

How to make quantitative data on the web searchable and interoperable part of the common vocabulary

Wolfgang Orthuber¹

Abstract: The elements of the common vocabulary (e.g. words) are quickly known by all participants of a conversation. It was possible to extend on the web the vocabulary effectively by HTTP URLs, because their content is quickly viewable. A further extension is possible, which is powerful but not used up to now: Elements (vectors) of metric spaces, identified by a HTTP URL *plus* a sequence of values, where the URL locates the definition of the metric space. All elements and with this all well defined quantitative data can become searchable and quickly viewable and so part of the common vocabulary. In this paper we introduce the concept and propose worldwide multidimensional "Domain Spaces", which can be defined online by all users and which can be efficiently used as container of a huge extended vocabulary, especially user defined searchable quantitative data of all kinds.

Keywords: Metric Space, Domain Space, DS, Domain Vector, DV, Searchable quantitative data, Vocabulary, Standards, Search process, Selection process, Distributed data structures, Feature evaluation and selection, Feature Vector, Metric Search, High Resolution Search, Numeric Search, Interoperable quantitative data

1 Introduction and state of the art

The development of the here shown approach for viewable, searchable and interoperable quantitative data on the web has been started already in 2006 and meanwhile is demonstrated online [Or6]. Though the concept is designed for efficient handling of quantitative data, it is not restricted to these. [Or5] contains a detailed description and this paper introduces the concept. Easy definitions of multidimensional searchable data by users and adapted search interfaces are considered as important. Identification makes quantitative data interoperable, searchable and efficiently evaluable. With this they become also quickly viewable and so a substantial extension of the common searchable vocabulary. We hope this also leads to a constructive discussion about efficient structured and identified multidimensional data on the web.

1.1 Problem

The term "common vocabulary" is not uniquely defined. Therefore we define it initially in a reasonable way. Most important is that every element of the common vocabulary is

¹ University Medical Center Schleswig-Holstein, Department of Orthodontics, Arnold Heller Str. 3 House 26, 24105 Kiel Germany, orthuber@kfo-zmk.uni-kiel.de

quickly known by conversing people. We know that clickable information was very successful and became even defining characteristic of the web, because it can be opened (usually within 1 second) and so viewed quickly. Therefore by "common vocabulary" we mean all *quickly recognizable information* on the web, especially well known words, symbols and clickable information. At this information is "clickable", if in case of appropriate hardware and software after clicking on a text or image a window with more detailed descriptive information opens within 1 second.

Well defined resources often have well defined quantitative properties or features. These features fully depend on the resource, and usually describe the resource more precisely. For example we know that a "car" has important quantitative features which describe it precisely. Also conventional words of language can have specific quantitative features, e.g. "color" can have three percentual parts of (by wavelength well defined) "red", "green" and "blue". Up to now there is no systematic concept which allows to combine text and other clickable resources on the web with individual multidimensional quantitative data in a way that after click the quantitative data are shown by definition or even graphically, so that they are quickly recognizable. The kind of data should be also recognizable by software (machine readable), so that they become searchable.

1.2 Current approaches, shortcomings

First we look at the current state of the art. There are approaches for representation of data on the web [McGui][W3RDF][HL7F] including quantitative data, but the found approaches do not allow slender and efficient representation of multidimensional data and no definition of world wide multidimensional (metric) spaces. This can be shown best using examples from latest literature, e.g. in medicine where quantitative data are important and decision relevant. The FHIR specification [HL7F] is a next generation standards framework created by HL7 [HL7]. At this interoperable data are coded in resources. We look at an exemplary resource example:

```
<Observation xmlns="http://hl7.org/fhir">
  <text>
    <status value="generated"/>
    <div xmlns="http://www.w3.org/1999/xhtml">
      Jan 30 2014: Body Weight = 185 lbs</div>
  </text>
  <name>
    <coding>
      <system value="http://loinc.org"/>
      <code value="3141-9"/>
      <display value="Weight Measured"/>
    </coding>
  </name>
  <valueQuantity>
    <value value="185"/>
    <units value="lbs"/>
    <system value="http://unitsofmeasure.org"/>
    <code value="[lb_av]"/>
  </valueQuantity>
```

```
</Observation>
```

Fig. 1: Excerpt of <http://www.hl7.org/fhir/observation-examples.html> [HL7FEX] which codes the observation of body weight.

