# Multi-Aspect Evaluation of Data Quality in Scientific Databases

Juliusz L. Kulikowski
Institute of Biocybernetics and Biomedical Engineering PAS
4, Ks. Trojdena Str., 02-109 Warsaw, Poland
e-mail: jlkulik@ibib.waw.pl

**Keywords**: scientific databases, data quality, semi-ordering of vectors, Kantorovitsch space

**Abstract**: It is here presented a method of multi-aspect data quality evaluation. The data can be evaluated from the points of view of their actuality, relevance, accuracy, credibility, irredundancy, completeness, legibility, etc. The quality factors can be extended on higher-order data structures. Vector representing the data structures' qualities can be compared and the best data structures can be chosen using the concept of semi-ordering of a linear vector space introduced by Kantorovitsch. The way of using this concept to data evaluation is described.

## 1. Introduction

The data quality evaluation problem arises both, when a database is to be designed and when database customers are going to use data in investigations, learning and/or decision making. However, it is not quite clear what the requirement of high data quality exactly means. Of course, it suggests that it exists a data quality evaluation method or, at least, that a method of comparison of two or more data sets' relative qualities is possible. Several other questions concerning the problem also arise, like:

a / Is it possible to characterise data quality by a single numerical parameter?

b/ What is the relationship between the qualities of separately taken data and the quality of a data set as a whole?

c/ Is the quality evaluated by a data supplier the same as this one evaluated by a customer?

d/ Is the data quality evaluated by different customers or by a given customer in various situations still the same?

e/ Is it possible to reach higher data quality by acquiring corresponding (similar) data from several independent data bases?, etc.

The aim of this paper is to present some suggestions of answering the above-formulated questions and of the relative problems solution.

## 2. Data quality characteristics

Before going to more detailed considerations it seems necessary to distinguish between the notions of *data quality*, *data value* and *data cost.* The two last notions correspond to commercial aspects of data management: data cost is the cost of data acquisition or the one proposed by a data supplier to the customer for data delivery, while data value reflects the profits expected or reached by an user due to data acquisition and a proper data usage. Data costs can be expressed in monetary units, while data value can be defined in a more general way as direct economical profits or by scientific, educational, cultural, social, medical, psychological, military, management or any other data "importance". It is evident that, in general, so widely defined data value not only cannot be expressed in monetary units but sometimes it also cannot be expressed by any other single numerical parameter. Next observation is that data value is relative, depending on various users' expectations or needs. For a given user (or a class of users) data value is time- and/or circumstances-dependent. And, at last, there is no direct correspondence between data value and data volume, while data cost is usually an increasing function of data volume.

The above-given remarks show us a basic difference between the points of view of data suppliers and their customers - final data users. A data supplier has to do with data acquisition costs on the one hand and with data distribution or delivery incomes on the other one. He compares these components expressed in the same monetary units and he is interested in maximisation of his profit being a difference between the two components. The situation of a data user is much more complicated: he has to do with data costs on one hand and with data value on the other one. However, any comparison of definite data costs with usually vague data value is meaningless: could it be reasonable to compare billions USD spent for reaching the one-bit information about existence or not-existence of any life-forms on the Mars' surface? In similar way the cost of data concerning effective medical treatment in a certain situation and the value of a saved human life are incomparable. This shows us that a data user very often is not able to justify a decision about data acquisition as a result of strong economical calculus. However, if he can be supplied by similar data from several sources, he is able to compare costs as well as qualities of data offered in different options or by different suppliers and to chose a better variant.

For this purpose the notion of data quality should be introduced. It should not be related to data content but rather it should reflect some more general data features, substantial in data using. It is also desirable that data quality is easy to be evaluated and subjected to strong mathematical rules.

