value "None".

Table 13. Attribute ranges of the ULD state data without time stamp

Attribute	Range
ID	ULD1, ULD3, ULD5, …, ULD265, ULD267
Place	Dolly1, Dolly2, …, Dolly40, W, FLZ1, …, FLZ6
Status	From Aircraft, To Warehouse, From Warehouse, To Aircraft, Mistake, Mistake patched
Refrigeration	Refrigeration ok, Refrigeration defect, Refrigeration repaired
Shock	None, Shock

The distance of two state data is equal to zero when the compared state data without time stamp are equal. This means that the distance is calculated of a state data with itself or when a ULD has the same state data at different times (note that the time stamp is not included in the calculation). The first case is seen on Figure 7 as a blue diagonal line at the bottom. Because distance levels overlap with each other on Figure 7, these are shown separately for distance equal to 0, 1, 2, and 3 again on Figure 8 and equal to 4 on Figure 9.

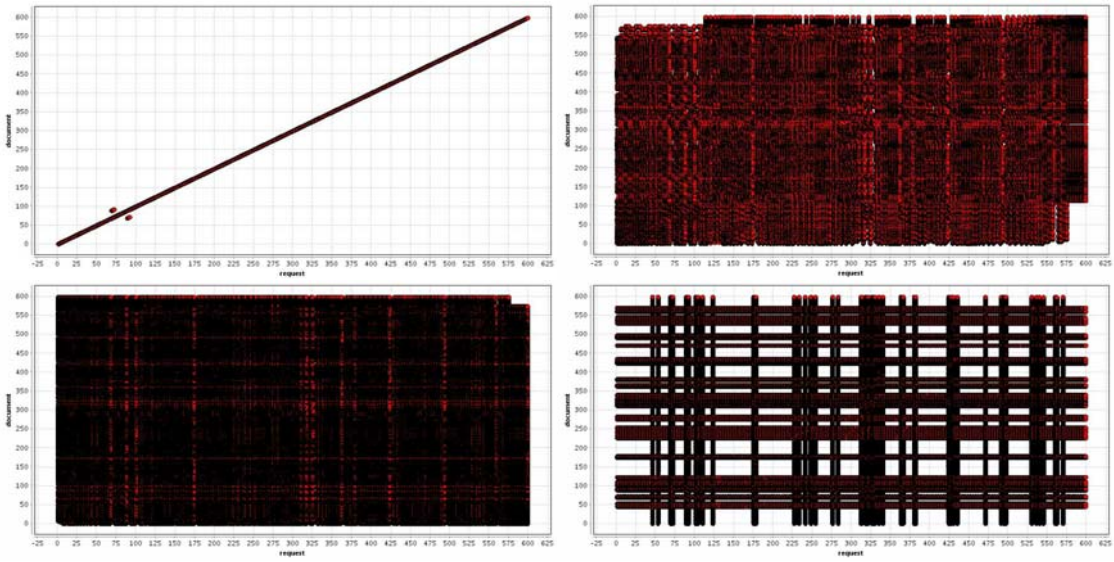

Figure 8. Pairs of state data of the ULDs which have the appropriate Hamming distance; top left: distance equal to 0, top right: distance equal to 1, bottom left: distance equal to 2, bottom right: distance equal to 3; *x*- and *y*-axis represent the state data labelled by row number

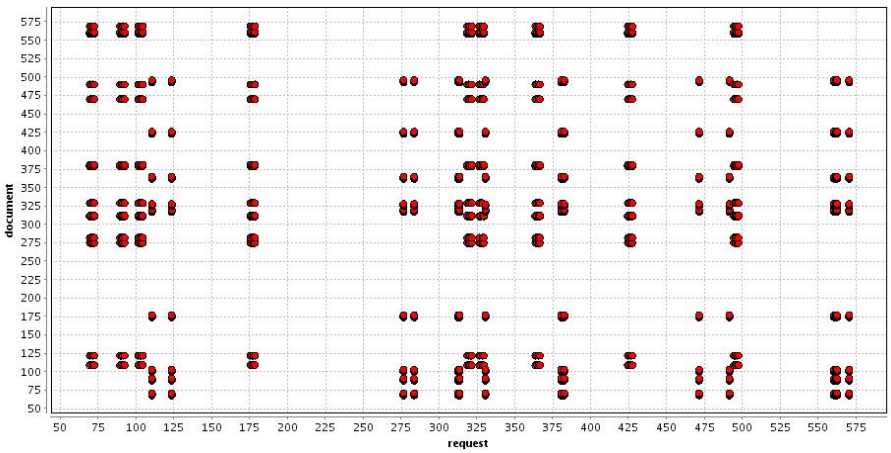

Figure 9. Pairs of state data of the ULDs which have the Hamming distance equal to 4; *x*- and *y*-axis mean the state data labelled by the row number