"<http://hl7.org/fhir>" in Fig. 1 specifies the namespace. According to [W3N] "It is not a goal that it be directly usable for retrieval of a schema (if any exists)." The URL "<http://hl7.org/fhir>" contains a human readable description of the general FHIR concept, *somewhere* in it the resource "observation" is defined. The URL "<http://loinc.org>" combined with the code "3141-9" yields the URL <http://loinc.org/3141-9> where the quantity "Body weight Measured" is defined in human readable form. The unit of measurement is not determined there, so it is specified subsequently together with the value "185" as "lbs" with <http://unitsofmeasure.org> and "[lb_av]". Fact is: The more defining information is necessary in the data, the less standardized are the data. So (varying) defining information together with data is unnecessary and even disadvantageous.

1.3 Proposed approach

Fig. 1 contains defining information (e.g. about the unit) but no URL to a standardized definition which is machine readable in uniform format. In current interoperability approaches we have not found usage of such *online definitions*. This means, that together with the (*variable*) data a URL is given which points to a complete (*constant and reliable*) definition in standardized machine readable form. The data are given as sequence of values (variables). They are following after the URL of their definition, there is no further defining or redundant information. The online definition must be reliable and complete, it can contain also extended information in multiple languages. We can use for the online definition a standardized form, e.g. HTML [MF] or in XML. The subsequently described DS definitions (see 3.1) are the online definitions which we propose. Fig. 5 shows an abbreviated example of a definition and Fig. 2 exemplary data. It contains the URL of such a online (DS) definition plus 2 numbers and represents the values of Fig. 1 in machine readable form.

```
<v http://numericsearch.com/bw.xml; 2014-01-30; 83.914>clickable text</v>
```

Fig. 2: Values of Fig. 1 in exemplary representation using the URL of an online definition which contains the definition of a metric space (DS, see 2.1). The complete line can be seen as element (DV, see 2.2) of this space. The online definition (DS definition) can represent very much information in standardized way, e.g. default clickable texts in different languages.

Fig. 2 shows on a single line a possible example of a short new standard (DV, see 2.2) which could represent the date and body weight of Fig. 1. Here 83.914 represents the weight (185 lbs) in kg (we recommend usage of SI units in basal dimension definitions). The compactness of Fig. 2 in comparison to Fig. 1 makes it practicable to place multidimensional quantitative data into HTML pages, and because the URL points to an online definition which completes the information of the quantitative data, after click the complete information can be displayed in human readable form. The approach can be

generalized for all multidimensional data. We will deepen details subsequently.

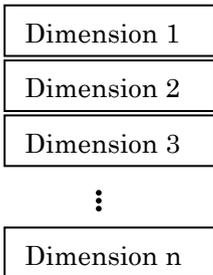
2 Data structures

Fig. 2 shows a HTTP URL plus a sequence of values. To make it searchable we can understand every value as a dimension of a metric space. Every pair of elements of a metric space has a well defined nonnegative distance which is the smaller, the greater the similarity is. Therefore in our approach multidimensional quantitative data (Fig. 2) are defined as elements (vectors) of one- or multidimensional metric spaces. This allows similarity search by minimizing the distance. To cover all interesting topics, all users must get the practicable possibility to compose own online definitions of metric spaces. Subsequently we call these metric spaces "Domain Spaces" (DSs) and their elements "Domain Vectors" (DVs). We selected these names because DSs can be defined by web users according to their domain of interest, and because of their applicability in Domain Ontologies (see [Haas]).

2.1 Domain Space (DS)

Fig. 3 shows the structure of a DS definition. Main content is an ordered sequence of dimension definitions. The DS definition and every dimension definition can be addressed using a HTTP URL.