It was proposed in [1,2,3] that information or data quality is represented by a real vector $\boldsymbol{v}$ whose components (quality factors) characterise such data properties as:

- actuality ($v_{act}$) ,
- relevance ($v_{rlv}$),
- irredundancy ($v_{irr}$),
- accuracy ($v_{acc}$),
- credibility ($v_{crd}$),
- completeness, ($v_{cpl}$)
- legibility ($v_{leg}$),

etc. Therefore,

$$\boldsymbol{v} = [v_{act}, v_{rlv}, v_{irr}, v_{acc}, v_{crd}, v_{cpl}, v_{leg}] \tag{1}$$

assuming that quality factors are expressed by real numbers. However, taking into account that the notion of quality, in general, can be used in relation to: a/ single data, b/ structured multi-component records, c/ multi-record files, d/ non-structured files, etc., it can be remarked that for various data structures different sets of quality factors are appropriate for data quality characterisation. For example, the notion of completeness is meaningless with respect to single data, accuracy suits better to numerical data, while credibility to the non-numerical ones. It also arises the problem of relationships between quality factors related to single data and the same ones related to higher-order data structures. For example, if a structured record has the form:

| | | | | | |
|---|---|---|---|---|---|
| *Identifier* | *D*1 | *D*2 | *D*3 | ..... | *Dn* |

where *D*1, *D*2,...,*Dn* denote some single data whose actuality, relevance, etc. have been evaluated then it arises the problem of evaluation actuality, relevance, etc. of the record as a whole. In similar way, if a file consists of a sequence of structured records with so or so evaluated actuality, relevance, etc., then it arises the problem of actuality, relevance, etc. evaluation of the file as a whole. Some suggestions how to solve the above-mentioned problems will be given below.

## 3. Quality factors of single data

There will be given below some proposals concerning the definition of single data quality factors, as a basis for further considerations. In general, they should satisfy the following general conditions:

1/ they should correspond to our intuitive understanding of actuality, relevance, etc.;

2/ the definitions should be constructive, i.e. they should suggest the way of the corresponding factors evaluation;

3/ they should be expressed by non-negative real and dimensionless numbers;

4/ they should be normalised with respect to some standard values adequate to application areas.

a/ Actuality

It is considered a data life-time and it is assumed that for a given event or process a certain type of data describe its current state at the time $\tau = 0$. Let us denote by $\tau_0$ the maximum admissible time-delay that the information contained in the data can be used for effective decision making or for using it in another way. Then the data actuality factor can be defined as

$$v_{act} = \tau_0 / \tau, \quad \tau > 0. \tag{2}$$

where $\tau$, $\tau > 0$, denotes the time of data using.

It can be easily proven that so defined $v_{act}$ is a non-negative real number such that $v_{act} < 1$ for $\tau < \tau_0$ and it decreases to 0 while $\tau$ is infinitely increased.

b/ Relevance.

It describes a correspondence between the data contents and the user's expectations or needs. For single data $v_{rlv}$ can be expressed in a finite scale of relevance factor values $[0,1,...,r]$, $r$ being a natural number $\geq 1$, where 0 corresponds to a minimal and $r$ to a maximal relevance. However, there is no exact way of assigning the $v_{rlv}$ values to given data excepting the ones based on the intuition of data users.

c/ Irredundancy.

This factor expresses the lack of needless information elements in the data. In numerical data it may mean the lack of superfluous digital positions in data code, in monochromatic visual data – only a necessary number of grey-levels, etc. If $N$ denotes a total number of information elements in a given datum expression and $N_0$ denotes the minimum necessary number of such elements then we can put:

$$v_{irr} = N_0 / N. \tag{3}$$

Of course, $v_{irr} = 0$ if all information elements are superfluous and $v_{irr} = 1$ if no such elements in the data occur.

d/ Accuracy.

This property corresponds to non-redundant elements of numerical or graphical data only. If $\xi$ denotes numerical data describing a real (physical, economical, etc.) continuous parameter whose real value with probability $p$, $0 < p \leq 1$, is contained in a confidence interval $[\xi - \delta, \xi + \delta]$, where $\delta > 0$, and $D$ denotes the length of numerical interval in which $\xi$ can be contained then we can put

$$v_{acc} = p \cdot D / 2\delta. \tag{4}$$

So defined data accuracy can take values from 0 to infinity; higher accuracy can be reached by narrowing the confidence interval or by extension of admissible data-values interval $D$.

e/ Credibility.

