

# Introduction to NumericSearch

---

## 1 Abstract

Human language is optimized to oral communication, not to digital communication, which allows much more. The domain "language vocabulary" as common set of possibilities (digitally selected by number sequences) can be extended substantially: The domain of **any piece of digital information (number sequence)** can be defined online. This is fully adaptable to the application and globally precisely comparable and searchable according to freely definable criteria, which by far exceeds the capabilities of language based search, where reality's features are categorized by words. All features can be quantified and represented more precisely by identified feature vectors or sequences of numbers or "Domain Vectors" (DVs), see <https://www.mdpi.com/1660-4601/17/8/2975/htm> . DVs are elements of user defined Domains of information which are metric Spaces and which are called "Domain Spaces" (DSs).

<http://NumericSearch.com/> demonstrates these in a local database in which every DS and all of its DVs are identified by a unique "Uniform Locator" (UL) which is its index (i7). Also its first keyword is unique and so could be used as UL.

(In a later web standard every DS and all of its DVs are identified by an "UL". The UL is a "Uniform Locator"<sup>1</sup> of the definition (the DS definition defines also all of its elements resp. DVs). So in a later web standard every DV has the form

"UL plus number sequence"

where the UL efficiently localizes the standardized online definition of the numbers resp. DV. *The UL is locator and also identifier like here the UL* (Uniform Locator) which identifies all DVs of the same DS and so makes these accessible to detailed similarity search. )

DS search consists of two systematic steps: 1. Selection of the appropriate DS e.g. by search of the ULs. 2. Similarity search of DVs within the selected DS or a part of it.

The technical potential of this combination is great. DS based description and search has larger range with much higher resolution and precision than usual word based description and search. It can handle an increasing number of very different DSs, defined by the users, in one and the same application. The first implementation is <http://NumericSearch.com/>

## 2 Motivation

A DV describes a resource. Its UL unambiguously localizes the containing Domain Space (DS), and by a sequence of numbers which (partially) represent an element (a vector) of this DS. DS based representation of resources and *DS search* promote precise communication.

---

<sup>1</sup> The "UL" has similar function like the well known "URL" (Uniform Resource Locator), but allows for maximal efficiency. To clarify this, the new term "UL" (Uniform Locator) has been introduced.

Similarity search on metric spaces or *Domain Spaces* (DSs) is a well known tool to find precisely resources in large databases. In *DS search* many DSs are accessible for similarity search. If there is no appropriate DS, it can be defined quickly. The number of DSs and the number of DVs (vectors) per DS can become very large (only limited by hardware) and every DS has an own well defined metric (distance function) for quantification of similarity. So independent of individual preferences<sup>2</sup>, from the technical point of view searching in DSs (which are numeric spaces) has important advantages. For the complete DS only one definition is necessary which can be very precise (more precise than usual word definitions). Then all DVs in this DS are well defined. There can be much more different DVs than words. To handle different topics<sup>3</sup>, the advantages of both concepts can be combined: A globally unique UL (like here the UL) for identification of the DS and so rough selection of the topic to cover the range of words already in the first step - and within every DS search in high resolution using its individual well defined metric for sorting the search result.

So instead of searching in sets of words we can search in sets of DSs which have larger range with much higher resolution and precision. More has been described e.g. in <http://numericsearch.com/wvdomainspaces.pdf> . Here this idea is realized on a server system.

The in <http://subs.emis.de/LNI/Proceedings/Proceedings176/469.pdf> recommended "Domain Vectors" (DVs) are implemented (index i4) and sometimes shortly called "vectors".

### **3 Data structures, operating instructions**

First some knowledge of the used data structures is necessary:

Every Domain Space (DS) contains as elements the DVs which are vectors. In the general approach every dimension can represent again a DS, so nesting is not limited. In this implementation for demonstration purposes the nesting level is fixed to 2, i.e. every DV represents combinations of "subvectors" to demonstrate partition and combinations of distance functions.

For systematic representation of DSs, DVs, subvectors, dimensions (and also users) the following indexes are used:

i0: index of a dimension within a subvector

i2: index of a subvector within a vector (DV)

---

<sup>2</sup> Up to now searching with numbers is unusual, but it's just a matter of practice. For DV search first of all brain needs to remember combinations of numbers with words (the vector space names and dimension names or dimension order), calculations are only necessary for advanced applications (example: "car 2p, 40g" as abbreviation of "car for 2 persons, with emission 40g CO2/km "). Such combinations of a word (vector space name "car") with numbers (precise description) are typical arguments for DV search. This can be used on the web also for ordinary purposes, not only for advanced applications. For advanced applications it is possible to provide complex search arguments and DVs by software.

<sup>3</sup> The range of topics is underestimated. For example values and results of surveys can be dimensions, which e.g. allows to increase comparability of law cases.

[i4](#): index of a DV within a DS

[i7](#): index over all DSs

[iu](#): index over all users

Clicking on the index symbols opens a table over the index, e.g. clicking on [i7](#) opens a table over all DSs.

Subsequently data structures are shown with exemplary content. This manual started in January 2012, please don't expect to find these examples in the current database, the content of the database is changing due to dynamic access. Also software updates can lead to changes of appearance.