Domain Space (DS):



Dimension of a DS:

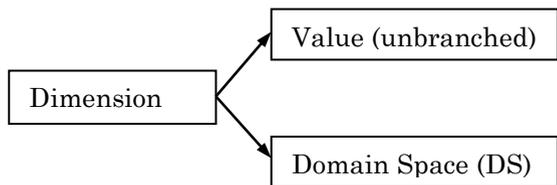


Fig. 3: Structure of a DS and DS dimension. The DS and every of its dimensions have a unique name (HTTP URL). Every dimension of a DS can represent an unbranched value or again a DS.

Fig. 3 shows that DS dimensions can represent a (unbranched) value or again a DS. The value is numeric by default. As special feature a value can also represent text or binary data using e.g. the discrete metric as distance function. If the dimension represents a DS, it is called branched. Because a dimension (of a DS) again can represent a DS, DS definitions can be nested and reused within new DS definitions.

2.2 Domain Vector (DV)

For identification of DVs (quantitative data) on the web we need a unique identifier (URI). It is efficient to use for this a Uniform Resource Locator (URL) [W3U] of the definition of the containing DS. This URL identifies the data *and* simultaneously locates their definition. Therefore the DV content starts with the HTTP URL of the DS definition. After this identifier the list of values follows usually in the same order like in the DS definition. So we get the above mentioned compact structure *HTTP URL plus a sequence of values* of a DV (data) on the web, as shown in Fig. 2. Additional identification of values by (abbreviated) HTTP URLs of dimension definitions is possible. Identification is necessary if the order is changed, e.g. if dimensions are bypassed.

3 Implementation

We made an online implementation which demonstrates on a local database important key steps: Guided definition of DSs and DVs, interoperable data formats which unambiguously depend on the definition, text search within parts of definitions, e.g. for selection of DSs for numeric search. Most interesting is multidimensional similarity search with and without additional conditions, where the search window is automatically adapted to the chosen DS (e.g. Fig. 7 and Fig. 9). In this paper we will show important steps and results of the procedure.

3.1 Definition of a DS

It is necessary that users can define own DSs. For demonstration purposes we show in Fig. 4 the definition of the unbranched simple space "BodyWeight". Details of every dimension can be defined online.

Definition of DS 1029 (BodyWeight) [owner](#)

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

0	BodyWeight
0	Date yyyy-mm-dd
1	Weight kg

Keyword: Link:

Comment:

This is: draft ok deprecated

Nested metric: Manhattan Euclidean Maximum

Fig. 4: Definition of a simple unbranched DS "BodyWeight".

Fig. 5 shows a possible XML representation of this DS definition [OrExDS].

```

<DS>
  <kw>BodyWeight</kw>
  <dim>
    <format>yyyy-mm-dd</format>
    <kw>Date</kw>
    <kw>yyyy-mm-dd</kw>
  </dim>
  <dim>
    <format>float</format>
    <kw>Weight</kw>
    <kw>kg</kw>
  </dim>
</DS>

```

Fig. 5: Possible abbreviated short XML representation of the definition of the DS "BodyWeight". This is only an example for explanation purposes. A later standard will deviate from this.

If "http://numericsearch.com/bw.xml" is the URL of this DS definition (Fig. 5), then Fig. 2 could represent a possible DV of this DS. It could be displayed on the screen as "clickable text" similarly like a clickable hyperlink. After click an adapted browser can show a detailed representation of the quantitative data, also our online implementation demonstrates this (Fig. 6) [OrExDV].

0	BodyWeight
0	2014-01-30 Date yyyy-mm-dd
1	83.914 Weight kg

Fig. 6: Exemplary representation of the DV (quantitative data) of Fig. 2 [OrExDV] using the DS definition (Fig. 4).

Fig. 7 shows an exemplary search (search window online [OrExDV]) in a small DS "BodyWeight" with 4 DVs.