This notion corresponds mostly to non-numerical data. It describes a level of confidence assigned to some statements about facts, qualities of objects, etc. Like relevance, the credibility factor $v_{crd}$ can be expressed in a finite scale [0,1,...,$c$], $c$ being a natural number $\geq 1$. The values of $v_{crd}$ can be assigned to data on intuitive basis or on a general confidence level assigned to a given data source.

f/ Completeness.

The notion of single data completeness is limited to those cases only when multi-component data are considered. Such data are complete if the values of all their components are available. For example, if data describe the values of pixels of a colour image in *RGB* (*red-green-blue*) or in *HSV* (*hue-saturation-value*) representation then a lack of some data component, say, *blue* or *saturation* makes possible image visualisation preserving geometrical forms but in non-adequate colours. If $m$ denotes the number of available data components and $m_0$ the desired number of such components, $0 \leq m \leq m_0$, then we can put:

$$v_{cpl} = m / m_0 \tag{5}$$

So defined completeness factor takes values from the interval [0,...,1].

g/ Legibility.

It is connected with the fact that some data cannot be used directly, before being submitted to some pre-processing: re-calculation of numerical scale, reformatting, reinterpretation, etc. Additional data pre-processing requires additional data processing time which can be compared with the time of data acquisition in the case when data can be directly used. Let us denote the last time by $T$ and by $\Delta T$ the additional one, both being given by some real non-negative numbers. Then we can put:

$$v_{leg} = T / (T + \Delta T). \tag{6}$$

So defined legibility factor takes values within the interval [0,...,1].

The costs of data usually are, in general, a non-decreasing function of data quality factors. In addition, the quality factors are not quite independent each on each other one. The relationships among them cannot be, in general, described analytically. However, they have approximately the forms plotted in Fig. 1 a ,b,c.

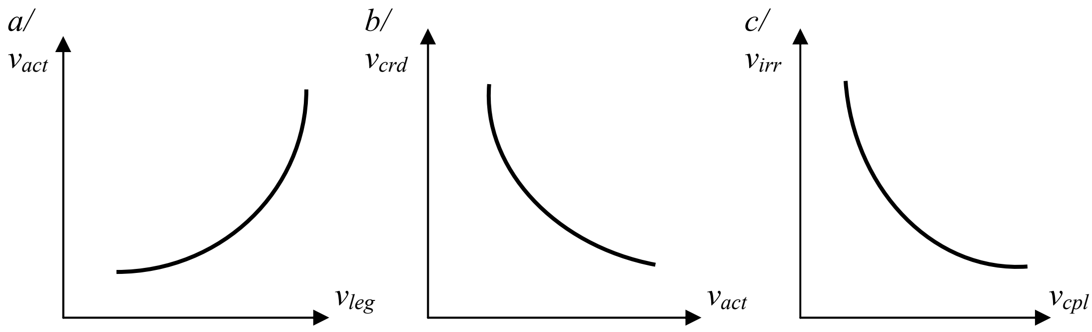


Fig. 1. Typical relationships among quality factors.

If data are highly legible they don't need additional processing and so, their actuality increases. On the other hand, data of high actuality may be, sometimes, less credible than those ones whose credibility need more time for being tested. Irredundant data may be incomplete if they have been compressed by bad-quality information-loss compression algorithms, etc.