As it is shown above on Figure 8 at the top left, for pairs of state data with Hamming distance equal to 0, the points must be on the diagonal axis of symmetry. All points outside of the diagonal axis are not compared with themselves. It can therefore be concluded that ULDs have the same state data at different times. Because of the range of the attributes, this is once and only the case if the ULD is shocked

at two times. In Table 14 the pairs of state data together with their time stamp are shown which are not on the diagonal axis of symmetry. Upon comparison of the pairs, the conclusion above is confirmed.

Table 14. Pairs of state data together with their time and different row number stamp which have Hamming distance equal to 0

No	ID	Time	Place	Status	Refrigeration	Shock
69	ULD51	12.10.2009 23:52:00	Dolly26	Mistake	Refrigeration ok	Shock
89	ULD51	12.10.2009 23:57:03	Dolly26	Mistake	Refrigeration ok	Shock
70	ULD53	12.10.2009 23:52:00	Dolly27	Mistake	Refrigeration ok	Shock
90	ULD53	12.10.2009 23:57:03	Dolly27	Mistake	Refrigeration ok	Shock
71	ULD55	12.10.2009 23:52:00	Dolly28	Mistake	Refrigeration ok	Shock
91	ULD55	12.10.2009 23:57:03	Dolly28	Mistake	Refrigeration ok	Shock
72	ULD57	12.10.2009 23:52:00	Dolly29	Mistake	Refrigeration ok	Shock
92	ULD57	12.10.2009 23:57:03	Dolly29	Mistake	Refrigeration ok	Shock

Upon comparison of the pairs of state data with Hamming distance equal to 1 and 2 on Figure 8, horizontal and vertical lines of varying thickness are recognizable. These lines are repeated by the pairs of Hamming distance equal to 3, Figure 8 bottom left, and as points on Figure 9, where pairs with Hamming distance equal to 4 are visualised.

Based on this knowledge it made sense to apply a clustering on these state data set to filter the outstanding state data. In this example, k-Means clustering as described above was applied. Desired are three clusters, one for the state data without mistakes, one for state data with a mistake in attribute "Shock" and one for state date where the "Refrigeration" does not show the value "Refrigeration ok".

The open source tool RapidMiner was used as data mining software. The tests were implemented with different values of k (the maximal number of clusters), 100 runs with different initial prototypes and maximal 100 iterations.

On Figure 10, the elements of the clusters are visualised dependent on "ID" and "Status" for the maximal number of clusters k equal to 3. The same results are generated for k equal to 4. It can be recognized that the desired clusters are generated. Comparison of the elements (state data) in cluster 0 and 1 with the pairs of state data of Figure 9 shows that these are the same. This means that by using k-Means clustering it is possible to identify the ULDs for which irregularities happen. ULDs with defective refrigeration are grouped in cluster 0. Because this mistake caused delays, it is clear that the downstream processes will also be delayed. With that knowledge it is obvious why all aircrafts are too late. On Figure 10, it is recognizable that one ULD is only once in cluster 2 with the status value "From Aircraft"; from this fact, one can conclude that the mistake happened after the unloading process. The FLZ5 was not finished, because the defect ULD was not loaded anymore in the observed time window.