NumericSearch in DS 1029 (BodyWeight)

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

sim | min max | g

0				BodyWeight
0				<input type="checkbox"/> Date yyyy-mm-dd
1	80			<input type="checkbox"/> Weight kg

Fig. 7: Possible search within DS "BodyWeight" [OrExS1], Fig. 8 shows the search result.

d	a
0 15	o clickable-text-1 2014-02-05, 80, 3.914
24	o clickable-text 2014-01-30, 83.914,
5 21	o clickable-text-2 2014-03-10, 75,
10 15	o clickable-text-3 2014-04-20, 70,

Fig. 8: Search result of Fig. 7. Column "a" shows the access count to the DV, Column "d" the distance (Manhattan metric with weight 1), and the last column the last dimension of the DV (weight).

Fig. 8 demonstrates the principle of similarity search in this simple example. As distance function default Manhattan Metric (in case of one dimension the absolute difference) is chosen. Let w represent the last dimension (weight) in a DV, d the distance. Because weight 80kg was searched, $d=|80-w|$. The search result is sorted that the rank is the higher, the smaller d is.

4 Evaluation

A DV represents quantitative data by a HTTP URL plus a sequence of values (Fig. 2). It is interoperable because data of the same kind have the same format and can be recognized by a unique URL. It is clickable like today a HTTP URL is clickable, if the URL points to a complete standardized definition of the values (a DS definition, e.g. Fig. 5) and the browser is programmed to combine it with the values. Also our implementation demonstrates clickable text and after click combination of values with a DS definition (Fig. 6).

Less trivial is search of DVs, especially its practicability. Because on the web all DVs of the same DS have the same identifier (the URL of the DS definition), a crawler can collect these into the same database. So DVs are searchable. But it is not clear to what extent search of DVs is possible in larger DSs. Therefore we generated a high dimensional nested DS and filled it with 100001 DVs. It combines according to 2.1 unbranched DSs. Two of them (a 4-dimensional DS with Euclidean Metric and a 140-dimensional DS with Manhattan Metric) were filled with pseudo random numbers which are equally distributed between 0 and 10. To demonstrate the effect of similarity search we searched in a 2 dimensional subspace (Fig. 9).

NumericSearch in DS 1023 (test-150dim)

		0..2..2		search	spar
	sim	min	max	g	
0				subv0	
0				<input type="checkbox"/>	dim0
1				<input type="checkbox"/>	dim1
1				subv1	
0	2			<input checked="" type="checkbox"/>	dim0
1	7			<input checked="" type="checkbox"/>	dim1

Fig. 9: Excerpt of the search window, Fig. 10 shows the search result.

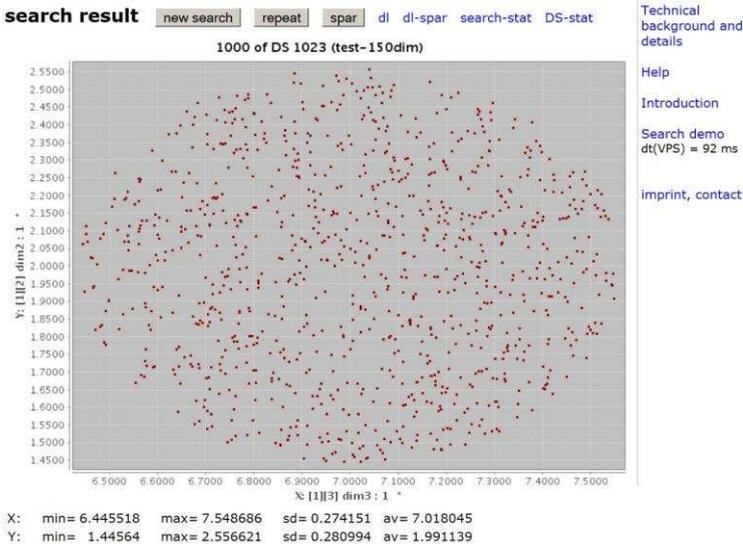


Fig. 10: Search result of Fig. 9, shown are the 1000 found DVs of 100001 DVs.

Fig. 10 shows the graphical result after search of Fig. 9. The content of the searched dimensions of the 1000 nearest of 100001 DVs is displayed. Due to the Euclidean Metric an elliptic shape around the searched coordinates results, because it represents in the given scale geometrically the set of nearest points around the point with the searched coordinates.