## 4. Quality factors of structured records and files

It will be assumed that a structured record consists of a finite sequence of data whose quality factors are given. Then it arises the problem of the whole record quality characterisation. For this purpose all the above-mentioned quality factors will be used, and their values will be defined as some positive non-decreasing functions of the corresponding quality factors of the component data. We shall denote by $w_{...}$ with the corresponding subscripts the quality factors of structured records. Therefore, we are looking for the functions

$$w_{...} = f(v^{1}_{...}, v^{2}_{...}, \dots , v^{n}_{...}) \tag{7}$$

real, positive and non-decreasing with respect to their arguments. It is also desirable that the values of *f* are not lower than its minimum and not higher than its maximum argument value. There is a large variety of such functions. However, it seems desirable to chose such ones that are meaningful from the applications point of view. We shall take into account three variants of such functions: the *minimum,* the *maximum*, and the *weighted mean value* of the arguments. The *minimum* variant leads to a "careful" philosophy of data using: if a record satisfies user's expectations then all its components do so as well. The *maximum* follows from an "optimistic" philosophy: when a record satisfies user's expectations as a whole then they are satisfied by at least one record's component. The *weighted mean value* variant takes the opportunity of record's components differentiation. Therefore, the general formula (7) can take one of the following forms:

$$w_{...} = min(v^{1}_{...}, v^{2}_{...}, \dots , v^{n}_{...}), \tag{8}$$

$$w_{...} = max(v^{1}_{...}, v^{2}_{...}, \dots , v^{n}_{...}), \tag{9}$$

$$w_{...} = \sum_{\mu=1}^{n} \alpha_{\mu} \cdot v^{\mu}_{...} \tag{10}$$

where the subscripts $_{...}$ may take the values: *act, rlv, irr, acc,crd, leg* and $\alpha_{\mu}$ are some non-negative weight coefficients such that $\Sigma_{\mu}\alpha_{\mu} \equiv 1$. If we put $\alpha_{\mu} = 1/n$ for all $\mu$ then the weighted mean value becomes an ordinary arithmetical mean value.

Several words should be said about the completeness factor $w_{cpl}$ of records. It is not enough to calculate it directly as a minimum, maximum or mean value of the completeness factors of component data. The record as a whole (say, when imported from a remote database) may contain not all components required by the user. Therefore, if we denote by $v^{\nu}_{cpl}$ the completeness factor of $\nu$th component data, where $\nu$ corresponds not only to the data components existing in the given record but also to the lacking ones, one should put $v^{\nu}_{cpl} = 0$ for the last ones and, only then, a mean value of $v^{\nu}_{cpl}$ -s taken over all $\nu$-s can be taken as a completeness factor of the given record.

The quality factors of structured files can be characterised, in similar way, as the minimum, maximum or weighed mean values of the quality factors of records' components.

Finally, using the above-described approach we are able to evaluate the quality of single data, of structured records consisting of data sequences and of files consisting of sequences of structured records as well. In all cases the data structures' quality is given by the vectors (1) of quality factors.

## 5. Comparison of data structures' qualities

It will be assumed that the user, in order to create his proper database, takes into account several possibilities of supplying the database with data structures imported from several remote databases. In such case he is interested in evaluation and comparison of the data values in order to chose the best variant of database creation and maintaining. From a formal point of view this leads to a problem of vectors' ordering in a multi-dimensional space.

The problem can be solved on the basis of linear semi-ordered vector spaces' (*K-spaces,* see Appendix 1).

There is a constructive method of a *K*-space construction based on a concept of *positive cone* $K^+$.It is based on the following assumptions:

1/ There is described in a linear vector space $X$ a cone $K^+$, $K^+ \subset X$, of *positive vectors* such that if there are given two vectors $\boldsymbol{x}, \boldsymbol{y} \in K^+$ and real numbers α, β then also $\alpha\cdot \boldsymbol{x} + \beta\cdot \boldsymbol{y} \in K^+$.

2/ If there are given any two vectors $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$ such that $\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{z}$ and $\boldsymbol{z} \in \mathrm{K}^+ \cup \theta$ then it is called that $\boldsymbol{x}$ *is not smaller than* $\boldsymbol{y}$ and this relation is shortly written as $\boldsymbol{x} \succ \boldsymbol{y}$ or as $\boldsymbol{y} \prec \boldsymbol{x}$ ($\boldsymbol{y}$ *is not greater than* $\boldsymbol{x}$).

A typical example of a positive cone $K^+$ (in a two-dimensional space $X$) is shown in Fig. 2.