A data structure called "keycomment" is used for short description of dimensions, subvectors, DVs, DSs and also users. It consists of one or more keywords and a comment. The first keyword ("keyword 0" or short "kw0") is most important; it is the topic of the comment (in case of a keycomment of a DS it represents the UL). Every keyword can be associated with an URL, i.e. every keyword can serve also as hyperlink. The total size of every keycomment is limited (currently to 2000 chars). The usual length should be shorter. Keycomments should contain compressed information to allow quick overview for reading users. Details can be made accessible via hyperlinks ("Link", see figure 1), if necessary.

Keyword:		Link:	
<input type="text" value="space-10000"/>	<input type="button" value="A"/>	<input type="text" value="http://url-with-descriptions-to-space-10000"/>	
<input type="text" value="keyword2"/>	<input type="button" value="A"/>	<input type="button" value="X"/>	<input type="text"/>
<input type="text" value="keyword3"/>	<input type="button" value="A"/>	<input type="button" value="X"/>	<input type="text"/>
Comment:	<input type="text" value="Comment to space 10000"/>		

**Figure 1: A keycomment in its input mask. Clicking on "A" appends a new line which can contain a further keyword, "X" deletes it. Deletion of the first keyword (kw0) is not possible.**

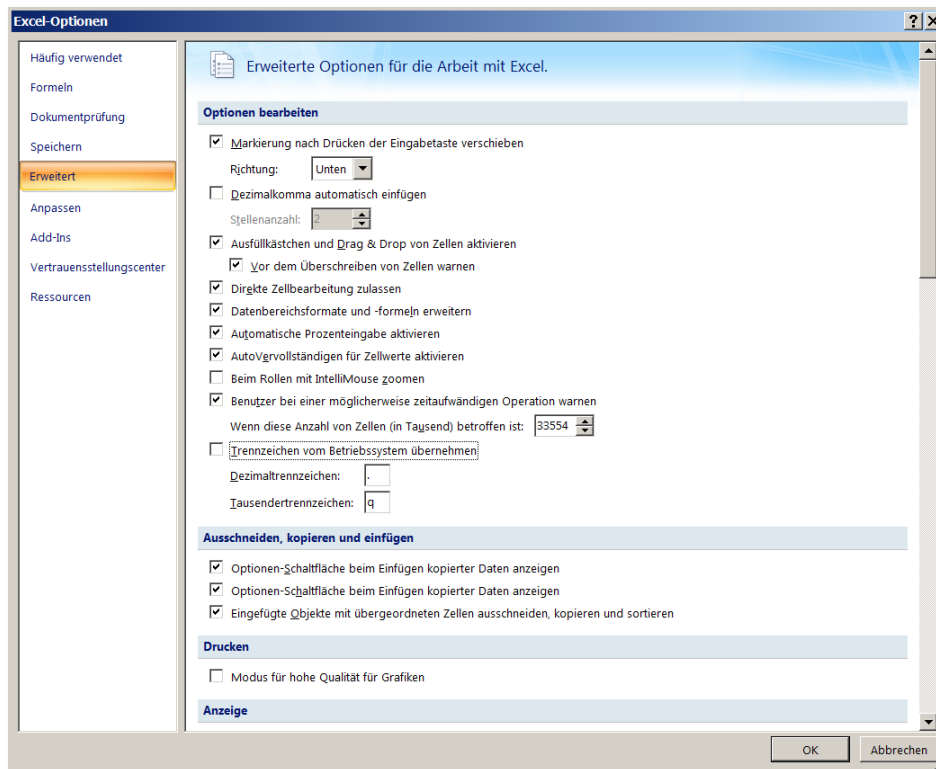
[space-10000](#) [keyword2](#) [keyword3](#) / Comment to space 10000

**Figure 2: In listings shown short representation of the keycomment given in figure 1.**

The content of keycomments and other descriptions are determined by the owners of the described DVs or DSs. Owners are registered users who provide DVs or define DSs. After registration you can make own contributions. Search is possible also without registration.

### **3.1 Decimal separator**

NumericSearch allows export .csv format, so that the data can be used in Excel. To allow international exchange of data we decided to use "." as decimal separator and we do not recommend to use thousands separator. We tested it with the following options in Excel:



The following letters are used as separators and should be avoided: ; | < >

### **3.2 Text search**

[{Kw0Search}](#) Text search is not focused, but search within the first keywords (keyword 0, kw0) is implemented (an additional reason to select a meaningful keyword 0). You can enter the (start of a) keyword to search it in kw0. This text search can be used for example to find a DS by its UL.

### **3.3 Numeric search**

Not all possibilities are explained in this manual. You can collect much information most efficiently by clicking on blue text and checking the result.

After selection of the DS (by clicking on "i7" and in the opened table on the DS number) numeric search, i.e. precise similarity search over the DVs of this DS is possible and the search mask appears.

enter numbers for similarity search (into the "sim" fields)

and/or enter the range (using the "min" and "max" fields)

\* i4 \* iu kw0 login reg own us home

---

\* i7=1001', o | 2013-02-09 ride

### NumericSearch in DS 1001 (ride) search-stat DS-stat

< << < > >> >| 0..2

	sim	min	max	g
0				gps-coordinates-start
0	54.32,10.12			<input type="checkbox"/> latitude   degree
1				<input type="checkbox"/> longitude   degree
1				gps-coordinates-destination
0	53.54			<input type="checkbox"/> latitude   degree
1	9.98			<input type="checkbox"/> longitude   degree
2				data
0	2014-07-31			<input type="checkbox"/> start-time   yyyy-mm-dd hh:mm
1	1.1			<input type="checkbox"/> duration   hours
2				<input type="checkbox"/> price-per-person   euro
3	3			<input type="checkbox"/> max-count-of-persons
4				<input type="checkbox"/> vehicle-load-capacity

o:  w:  pcnt:

**Figure 3: Search mask; here search is done within DS 1001 with name ride. The comma is interpreted as separation, so values in one input field can be given also for the following fields of the same column. Clicking on "search" starts DS search.**

At this for every dimension the searched numeric value can be given and/or minimum and maximum. For the search only those DVs are taken into consideration, which contain values at the searched dimensions. Therefore it is recommendable for the owner of a DV to provide as much as possible dimensions values to maximize the chance that the DV can be found in numeric search.