Now we test the performance of multidimensional similarity search in a more complex example with combined Euclidean and Manhattan metric. Let x_m denote a coordinate of a DV within the Euclidean Space and x_{s_m} the corresponding searched coordinate and y_n a coordinate of a DV within the Manhattan Space and y_{s_n} the corresponding searched coordinate, then the distance d is:

$$d = \sqrt{\sum_{m=1}^4 (x_m - x_{s_m})^2 + \sum_{n=1}^{n_{\max}} |y_n - y_{s_n}|} \cdot (1)$$

Using this distance function we performed similarity search with $n_{\max} \in \{0,10,20,30,40,50,60,70,80,90,100,110,120,130,140\}$. We want to emphasize that due to the curse of dimensionality [Ag] high dimensional similarity search ($n_{\max} > 10$) tends to become inefficient in case of independent dimensions, else (in case of dependent dimensions) dimension reduction techniques are recommendable (e.g. [Indyk]). If the dependencies of dimensions are hidden, initially this can be difficult. High dimensional search was tested here to examine the performance of the system also in case of difficult situations.

According to (1) the total count of searched dimensions is $n_{\max} + 4$. Fig. 11 shows the average search times in case of 4.144 searched dimensions. The implementation was programmed using java jdk-7u4-linux-x64.rpm, apache-tomcat-7.0.27.zip, synchronized

index [Or5]. To save costs this implementation was done on a virtual server. It uses 14GB RAM and 4 cores of an Intel Core i7 CPU with 3.2 GHz, 1000 GB HDD with Linux Cent OS 6.3.

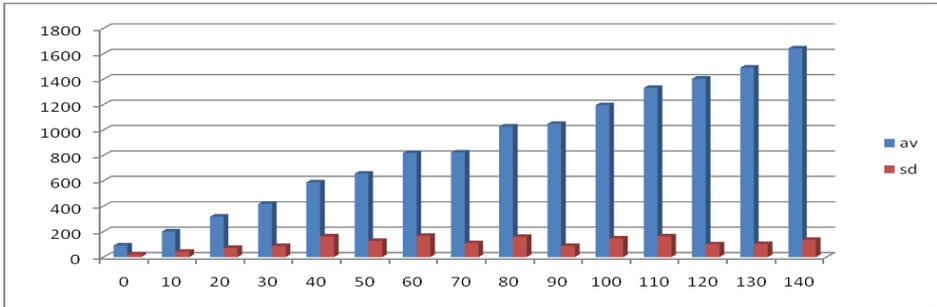


Fig. 11: Horizontal axis: nmax (=Count of searched dimensions in Manhattan Space)
Vertical axis: av (=average search time of 20 similarity searches within 100001 DVs in milliseconds) and sd (=standard deviation of search time)

Fig. 11 shows that it is possible to search through 100001 DVs on the average within 100ms in case of 4 dimensions, within 200ms in case of 14 dimensions and within 1700ms in case of 144 dimensions. Independently of this search performance could be enhanced by parallel processing, which is possible after splitting the space and reunion of the search results.

5 Discussion

According to [W3LT] at the June 2011 Semtech conference it was announced that a query of 1,009,690,381,946 triples was done in 338 hours for an average rate of 829556 triples per second. The driving force has been Amdocs and their AIDA platform. So averagely per 100ms 82956 triples have been processed. In another source [Erl] was stated "Loading speed for data in the Turtle syntax is up to 38000 triples per second." (i.e. 3800 triples within 100ms) The here presented implementation searched in case of 4 dimensions on the average within 100ms through 100001 DVs, using a cheap virtual server. So it becomes clear that DV search is practicable and very efficient compared to existing approaches.

The in Fig. 3 shown structure of DSs definitions is similar to that of an ordered directory tree: A DS combines dimensions which represent values or again DSs (like a directory contains files or again directories). At this in a DS definition the order of dimension definitions is important because it determines the order in which values can be given in a DV without necessary dimension identifier. Moreover in case of DSs circular definitions are possible, i.e. a DS can contain somewhere in its own tree a dimension which represents (points back to) an own definition. Therefore software has to stop after a predefined count of circular extensions.