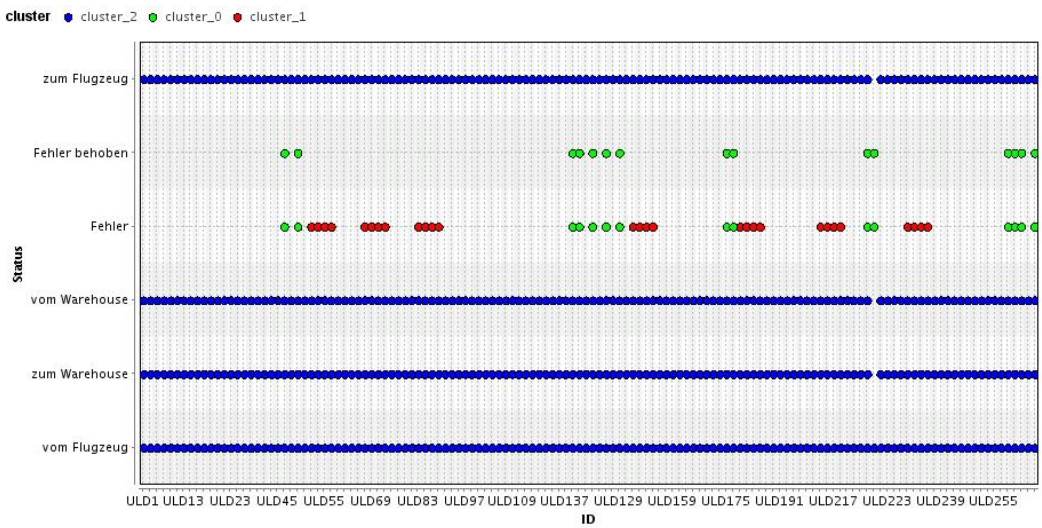

Figure 10. Elements of clusters in dependency of ID and Status by $k = 3$ and $k = 4$.
(Translation: Zum Flugzeug=To Aircraft, Fehler behoben=Mistake patched, Fehler=Mistake, Vom Warehouse=From Warehouse, Zum Warehouse=To Warehouse, Vom Flugzeug=From Aircraft)

5. Discussion, Conclusion and Further Research

Within this work, a general concept to model and analyse logistical state data to find irregularities and their causes and dependences was introduced and illustrated using an application example. It was presented that the k-Means clustering algorithm is suitable to identify irregularities on logistical state data with nominal attributes. Given these results, this work is consistent with the thesis of [13], which postulates that knowledge about the processes is needed to understand the results, and that the results of the data mining algorithm depend on the choice and ranges of attributes. Based upon these first results, more data analyses with different data mining algorithms will be performed and also applied to other examples. Furthermore, the present example is rather small with a total of 4228 state data. For further research it is necessary to extend the same analysis to a bigger data volume. As next steps general rules should be developed for the attribute choice to identify irregularities and their causes. It was also shown that data mining methods are useful to handle the current and further overload of data streams which is generated by identification, location and sensor technologies. These methods will support the operational logistics planner in monitoring a logistics hub.

References

1. Kruber, K.-P., Mees, A.-L., Meyer, C. (2013). Weltwirtschaftliche Entwicklungen zu Beginn des 21. Jahrhunderts. In *Informationen zur politischen Bildung*, No 299, 2008. Retrieved April 4, 2013, from http://www.bpb.de/izpb/8165/weltwirtschaftliche-entwicklungen-zu-beginn-des-21-jahrhunderts?p= all
2. Ziems, D. (2004). *Skript zur Vorlesung Technische Logistik*. Magdeburg: Institute of Logistics and Material Handling Systems of Otto von Guericke University.
3. Illés, B., Glistau, E. and Machado, N. I. C. (2007). *Logistik und Qualitätsmanagement*. Hungary: Miskolc,
4. Schenk, M., Wirth, S. and Müller E. (2010). *Factory planning manual: situation driven production facility planning*. Berlin Heidelberg: Springer-Verlag.
5. Brandau, A., Tolujevs, J. (2012). Logistics Event Management - An overview of concepts to interpret logistical real time data. In Proceedings of the 12th International Conference RelStat 12, October 17-20. Riga, Latvia: TTI.
6. Bretzke, W.-R. (2002). SCEM – Entwicklungsperspektive für Logistikdienstleister. *Supply Chain Management*, III, 27–31.
7. Nyman, H.J. (2012). An exploratory study of supply chain management it solutions. In *Hawaii International Conference on System Sciences* (pp. 4747–4756). Hawaii.
8. Heusler, K.F., Stölzle, W. & Bachmann, H. (2006). Supply Chain Event Management: Grundlagen, Funktionen und potenzielle Akteure. *WiSt*, 1,24.
9. Kurbel, K. & Schreber, D. (2005). Mobile SCEM Agentenbasiertes Event-Management in globalen Logistiknetzwerken. *Industrie Management*, 21(5), 21–24.
10. Yufei, Y. & Detlor, B. (2005). Intelligent mobile crisis response systems. *Communications of the ACM*, 48(2), 95–98.
11. Teuteberg, F. (2006). Diffusion und Anwendung von Technologien des Ubiquitous Computing zur Selbststeuerung im Supply Chain Event Management. In Ijioui, R., Emmerich, H. and Ceyp, M. (Ed.), *Supply Chain Event Management: Konzepte, Prozesse, Erfolgsfaktoren und Praxisbeispiele* (pp. 15–36). Physica Verlag.
12. Schenk, M., Tolujew, J., Barfus, K. and Reggelin, T. (2007). Grundkonzepte zu logistischen Echtzeitsystemen: Monitoring, Event Management und Frühwarnung. In Wolf-Kluthausen, H. (Ed.), *Jahrbuch Logistik 2007*, free beratung GmbH.
13. Windt, K., Knollmann, M. and Meyer, M. (2011). Anwendung von Data Mining Methoden zur Wissensgenerierung in der Logistik – Kritische Reflexion der Analysefähigkeit zur Termintreueverbesserung. In Spath, D. (Ed.), *Wissensarbeit – zwischen strengen Prozessen und kreativem Spielraum*, GITO.
14. Chen, P. P.-S. (1976). The entity-relationship model toward a unified view of data. *ACM Transactions on Database Systems (TODS),* 1(1), 9-36.
15. Wieser, O. and Lauterbach, B. (2001). Supply Chain Event Management mit mySAP SCM (Supply Chain Management). In Meinhardt, S. (Ed.), *Fertigungsmanagement in der Supply Chain,* no. 219 in HMD Praxis der Wirtschaftsinformatik. dpunkt-Verl.