It is usually recommendable to make multidimensional similarity search only over fields (s="similarity") which are closely related (e.g. GPS coordinates<sup>4</sup>). If this is not the case, e.g. if the fields have different units, it can be difficult or impossible for the owner of the DS to estimate the relative weight of the fields reliably when defining the distance function. In this case it is recommendable to use for search the s field only once (or for closely related dimensions) to order the search result, and else the min max fields (figure 3).

For quick overview we now explain a search example on a DS of cupboards.

<sup>4</sup> From <http://maps.google.com/>, after right click and click on "What's here?", GPS coordinates for search and for resources can be directly copied and pasted into the latitude field as visible in figure 3.

enter numbers for similarity search (into the "sim" fields)

and/or enter the range (using the "min" and "max" fields)

\* i4 \* iu kw0 login reg own us home

\* i7=1006', o | 2013-02-09 Cupboard | Schrank

## NumericSearch in DS 1006 (Cupboard)

|< << < > >> >| 0..1

	sim	min	max	g
0	<input type="text"/>	<input type="text"/>	<input type="text"/>	Finances
0	0.00	<input type="text"/>	<input type="text"/>	<input checked="" type="checkbox"/> Price   Euro
1				Size
0				<input checked="" type="checkbox"/> Width   cm
1				<input type="checkbox"/> Depth   cm
2				<input type="checkbox"/> Height   cm

o:  w:  pcnt:

**Figure 4: Search of the cheapest cupboard (price nearest to 0) with graphic output of price in dependence of width.**

Figure 4 shows the search of the cheapest cupboards (those nearest to price 0). The checkboxes "o" for "offer" or "w" for "wanted" are not checked, therefore within all DVs is searched (not only within those with checked "o" or "s"). Two checkboxes in column "g" are checked, so together with the search result also a graph over the checked coordinates is shown. Because no maximum pcnt is chosen, the graph of figure 5 contains all cupboards (if the database contains more than 1000, max. 1000 are shown).

search result

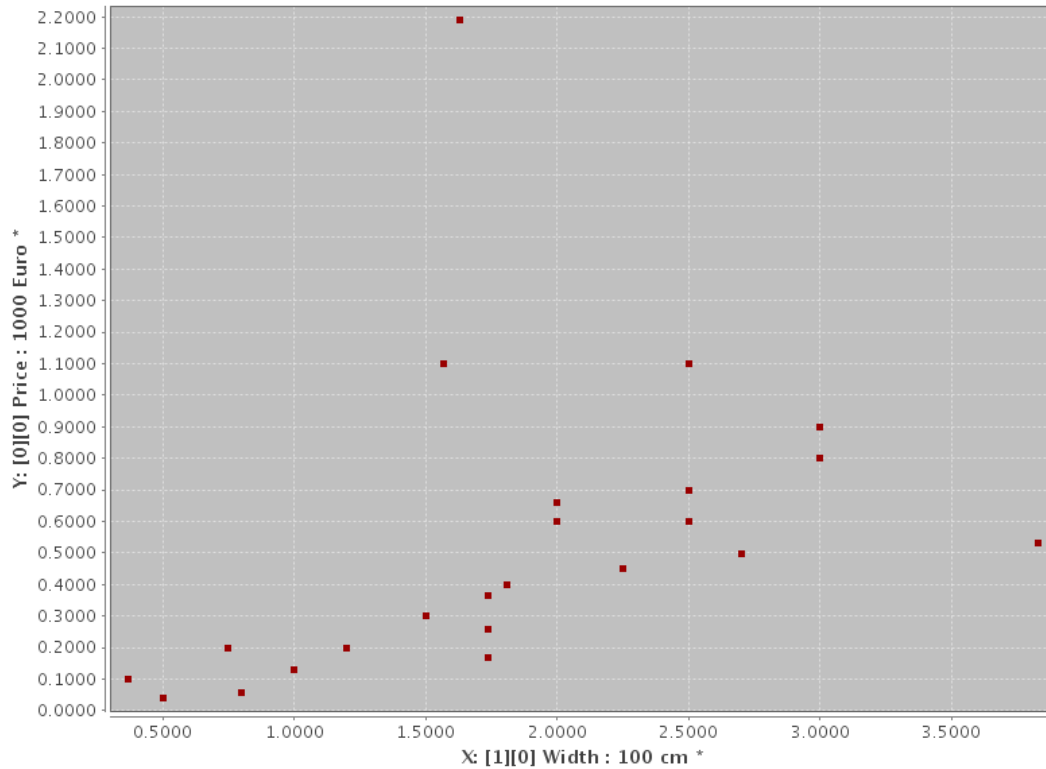
new search

repeat

par

search-stat TS-stat

24 of TS 1006 (Cupboard)