"Why does this extend the searchable vocabulary?" is a natural question due to the topic and explicitly answered: The extension of the common searchable vocabulary by quantitative data means, that these are quickly viewable (quickly known) and searchable by readers. Fig. 6 shows an example for this. With this not only the clickable text (e.g. "BodyWeight") but also underlying quantitative data become part of the quickly viewable and searchable common vocabulary. The definitions of the numbers and names can be translated also into other languages, so that after clicking on a DV the definition can be shown in the language of the reader (if known by the system). Words within conventional text could be used as clickable text of a DV (Fig. 2) and get additional searchable quantitative features, e.g. "drive" could get the additional feature "speed in km/h". If wished, such "quantified text" could contribute to more precise conversation without being more difficult to read than conventional text. We hope the scope of such an extension of conventional language (for technical, medical, legislative texts etc.) will be recognized. A detailed discussion [Or5] would exceed the scope of this paper, here some notes:

- Syntax for definition of DSs and DVs must be standardized.
- Definition of a DS shows data which are interesting in a certain domain. This increases motivation for writers to provide these data, to make the web more informative.
- The Definition of a DS (with quantitative dimensions) defines all elements with their connection and with this much more than the definition of a word. Therefore range and resolution of DV based description and search is much higher than word based description and search.
- Though the definition of a DS can be very elaborate, patents on DS definitions are counterproductive. DS definitions and DVs can be seen as part of language and of course patents on parts of language should not be possible.
- It is important that DS definitions are stable and reliable. The owner of a certain domain can provide some warranty, moreover DS definitions can be stored by search engines to detect possible incompatible changes.
- Redundant definitions should be avoided or linked together with equivalent definitions using "sameAs" directives [W3Ont]. For this focused text search within DS definitions is necessary.
- It possible to derive from every DS automatically an "evaluation DS" in an open database to give users the possibility to evaluate every DV of the DS with an own "evaluation DV" of the "evaluation DS".
- DSs with textual dimensions allow construction of sentences with given structure, also triples of the semantic web [BeSe].
- DSs whose dimensions describe preconditions (1), a following decision (2) and the following results (3) can be used for decision support e.g. in medicine.

6 Conclusion

Introduction of online defined standardized metric spaces (DSs resp. "Domain Spaces") is an efficient means for extending the common vocabulary by user defined quantitative data (DV's) and for making quantitative data interoperable and searchable. Due to the far reaching technical potential it is recommendable to deepen this topic by science and economy.

Literature

- [Ag] Aggarwal, C. Hinneburg, and A. Keim, D.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. ICDT 2001, Vol. 1973, 420-434. DOI: http://dx.doi.org/10.1007/3-540-44503-x_27, 2001.
- [An] Andoni, A. and Indyk, P.: Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. Communications of the ACM, vol. 51, no. 1, 2008, p. 117-122. <http://mags.acm.org/communications/200801/#pg119>, 2006.
- [BeSe] Berners-Lee, T., .Hendler, J., and Lassila, O. 2001. "The semantic web." Scientific american 284.5, 28-37.<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>
- [Erl] Erling, O., & Mikhailov, I. RDF Support in the Virtuoso DBMS. In Networked Knowledge-Networked Media (pp. 7-24). Springer Berlin Heidelberg. 2009.
- [Gae] Gärdenfors, P. Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge, MA, 2000.
- [Haas] Haas, P.: Medizinische Informationssysteme und Elektronische Krankenakten. Springer, Berlin, Heidelberg, 2005.
- [HL7] HL7 Health Level Seven® INTERNATIONAL, <http://www.hl7.org/>, viewed June 2015
- [HL7F] HL7 INTERNATIONAL: FHIR, <http://www.hl7.org/fhir/>, viewed June 2015
- [HL7FEX] HL7 INTERNATIONAL: Resource Observation - Examples, Simple Weight Example, <http://www.hl7.org/fhir/observation-examples.html>, viewed June 2015
- [Indyk] Indyk P.: Dimensionality Reduction Techniques for Proximity Problems, Stanford University. (650-723-4532). <http://people.csail.mit.edu/indyk/proxy99.ps>, 1999.
- [McGui] McGuinness, D. L., & Van Harmelen, F. OWL web ontology language overview. W3C recommendation, 10(10), 2004.
- [MF] Microformats: Get started, <http://microformats.org/wiki/get-started>, viewed June 2015
- [Na] Nakao M. A., Axelrod S.: Numbers are better than words. Verbal specifications of frequency have no place in medicine. Am J Med 1983, 74, 1061–1065, 1983.
- [Or1] Orthuber, W., Fiedler, G., Kattan, M., Sommer, T., Fischer-Brandies, H.: Design of a global medical database which is searchable by human diagnostic patterns. The Open Medical Informatics Journal 2, 21-32, 2008.