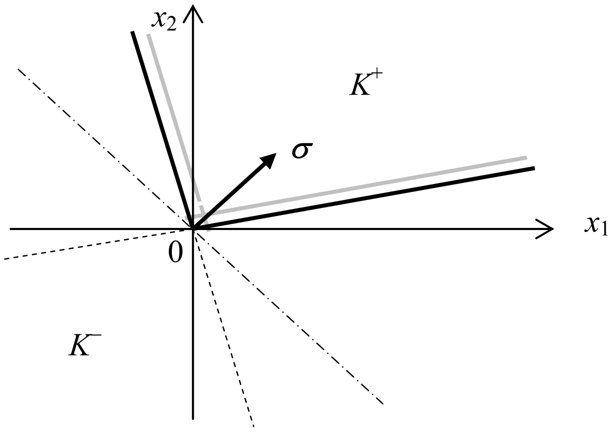


Fig. 2. Positive ($K^+$) and negative ($K^-$) cones in a two-dimensional *K*-space.

Let us define a *standard positive vector* $\boldsymbol{\sigma}$ in a *n*-dimensional space as a one having all positive and equal components and the norm equal to 1:

$$\boldsymbol{\sigma} = [s,s,\ldots,s], \tag{11}$$

where $s = n^{-1/2}$ . Then, in the simplest case, we can define the positive cone $K^+$ as the one consisting of all vectors $\boldsymbol{v} \neq \theta$ and such that the angle $\angle\, (\boldsymbol{v}, \boldsymbol{\sigma}) \leq \gamma$, $0 \leq \gamma < \pi/2$. Otherwise speaking, if

$$\beta = cos\, \gamma \tag{12}$$

then the vectors $\boldsymbol{v}$ belonging to $K^+$ should satisfy the inequality:

$$\frac{(\mathbf{v},\sigma)}{\|\,\mathbf{v}\,\|} \geq \beta \tag{13}$$

where (*,*) denotes a scalar product of vectors while ||*|| denotes a vector's norm. We also have used here the fact that $\|\boldsymbol{\sigma}\| = 1$.

The positive cone $K^+$ in this case has a circular symmetry around $\boldsymbol{\sigma}$, as shown in Fig. 3a.

Taking into account an extended form of $\boldsymbol{v}$:

$$\boldsymbol{v} = [v_1, v_2,...,v_n] \tag{14}$$

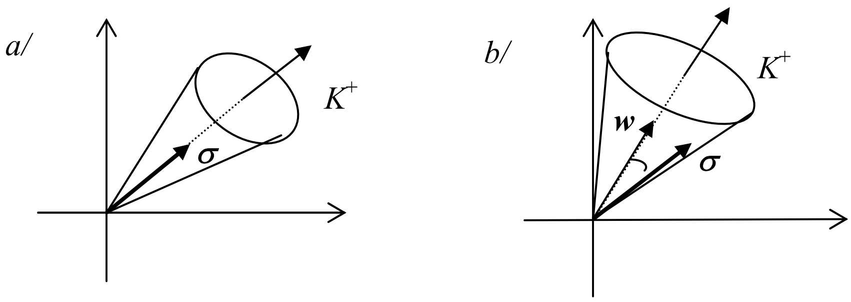


Fig. 3. Different types of positive cones: a/ axial with respect to $\boldsymbol{\sigma}$, b/ asymmetrical.

from (13) we obtain an inequality that should be satisfied by the components of $\boldsymbol{v}$, $\boldsymbol{v} \in K^+$:

$$\frac{\sum_{\mu=1}^{n} v_\mu}{(\sum_{\mu=1}^{n} v_\mu^2)^{1/2}} \geq \frac{\beta}{s} \tag{15}$$

Before going to a description of a more general classes of positive cones let us remark that if $K^+_1$, $K^+_2$, are two such cones then due to their convexity an intersection $K^+_1 \cap K^+_2$ is a positive cone, as well.