X: av= 184.875 sd= 84.77608 min= 37 max= 383  
 Y: av= 529.1829 sd= 467.6006 min= 39 max= 2189

< << < > >> >| page 1

i4	d	a
2	39	30 o ikea-PAX-schmal   39.00, 50, 60, 236
9	59	0 o ikea-IVAR   59.00, 80, 30, 83
1	99	24 o Ikea-PS   99.00, 37, 40, 190
10	130	16 o ikea-PAX    Schrank mit 2 Tueren, weiss, Ballstad weiss   130.00, 100, 60, 236
5	170	3 o ikea-IVAR    2 Elem/ Boeden/Schrank   170.00, 174, 30, 179
3	199	5 o ikea-ISALA   199.00, 75, 44, 131
4	199	11 o ikea-TROLLSTA    Sideboard   199.00, 120, 50, 76
8	258.9	12 o ikea-IVAR    2 Elem/ Boeden/Schrank   258.90, 174, 50, 179
22	299	0 o Home24-Austin    Schwebetuerenschrank - verschiedene Groessen - Weiss mit Mattglas   299.00, 150, 68, 216
0	362.9	8 o ikea-IVAR    2 Elem/ Schrank/Kommode   362.90, 174, 50, 179
6	362.9	2 o ikea-IVAR    2 Elem/ Schrank/Kommode   362.90, 174, 50, 179
11	399	14 o home24-Quadra    schwebetuerenschrank, Korpus alpinweiss/Front alpinweiss   399.00, 181, 58, 210
12	449	4 o home24-Quadra    Schwebetuerenschrank, Korpus alpinweiss/Front alpinweiss   449.00, 225, 58, 210
13	499	0 o home24-Quadra    Schwebetuerenschrank, Korpus alpinweiss/Front alpinweiss   499.00, 270, 58, 210
7	532.7	16 o ikea-IVAR    5 Elem/ Boeden/Schrank   532.70, 383, 50, 226
14	599	1 o home24-Mission-4    Schwebetuerenschrank - Alpinweiss, Abs, Pearlglanz Softwhite   599.00, 200, 65, 218
15	599	0 o home24-Mission-4    Schwebetuerenschrank - Alpinweiss, Abs, Pearlglanz Softwhite   599.00, 250, 65, 218
19	659	0 o home24-Lowisville    Schwebetuerenschrank - verschiedene Groessen Polarweiss/Weisslack   659.00, 200, 68, 216
20	699	1 o home24-Lowisville    Schwebetuerenschrank - verschiedene Groessen Polarweiss/Weisslack   699.00, 250, 68, 216
21	799	0 o home24-Lowisville    Schwebetuerenschrank - verschiedene Groessen Polarweiss/Weisslack   799.00, 300, 68, 216
16	899	0 o home24-Mission-4    Schwebetuerenschrank - Alpinweiss, Abs, Pearlglanz Softwhite   899.00, 300, 65, 218
18	1099	1 o home24-Maxi-Eleven    Kleiderschrank - verschiedene Varianten - Erle Massiv   1099.00, 157, 57, 203
23	1100	8 o RodrigoSchrank   1099.99, 250, 63, 223
17	2189	25 o home24-Rivoli    Kleiderschrank - mit Dekor - Fichte, antik lackiert, Standardeinteilung   2189.00, 163, 60, 197

387 th search in TS 1006 (Cupboard) , 24 (100%) within 24 found

new search

Figure 5: Search result following the search of figure 4. Data in every line:

i4 distance accesscount link-to-owner kw0 || keycomment | numeric content



Figure 5 shows the search result. The graph shows the dependence of price and width. Every line in the table contains most important data of every DV:

Blue number:  $i_4$ =DV index in DS

Green number:  $d$ =distance (if numbers have been given for similarity search)

Black number: access count of DV

Blue "o": link to owner

Text: keycomment of DV

| Numbers: numeric content of DV

Statistical information about the dimensions of the graph is shown together with the graph („av“ means average, „sd“ the standard deviation). The table with the search result lists all DVs with "dim0" of subvector 0 nearest to 0 (the cheapest are listened first, because they are nearest to 0). The green column shows the distances which are in this case simply the prices of the cupboards. The DS owner can determine the distance function (see below), additionally users can (after expansion of the search mask) change the weight of the dimensions for their search.

The definition of the distance function is part of the DS definition. This definition is accessible by clicking on the date on the short info about the DS (line after „i7“, see figure 6).

[i7=1006, o | 12-09-03 Cupboards/Schraenke](#)

**Figure6:** Excerpt from upper part of figure 4 or figure 5. Click on „i7“ opens the DS table, click on „o“ shows information about the owner, click on the definition date „12-09-03“ (yy-mm-dd) opens the DS definition window, [figure 12](#) shows the view for the owner.

### **3.4 Advanced search**

Advanced users can try additional search possibilities by clicking on min, max, |. Clicking on min or max fills the min-max text fields after first click with lower resp. upper 5% percentiles (maximal the 10000th resp.  $n-10000$ th value for DSs with size  $n > 200000$ ), further click with the minimum and maximum, further click restores the initial view. This can be used to quickly orientate about the range of values of every dimension and to quickly use this information. Because all min and max fields are automatically filled with values in this mode, it is necessary to check the checkbox below “p” at those dimensions, in which these values should be used for determination of a minimum or maximum for a search. In default mode this is not necessary, because in default mode no fields are automatically filled, therefore all fields with given values are (filled by the user and therefore) used for determination of the search result.