- [Or2] Orthuber, W., & Dietze, S.: Towards Standardized Vectorial Resource Descriptors on the Web. In GI Jahrestagung (2) (pp. 453-458), 2010.
- [Or3] Orthuber, W., Papavramidis E.: Standardized Vectorial Representation of Medical Data in Patient Records, Medical and Care Compunctics 6, 153-166, 2010.
- [Or4] Orthuber, W.: General approach to similarity search of resources with numeric features on the web. Open Data on the Web. <http://www.w3.org/2013/04/odw/papers#all>, 2013.
- [Or5] Orthuber, W.: Worldwide Domain Spaces make quantitative data searchable and prepare these for interoperable exchange, <http://numericsearch.com/wdomainspaces.pdf>, viewed 15.06.2015
- [Or6] Orthuber, W.: NumericSearch - online implementation of vectorial description and search, <http://NumericSearch.com>, viewed 24.04.2015
- [OrExDS] Orthuber, W.: Temporary DS example, <http://numericsearch.com/wcomedit.jsp?i9=77>, viewed 14.06.2015
- [OrExDV] Orthuber, W.: Temporary DV example, <http://numericsearch.com/wcomedit.jsp?i9=72&i7=1029&i4=0>, viewed 14.06.2015
- [OrExS1] Orthuber, W.: Temporary search example 1, <http://numericsearch.com/w7s.jsp?i7=1029>, viewed 14.06.2015
- [OrExS2] Orthuber, W.: Temporary search example 2, <http://numericsearch.com/w7s.jsp?i7=1023>, viewed 14.06.2015
- [W3M] W3C: Metric Spaces on the Web, Community Group, in <https://www.w3.org/community/numericweb/>, viewed May 2015.
- [W3N] W3C: Namespaces in XML 1.0 (Third Edition), W3C Recommendation 8 December 2009, in <http://www.w3.org/TR/REC-xml-names/>, 2009
- [W3Ont] W3C: OWL Web Ontology Language, W3C Recommendation, owl:sameAs, in <http://www.w3.org/TR/owl-ref/#sameAs-def>, 10 February 2004
- [W3LT] W3C: LargeTripleStores, AllegroGraph, <http://www.w3.org/wiki/LargeTripleStores>, viewed June 2015
- [W3U] World Wide Web Consortium: URIs, URLs, and URNs: Clarifications and Recommendations 1.0, <http://www.w3.org/TR/uri-clarification/>, 2001
- [W3RDF] World Wide Web Consortium. RDF 1.1 Concepts and Abstract Syntax. <http://www.w3.org/TR/rdf11-concepts/>, 2014
- [Za1] Zander S. and Schandl B. Context-driven rdf data replication on mobile devices. Semantic Web Journal Special Issue on Real-time and Ubiquitous Social Semantics, 1(1), 2011
- [Zezula] Zezula, P., Amato, G., Dohnal, V., Batko, M. Similarity Search. The Metric Space Approach. Series: Advances in Database Systems, Vol. 32., Springer, Berlin, Heidelberg. 2005.