Let us denote by $\boldsymbol{w}$ a vector such that $\|\boldsymbol{w}\| = 1$ and $\angle(\boldsymbol{w}, \boldsymbol{\sigma}) < \pi / 2$. Therefore, its components, as follows from (13) and from the fact that $\beta = cos(\pi / 2) = 0$, should satisfy the inequality:

$$\sum_{\mu=1}^{n} w_\mu \geq 0 \tag{16}$$

Then, we can describe a circular cone $K^+$ whose axis is indicated by $\boldsymbol{w}$, as shown in Fig. 3 b; we call vector $\boldsymbol{w}$ an *indicator* of the circular cone $K^+$. Therefore, a larger class of positive circular cones is given by their indicators $\boldsymbol{w}$ satisfying the inequality (16) and vertical angles $\gamma$ not exceeding $\pi / 2$. In particular, if for a given $\boldsymbol{w}$ there is $\gamma = \pi / 2$ then the circular cone takes the form of a *hemi-space* (a hemisphere of infinite radius), as shown in Fig. 4.

The hemi-space indicated by $\boldsymbol{w}$ ***wi**ll* be denoted by $S_w$. Excepting the case when $\angle(\boldsymbol{w}, \boldsymbol{\sigma}) = 0$ such a hemi-space considered as a cone does not satisfy the formerly given general conditions of positive cones. However, taking into account an intersection $S_\sigma \cap S_w$ we obtain a positive cone $K^+$. Vectors $\boldsymbol{v}$ belonging to it and having a norm $\| \boldsymbol{v} \| > 0$ should satisfy the following inequalities:

$$\sum_{\mu=1}^{n} v_\mu > 0 \tag{17a}$$

$$\sum_{\mu=1}^{n} v_\mu \cdot w_\mu > 0 \tag{17b}$$

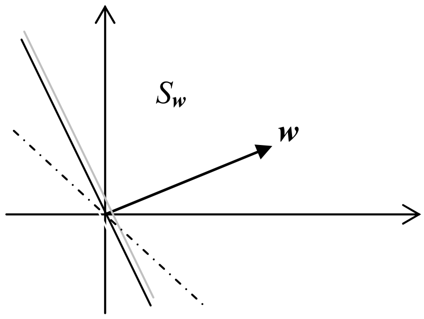


Fig. 4. A hemi-space $S_w$ indicated by a vector $\boldsymbol{w}$.

In the case when a positive cone $K^+$ is an intersection of more than two hemi-spaces $S_w$ corresponding to $\sigma$ and a finite set of indicators $\boldsymbol{w}^{(\beta)}$, $\beta = 1,2,...,b$, as shown in Fig. 5, vectors $\boldsymbol{v}$ belonging to it should satisfy the inequality (17a) and a series of inequalities:

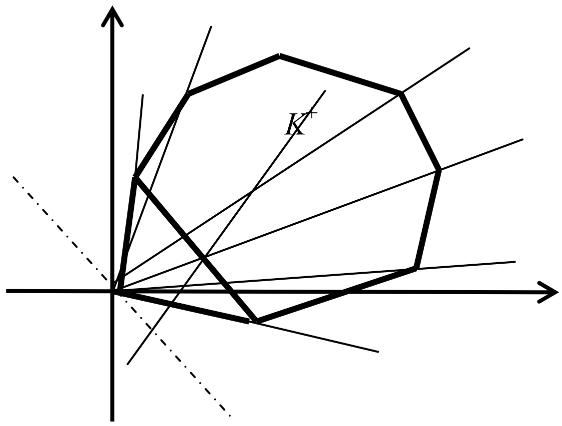


Fig. 5. A positive cone bounded by a set of hyperplanes.

$$\sum_{\mu=1}^{n} v_\mu \cdot w_\mu^{(\beta)} > 0, \quad \text{for } \beta = 1,2,...,b. \tag{17c}$$

The above-described formalism will be used to an evaluation and comparative examination of data quality. Let us assume that $X$ is a linear space whose elements $\boldsymbol{v}$ are multi-component vectors describing quality of some data structures. Let $\boldsymbol{v}^{(i)}$ and $\boldsymbol{v}^{(j)}$ be two vectors describing the qualities of two data structures that are to be compared. Then the problem, which (if any) of them is higher than the other one can be solved using the concept of semi-ordering of the vector space. The solution consists of the following steps:

1. Assign quality factors to the components of vectors $\boldsymbol{v}$ in $X$.
2. Define a positive cone $K^+$ in $X$ choosing a set of indicators $\boldsymbol{w}^{(\beta)}$, $\beta = 1,2,...,b$ and formulating a set of $b$+1 inequalities of the type (17a) and (17c).
3. Calculate the difference of vectors:

$$\Delta_{ij} = \boldsymbol{v}^{(i)} - \boldsymbol{v}^{(j)} \tag{18}$$

4. Check if $\Delta_{ij}$ satisfies the inequalities (17a) and (17c) (i.e. if $\Delta_{ij} \in K^+$). If so, then $\boldsymbol{v}^{(i)} \succ \boldsymbol{v}^{(j)}$, otherwise:

5. Check if (-1)· $\Delta_{ij}$ satisfies the inequalities (17a) and (17c). If so, then $\boldsymbol{v}^{(i)} \prec \boldsymbol{v}^{(j)}$, otherwise $\boldsymbol{v}^{(i)}$ and $\boldsymbol{v}^{(j)}$ are mutually *incomparable*.
6. STOP.

A decision "$\boldsymbol{v}^{(i)} \succ \boldsymbol{v}^{(j)}$" means, of course, that $\boldsymbol{v}^{(i)}$ is preferred with respect to $\boldsymbol{v}^{(j)}$ or the corresponding data structure is of higher quality than the other one.

## 6. Final comments

The concept of *K*-space offers a large variety of semi-orders that can be imposed on the sets of vectors representing data structures' quality. In particular:

a/ assuming that $K^+$ is traced by a single vector $\boldsymbol{\sigma}$ and $\beta = 1$ (see (11)-(13)) one obtain the strongest way of semi-ordering requiring domination of the preferred quality-vector over the alternative one in all quality factors;

b/ on the opposite side one can put $K^+$ being traced by a single vector $\boldsymbol{\sigma}$ and $\beta = 0$. In this case the preferred quality-vector should dominate at least in one quality factor over the alternative one, the rest quality factors being pair-wise comparable.

c/ In a more general case it is possible to narrow the positive cone $K^+$ by arbitrary choosing the admissible angle $\angle(\boldsymbol{\sigma}, \boldsymbol{v})$. For this purpose, according to (12), we should put $0 < \gamma < \pi/2$; the larger is $\gamma$ the less restrictive is the preference rule and the lower is the rate of mutually incomparable pairs of vectors.

d/ Using additional inequalities (17c) one reach the possibility of adjusting the vector preference rules to practical data quality requirements. As an example let us take into account the relationships shown in Fig.1. Looking at Fig. 1 b it seems reasonable to require that not only actuality and credibility are maximised separately but also their weighed sum should be maximised because of their mutual opposition. For similar reasons, a weighed sum of completeness and irredundancy should be maximised, as it follows from Fig. 1 c. Therefore, we obtain the requirements:

$$p \cdot v_{act} + (1-p) \cdot v_{crd} > b_1, \quad 0 < p < 1, \quad 0 \le b_1, \tag{19a}$$

$$q \cdot v_{cpl} + (1-q) \cdot v_{irr} > b_2, \quad 0 < q < 1, \quad 0 \le b_2. \tag{19b}$$

Here $p$ and $q$ are some relative weights assigned to $v_{act}$ and $v_{cpl}$, correspondingly, with respect to $v_{crd}$ and $v_{irr}$. The vectors

$$\boldsymbol{w}^{(1)} = [p,0,0,0,1-p,0], \tag{20a}$$

$$\boldsymbol{w}^{(2)} = [0,0,1-q,0,q,0] \tag{20b}$$

thus the role of indicators of the $K^+$ bounds while $b_1$ and $b_2$ are threshold parameters restricting the angles between indicators and the differences of quality vectors. In similar way, higher-order (multi-argument) constraints on the differences of quality vectors can be imposed.