Clicking on the left | shows average values, clicking on the right | opens fields for determination of individual weight factors (multiplied with those given in the DS definition) of every dimension for a search (figure 7).



enter numbers for similarity search (into the "sim" fields)

and/or enter the range (using the "min" and "max" fields)

\* i4 \* iu kw0 login reg own us home

---

\* i7=1006', o | 2013-02-09 Cupboard | Schrank

### NumericSearch in DS 1006 (Cupboard)

< << < > >> >| 0..1

	sim	min	max	w	g
0	<input type="text"/>	<input type="text"/>	<input type="text"/>	1	Finances
0	<input type="text"/>	<input type="text"/>	<input type="text"/>	1	<input type="checkbox"/> Price   Euro
1	<input type="text"/>	<input type="text"/>	<input type="text"/>	1	Size
0	<input type="text"/>	<input type="text"/>	<input type="text"/>	1	<input type="checkbox"/> Width   cm
1	<input type="text"/>	<input type="text"/>	<input type="text"/>	1	<input type="checkbox"/> Depth   cm
2	<input type="text"/>	<input type="text"/>	<input type="text"/>	1	<input type="checkbox"/> Height   cm

o:  w:  pcnt:

Figure 7: Search mask of figure 4, but with weight fields (after clicking on the right "/").

The column below "g" contains a checkbox for every dimension. You can click on 1 or 2 of these to activate a 1D or 2D graphic representation of these dimensions in the search result. Try the possibilities to check the results.

### 3.5 Search data as string

It is possible to store the search order as string which is valid as long as the DS definition is unchanged. For this press twice on the leftmost pager arrow, then in the first input field in the first blue bar (below "sim") a string appears which you can copy, store in a text editor and paste later into this field to repeat the search order. At this upper fields overwrite lower, i.e. if you paste in this field search data, all following input will be overwritten by this. Pressing "repeat" in the search result window shows you the search order in readable form. It is recommendable to check this to avoid errors due to possibly changed (shifted) DS definitions.

### 3.6 Registration

Registration is necessary if you want to provide own contributions (DVs, DS definitions). The registration form opens after clicking on "reg" in the upper line. First it is necessary to provide obligatory data, after clicking on "check" the system proposes a first version of your keycomment. Your login-name is keyword 0 (kw0). Initially you can modify kw0 and with this your login name, after first logout this no more possible, i.e. kw0 becomes your permanent login name (figures 9 and 11).

Click on "check" to review your data. Check also the suggested login name, you can change it in this state, but no more after logout. If all is ok, click on "logout" and you get an email with your password. Please login again using this information to confirm your account.

### 3.7 Login

Click on “login” in the upper line to open the login form (figure 8)

i4 iu search kw0 reg own

---

i7=1002, o | 12-01-09' space2d

### Login

username :

password :

you can enter only your username and press  or [register](#) if you have no username

**Figure 8: Login form. The system recognizes “,” as separator, so using “,” as separator, you can copy and paste login-name and pin also in one textbox (of username or password – using the password textbox your input is unreadable on the screen) to save time.**

i2 search kw0 I logout[10003] up own icalc

---

iu=10003 12-02-07 login-name / I buy Messbares  
i7= 1000, | 12-02-07 space-of-spaces / v-1

### user 10003

Keyword:    Link:

Comment:

Text fields are published, if checked as visible (v),  
Obligatory fields are marked by \*

(v)

\*forename:

\*surname:

\*emailadr:

\*birthdate:   (yyyy-MM-dd)

\*country:

\*city:

postalcode:

street:

telephone:

**Figure 9: Exemplary user data after login. The first keyword (kw0) is your login name, the comment is optional. Login-name, forename, surname and birthdate are fixed after registration.**

It is recommendable that you then change your password. Forename and surname are visible by all, using the other checkboxes (figure 9) you can determine which

additional fields are published. You can also give additional information concerning email exchange to avoid misunderstandings. It is possible that you use a pseudonym. This may be appropriate e.g. for tests.

- optional published information for email exchange:
- I can answer all individual emails sent to me
  - I can read all individual emails sent to me
  - I have not enough time to answer all individual email questions
  - I can only answer less than 10% of all individual email questions
  - I cannot read all individual emails sent to me
  - I not allow individual emails via message window

**Figure 10: Additional email information. In case of own DSs the last option is not shown. You can set DSs free for sale if there is not enough time for supervision of these.**

If in figure 9 no checkboxes are selected, only minimal information is visible by all: these contain login name, forename and surname with optional comment (Figure 11). Publishing more information can be recommendable for more specific information exchange or to avoid misunderstandings, e.g. in case of frequent names like “john smith”.

i4 search kw0 login reg own icalc

---

iu=10003 [12-02-07](#) login-name / I buy Messbares  
i7= 1000, | [12-02-07](#) space-of-spaces / v-1

**user 10003**

Keyword:  Link:

Comment:

\*forename:   
\*surname:

**Figure 11: user data of figure 8 which are visible by all, if no additional checkboxes are selected.**

After login clicking on “own” and allows definition of own DVs or DSs. Owing is possible for test purposes.