16. Scholz-Reiter, B., de Berr, C., Freitag, M., Hamann, T., Rekersbrink, H. and Tervo, J. T. (2008). Dynamik logistischer Systeme. In Nyhuis, P. (Ed.), *Beiträge zu einer Theorie der Logistik* (pp. 109-138). Springer-Verlag.
17. Fleischmann, B. (2008). Grundlagen: Begriff der Logistik, logistische Systeme und Prozesse – Begriffliche Grundlagen. In Arnold, D., Isermann, H., Kuhn, A., Tempelmeier, H. and Furmans, K. (Ed.), *Handbuch Logistik* (pp. 3–12), 3 ed., chap. A 1.1. Berlin Heidelberg: Springer-Verlag.
18. Tolujew, J., Reggelin, T. & Sermpetzoglou, C. (2007). Simulation und Interpretation von Datenströmen in logistischen Echtzeitsystemen. In Engelhardt-Nowitzki, C. (Ed.), *Management komplexer Materialflüsse mittels Simulation: State-of-the-Art und innovative Konzepte* (pp. 215–231). Wiesbaden: Leobener Logistik cases. Dt. Univ.-Verl.
19. Brandau, A. & Tolujew, J. (2010). Logistics event management. In *XXIV. MicroCAD; Pszekció: Anyagáramlási rendszerek, logisztikai informatika (*pp. 7–18). Hungary: Miskolc.
20. Brandau, A. & Tolujew, J. (2011). Logistics event management. In Schenk, M. (Ed.), *9./10. IFF-Kolloquium Forschung vernetzen - Innovationen beschleunigen* (pp. 47–52). Magdeburg.
21. Fayyad, U.M., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery: An Overview. In Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy R. (Ed.), *Advances in Knowledge Discovery and Data Mining* (1–30). Californien, Menlo Park: AAAI Press.
22. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases, *AI Magazine,* 17, 37–54.
23. Petersohn, H. (2005). *Data Mining: Verfahren, Prozesse, Anwendungsarchitekturen*. München: Oldenbourg Wissenschaftsverlage GmbH.
24. Runkler, T.A. (2010). *Data Mining: Methoden und Algorithmen intelligenter Datenanalyse*. Wiesbaden: Vieweg, Teubner, GWV Fachverlage GmbH.
25. Maimon, O., Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook.* Springer.
26. Berthold, M.R., Borgelt, C., Höppner, F. & Klawonn, F. (2010). *Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data.* London: Springer-Verlag.
27. Bramer, M. (2007). *Principles of Data Mining.* London: Springer-Verlag.
28. Brandau, A. (2012). Konzept zur Modellierung eines Systems für die Analyse von Zustandsdaten. In Schenk, M. (Ed.), *13./14. IFF-Kolloquium Forschung vernetzen – Innovationen beschleunigen* (pp. 69–76). Magdeburg.