## Appendix 1

## Kantorovitsch spaces (*K*- spaces)

The concept of *K*-spaces has been introduced by L.V. Kantorovitsch et al. [4].

Roughly speaking, *K*-space is a linear vector space $\boldsymbol{X}$ with a null-element $\theta$ and introduced ordinary operations of vectors' addition and multiplication by real numbers. In addition, it is assumed that the vectors are semi-ordered in the sense that for certain pairs of vectors $\boldsymbol{x}, \boldsymbol{y} \in X$ it holds at least one of the relationships: $\boldsymbol{x} \prec \boldsymbol{y}$ (read: ***x** is not greater than **y***) or $\boldsymbol{y} \prec \boldsymbol{x}$. Both: $\boldsymbol{x} \prec \boldsymbol{y}$ and $\boldsymbol{y} \prec \boldsymbol{x}$ hold if and only if $\boldsymbol{x} = \boldsymbol{y}$.

The relation $\prec$ should satisfy the following conditions:

1/ If it is $\boldsymbol{x} \prec \boldsymbol{y}$ and $\alpha$ is a positive real number then $\alpha\cdot \boldsymbol{x} \prec \alpha\cdot \boldsymbol{y}$;

2/ if it is $\boldsymbol{x} \prec \boldsymbol{y}$ and $\alpha$ is a negative real number then $\alpha\cdot \boldsymbol{y} \prec \alpha\cdot \boldsymbol{x}$;

3/ if it is $\boldsymbol{x} \prec \boldsymbol{y}$ and $\boldsymbol{u} \prec \boldsymbol{v}$ then $\boldsymbol{x} + \boldsymbol{u} \prec \boldsymbol{y} + \boldsymbol{v}$.

A pair of elements $x, y \in X$ is called *mutually incomparable* if neither $\boldsymbol{x} \prec \boldsymbol{y}$ nor $\boldsymbol{y} \prec \boldsymbol{x}$.

Next property of a *K*-space is connected with a notion of *supremum* of a bounded subset $A \subset X$. It is called that the subset has an *upper bound* $\xi$ if $\xi \in X$ is such that for any $x \in A$ , $x \neq \xi$, there is $x \prec \xi$.

The element $\xi$ is called a *strong upper bound* of *A* (a *supremum*, *sup A*) if it is an upper bound of *A* and there is no other upper bound $\zeta$ of *A* such that $\zeta \prec \xi$.

It is assumed in the definition of a *K*-space that each subset $A \subset X$ having its upper bound has its *sup A*.

The definitions of upper bound of *A* and of its strong upper bound can be easily used, by analogy, to the notions of a *lower bound* and of a *strong lower bound* (an *infimum, inf A*).

An element $\xi$, $\xi \in A$, is called *a maximum of A* if there is no other element $x$, $x \in A$, such that $\xi \prec x$. A subset *A* may have, in general, more than one *maximum.* If $\xi$, $\zeta$ are some *maxima* of *A* then it is neither $\xi \prec \zeta$ nor $\zeta \prec \xi$. Otherwise speaking, any two (or more) *maxima* of the given set *A* are *mutually incomparable.*

### References


1. Kulikowski J.L. *Notes on the Value of Information Transmitted through a Communication Channel.* In: 2nd International Symposium on Information Theory, Cachkadsor, ASSR, 1971. Conference Proceedings, Akademiai Kiado, Budapest, 1972.

2. Kulikowski J.L. *Value of Information in Computer Systems* (in Polish). Institute of Computer Sciences PAS, Reports No 472, Warsaw, 1982.

3. Kulikowski J.L. *Backgrounds of Data Value Evaluation* (in Polish). In: Technology and Methods of Distributed Data Processing, Part 1. Vroclav University of Technology, Vroclav, 1986.

4. Kantorovitsch L.V., Vulich B.Z., Pinsker A.G. *Funkcyonalnyj analiz v poluuporyadočennych prostranstvach* (in Russian). GITTL, Moscow, 1950.