### **3.8 Definition of a DS**

Recall: **Please download your definition (possible in XML form) so that you can restore the former version in case of error.**

In the current version changing of dimensions and even deletion is possible, but these options should be only used in draft state, for quick corrections. As soon as non-draft definitions are used, they must not be changed.

The definition window of a new DS appears after login, clicking on "own" and on "define a new DS". (The definition window of an already defined DS is accessible by clicking on the date in the short info about the DS (line after „i7“, see [figure 6](#) ) and is editable, if you are owner of the DS.)

Please contact us if you need help. In this initial state we offer free support. In the initial state owning of a DS name (with definition of a DS and determination of the links in the definition) is free for 6 months. Interesting DS definitions are welcome. If DSs are (nearly) unused over long time, these can be deleted by the system. Before defining a DS with certain keyword "newtopic", look into the DS table and use [\[Kw0Search\]](#) with "newtopic" (and synonyms) to find a DS to this keyword0. If you find such a DS with satisfying definition, don't define it again!

From the technical point of view it is desirable that DVs to one topic concentrate in one DS. Search is done over one DS (in the current version); this is well defined and efficient. It is contra productive to spread DVs to analogous topics over many DSs, because search in one DS would be incomplete. If there are multiple similar DSs to nearly the same topic, we recommend to use for own DVs and search in case of doubt the DS with the earliest definition.

If you think, the found definition is incomplete, i.e. some interesting dimensions or subvectors are missing, tell your suggestions to the owner of the DS. If his email address is not visible you can contact the owner after login and clicking on "email".

If you think that your ideas are significant and if the found definition cannot be completed according to your ideas, then consider definition of a new own DS.

The dimensions of the DS should be interesting (decision relevant) and mutually independent. Every dimension should be well defined and clear, reproducible at least by experts. Prefer most frequently used units, in case of doubt prefer decimal powers of [Si derived units](#) .

Figure 12 shows the DS of figure 4. It is a DS with 2 subvectors; one contains the price, the other measurements.

\* [iu search kw0 I logout\[10001\] up own us home](#)

---

\* [i7=1006', o | 2013-02-09 Cupboard | Schrank](#)  
 \* [i4= 0, o 2014-01-09 ikea-ivar || 3 Elem/Boeden/Schrank, Kiefer | 175.40, 258, 30, 124,](#)

**Definition of DS 1006 (Cupboard)** [owner](#) [Append subv](#) [Del last subv](#) [download-def](#)

< << < > >> >| 0..1

0	Finances
0	Price   Euro
1	Size
0	Width   cm
1	Depth   cm
2	Height   cm

Keyword:

Link:

Comment:

This is:  draft  ok  deprecated  
 Nested metric:  Manhattan  Euclidean  Maximum

DVs can be added into this DS by all users ([change](#))

**Figure 12:** Definition of the DS in which the search of figure 4 has been done. The state of the text boxes is read-write, because the owner of the DS is logged in.

We now show the definition of a DS like in figure 4.

First login, then click on “own” in the upper line and then under "own DSs" and "preparation" and then on “define own DS”. This opens a window with a minimal (one dimensional) DS definition. After typing a name (e.g. Cupboards/Schraenke) as first keyword (kw0) it looks like figure 13 (the index 1008 is the first free i7 of the database at time of the manual - meanwhile changes of appearance are possible).

\* [iu search kw0 I logout\[10001\] up own us home](#)

---

\* [i7=1006', o | 2015-01-22 Cupboard | Schrank](#)

### Definition of DS 1006 (Cupboard) [owner](#) [Append subv](#) [Del last subv](#) [download-def](#)

|< << < > >> >| 0..1

0	subv0
0	dim0
1	subv1
0	dim0

Keyword:      Link:

Comment:

This is:  draft  ok  deprecated  
Nested metric:  Manhattan  Euclidean  Maximum

DVs can be added into this DS only by you as DS-owner ([change](#))

**Figure 13: Start of a DS definition "Cupboards/Schraenke" with link for the first (kw0) field and clicking once on "Append subv"**

Because the DS which we want to define has two subvectors, it is necessary to click once on “append subvector”. To change the definition of a subvector we click on the number of the subvector (in figure 13 the “0” before “subv0”). Then opens a window which allows to change the definition of this subvector. Figure 14 shows it after typing "" as first keyword (deleting default name "subv0").

### Keycomment to Sub Vector definition

owner Append dim Insert subv Del subv

|< << < > >> >| 0..0

0	subv0
0	dim0

Keyword:  Link:

Comment:

Weight:

Metric:  Manhattan  Euclidean

Figure 14: Definition of the subvector of figure 13 after typing "" as first keyword (kw0) and submit.

Now it is necessary to define the dimensions. Clicking on the number of the dimension ("0" before "dim0" in figure 14) opens the dimension definition window. Figure 15 shows it after typing "Price/Preis" as first keyword and "Euro" as second keyword (unit). Every dimension window contains automatically two keywords: the first is kw0 (the topic) and the second is the unit. We decided for this convention, so that the unit of a dimension is reproducibly readable as second keyword.

### keycomment to dimension

Keyword:  Link:

Unit:  Link:

Comment:

Min:  Max:  Weight:

representation:  list  tux  
 integer  money  floating point: medium length  floating point: max. length

date in:  yyyy-mm-dd hh:mm:ss  yyyy-mm-dd hh:mm  yyyy-mm-dd hh  
 yyyy-mm-dd  yyyy-mm  yyyy  hh:mm:ss  hh:mm

input necessary in resource

Figure 15: Definition of a dimension the subvector of figure 13. The second keyword contains the unit.

After defining subvector 1 with 3 dimensions one gets the DS definition of figure 16.

\* iu search kw0 I logout[10001] up own us home

\* i7=1006', O | 2015-01-22 Cupboard | Schrank

\* i4= 0, o 2014-01-09 ikea-ivar || 3 Elem/Boeden/Schrank, Kiefer | 175.40, 258, 30, 124,

## Definition of DS 1006 (Cupboard) [owner](#) [Append subv](#) [Del last subv](#) [download-def](#)

< << < > >> >| 0..1

0	Finances
0	Price   Euro
1	Size
0	Width   cm
1	Depth   cm
2	Height   cm

Keyword:

Cupboard

A

Link:

<http://en.wikipedia.org/wiki/cupboard>

Schrank

A

X

<http://de.wikipedia.org/wiki/schrank>

Comment:

This is:  draft  ok  deprecated

Nested metric:  Manhattan  Euclidean  Maximum

DVs can be added into this DS only by you as DS-owner ([change](#))

**Figure 16: Complete definition of the DS "Cupboards/Schraenke".**

It is possible to define minimum and maximum of every dimension. Also the representation can be adapted to the user's needs:

list:

Appropriate if the dimension is simply a small unordered list of items, e.g. a selection between two possibilities like "yes" and "no". In case of ordered dimensions (which are more efficient) you should (instead of "list") use "integer" or some "floating point" or "date" format, because these allow greater or smaller comparisons in case of a search.

tux:

(short alphanumeric text) appropriate if the dimension is a large or increasing unordered list of names; tux can contain up to 8 lowercase letters a..z or digits 0..9. It is introduced because it can be easily remembered. Similarity search is defined so that all DVs are found whose initial letters are at this dimension identical to the searched tux. So the initial letters should be most significant.

integer:

useful e.g. for a count of items

floating point:

e.g. for measurement results

date:



for representation dates (highest significant numbers first)

If you think that everyone who provides a DV would also provide a value for this dimension, you can select "input necessary in DV".

You can provide individual relative weights. The higher the weight of a dimension (or subvector), the more it contributes to the distance, and the more important it is for ordering the search result.

The up to now supported distance functions:

GPS metric:

If not both subvectors contain elevation:

Let be  $\theta_1, \varphi_1$  and  $\theta_2, \varphi_2$  = latitude and longitude of 2 subvectors

and  $\ell$  = Diameter of earth,  $f$  the *atan2* function and  $r_{GPS}$  the distance

$a = \sin((\theta_2 - \theta_1)/2)$  ,  $b = \sin((\varphi_2 - \varphi_1)/2)$  ,  $c = a^2 + b^2 \cos(\theta_1) \cos(\theta_2)$  , then

$$r_{GPS} = \ell \cdot f(\sqrt{c}, \sqrt{1-c})$$

If both subvectors contain elevation:

Let be  $\theta_1, \varphi_1, h_1$  and  $\theta_2, \varphi_2, h_2$  = latitude and longitude and elevation of 2 subvectors

and  $\ell$  = Diameter of earth,  $f$  the *atan2* function and  $r_{GPS}$  the distance

$a = \sin((\theta_2 - \theta_1)/2)$  ,  $b = \sin((\varphi_2 - \varphi_1)/2)$  ,  $c = a^2 + b^2 \cos(\theta_1) \cos(\theta_2)$  , then

$$r_{GPS} = \sqrt{(h_1 - h_2)^2 + (\ell \cdot f(\sqrt{c}, \sqrt{1-c}))^2}$$

Manhattan and Euclidean metric:

Let be  $x_0 \dots x_n$  and  $y_0 \dots y_n$  coordinates of 2 subvectors,  $v_i$  the weights of the dimensions (figure 15) and  $w_i$  the weights given during search (figure 7, values below "w"), then

the Manhattan distance is

$$r = \sum_{i=0}^n |v_i w_i (y_i - x_i)|$$

and the Euclidean distance is

$$r = \sum_{i=0}^n \sqrt{(v_i w_i (y_i - x_i))^2}$$

and the Maximum distance is

$$r = \max_{i=0}^n |v_i w_i (y_i - x_i)|$$

The distance  $d$  of 2 vectors (DVs) is calculated from the distances  $r_k$  of the subvectors (0..m) using the selected "connected" metric (selection after "conn:" in figure 13 ) and weights  $u_k$  of the subvectors (figure 14) and optional further subvector weights  $q_k$  given directly before search (figure 7, field left of "size")

In case of connected Manhattan metric:

$$d = \sum_{k=0}^m |q_k u_k r_k|$$

In case of connected Euclidean metric:

$$d = \sqrt{\sum_{k=0}^m (q_k u_k r_k)^2}$$

In case of connected Maximum metric:

$$d = \max_{k=0}^m |q_k u_k r_k|$$

The distance  $d$  is listed and used for sorting the search result. The smaller the distance, the higher is the rank in the search result.

### **3.9 Draft state of a DS**

Newly defined DSs are automatically in draft state (Figure 13). If you are not satisfied with your definition, you can delete your DS, if the DS is in draft state. We recommend to keep draft state not longer than a week, and to delete the DS within this draft-time when you don't find a satisfying definition.

DSs in draft state are marked by an apostrophe ( ' visible e.g. in the second line of figure 13).

When you find a satisfying definition and think it should be used by all, set the state of your DS to ok (figure 12). After this deletion of the DS is only possible, if it is empty.

Write access to definitions of dimensions is allowed also in a non-draft DS for the owner for clarifications or comments. Definitions should (of course) not be changed in their meaning, if the DS contains already DVs. The original definition at the time of a DV's creation remains accessible (option "vieworg").

### **3.10 Own DVs (temporarily possible for test purposes)**

After definition of a DS it can be filled with DVs. Owning is temporarily possible for test purposes. After login (and selection of the DV) click on "own" and then on "define own DV for DS ...". This opens the DV definition window (figure 17). It is recommendable to provide values for as much as possible dimensions to maximize the chance that the DV can be found in numeric search.

\* iu search kw0 I logout[10001] up own us home

\* i7=1006', o | 2015-01-22 Cupboard | Schrank

\* i4= 0, o 2014-01-09 ikea-ivar || 3 Elem/Boeden/Schrank, Kiefer | 175.40, 258, 30, 124,

## DV 0 (ikea-ivar) of DS 1006 (Cupboard)

owner

download DV

|< << < > >> >| 0..1

0	Finances
0	175.40 Price   Euro
1	Size
0	258 Width   cm
1	30 Depth   cm
2	124 Height   cm

o:  w:

Keyword:

Link:

ikea-ivar

A

http://www.ikea.com/de/de/catalog/...

Comment:

3 Elem/Boeden/Schrank, Kiefer

Submit

*Figure 17: Example of a DV of the DS defined in figure 16. The selection of "o" means that the resource is offered (for sale). Selection of "s" would mean that such a resource is searched (to buy).*

Figure 17 shows a typical example. If your DV describes something for sale ("o"="offer" checked), don't forget to delete it after sale. Similarly, if your DV describes something which you search ("s"="search" checked), don't forget to delete it after you have found what you search.

### 3.11 Uploading own DVs

Own DVs can be also uploaded using interchangeable XML format (which can be studied using the "download DV" option, figure 17). It can be interesting for manufacturer of devices which provide nontrivial numeric measurement results to define a DS with a subvector which contains these measurement results, and to offer export of their data in this XML format (as additional option).

Every company which fabricates devices for making measurements is invited to define custom-designed DSs whose dimensions represent (decision relevant) numbers which are derived from the measurement results - as long as analogous DSs are not already defined. If analogous DSs are already defined, it should be tried to adapt to this standard. If there are ideas about new interesting dimensions for these DSs, we recommend to contact the owner of the existing DS, who should consider ideas about interesting expansions of his DS by new dimensions.

Also programmers of all kinds of software which produces reproducibly sequences of decision relevant numbers (e.g. feature extraction software) can determine DSs and XML export options according to their needs.

### **3.12 Web interface depending on user interest**

Users can provide numbers and other data not only manually. Software (of DS owners) can provide DVs for own DSs and search data via web interface. Due to the flexibility of software this offers very flexible solutions, e.g. determination of coordinates and other values via pointing device or on touch screen, or direct input from measurement devices.

Let us know your suggestions concerning this. We collect proposals to find the best solution. The possibility to upload and download DVs via exchangeable XML format is an example.

But this should be only for demonstration purposes, we still hope for a web standard for worldwide valid DS definitions and DVs, so that these need not to be stored in the database of this DS search engine (with possible XML exchange), but directly on the web. Then all web search engines have the possibility to realize DS search.

## **4 Concluding remarks**

The inducement for this came from medicine:

1. Medical treatment has in the average strongly increased in quality, but it still depends very much on the experience of the chosen doctor (i.e. coincidence plays a relevant role).
2. Therefore a database would be welcome, which allows unification of the experiences.
3. The storage of findings in metric Domain Spaces (DSs) allows access to similar findings and so quick access to (worldwide) experiences with such findings. To get a helpful result, most important decision relevant parameters must be mapped reproducibly onto these DSs. This is recognizable, if enough "similar" cases (i.e. cases within a given distance or min/max constellation) can be found, and if the results (treatment results of a certain treatment decision) show a standard deviation which is small enough to get significant information. Then the results can be valuable for finding the best available treatment decisions.

Future medicine can scan human beings and collect systematically multidimensional genotypic and phenotypic data already in an early stage of life and compare these (after collection of enough data) with those of former patients. So it will be possible to derive from this already early in life important codes of behavior for the scanned person to efficiently prevent later diseases, for which the scanned person is predisposed.

So the medical history of every person is, if stored in analyzable and searchable way, very valuable for all future generations. Because now there are technical possibilities to use it, it seems not responsible to delete it. There are substantial advantages, if most important parts of the medical data and history of a patient are made available to the general public. The patient can decide together with his doctor, which parts are valuable and can be published (e.g. every 10 years one or more pages with most

important experiences and numeric data). These important experiences can be shared if there are reproducible (and enough precise) diagnostic and treatment descriptions. This is also a stimulus for precise diagnostics. The name of the patient (and also the name of the doctor) can be replaced by a number.

Suppose such a database would already exist. Then it would be possible (for your doctor) to search using your diagnostic data worldwide for similar cases and for best treatment decisions.

It is clear, that such systematic worldwide exchange of medical experience can be very advantageous (and also save much money). Whether we can or cannot use these advantages, depends on our decision.