



(REVIEW ARTICLE)



## Internet since decades - but no globally searchable data

Wolfgang Orthuber \*

*Department of Orthodontics, UKSH, Kiel University, 24105 Kiel, Germany.*

Publication history: Received on 07 August 2020; revised on 19 September 2020; accepted on 21 September 2020

Article DOI: <https://doi.org/10.30574/gjeta.2020.4.3.0053>

### Abstract

The usability of digital information on the internet strongly depends on its searchability. There are search engines for global language based text search, but language based information representation has only limited value, e.g. concerning reproducibility, comparability, precision. Searchable digital information can be much more. Any piece of digital information is a number sequence. We can define the binary format and value set resp. domain of every number online. Then universal information transport is possible by the online defined "Domain Vector" (DV) data structure: "UL plus number sequence". At this "UL" is an efficient link to the machine readable online definition of the number sequence. The UL also is global identifier of a certain kind of data. The online definition includes additional information, e.g. for similarity comparison of the number sequence. With this a universal numeric search engine can provide precise user defined worldwide search of DVs like in a globalized database. All users can participate. Together we can optimize online definitions and provide defined data as DVs. Initially it is necessary to build a first attractive online presence where users can provide online definitions. We should stay in contact to optimize together the common standard for machine readable online definitions and DVs.

**Keywords:** Information; Online definition; Domain; Domain Vector; DV

### 1. Introduction

Text search has become indispensable for every user of the internet. It is necessary, especially if we search for names and other distinctive text parts. But often we search not a priori text but the digital representation of some original. This includes any information about reality, for example the description of an exact medical finding or of a product with technical data. There are many possibilities for translation of such nontrivial language independent original information into language, even if we use the same language vocabulary. As consequence, the same original information is represented in very variable way digitally as text, and often this representation is not precise. This clarifies the limited value of text representation and of text search.

### 2. Globally searchable information

To overcome the limits of language, we have to recall the basics of information representation. As described in detail [1] [2], "information" always is represented as selection from a "domain" which is a "common set of possibilities" for sender and all receivers of information. Also the bits of digital information are parts of numbers which are selections. They may select e.g. letters from Unicode and so words and phrases from language vocabulary. Frequently this (detour) is preferred because our brain is adapted to it. However, "language vocabulary" is only one possible domain. We can adapt the domain to the application. After building an online presence, *all users* could define adapted domains online [1] [2] [3] [4] [5]. Such adapted domains aim to represent bijectively especially the important features of a certain application and can be extended afterwards. The digital information is transported as "Domain Vector" resp. "DV".

\* Corresponding author: Wolfgang Orthuber  
Department of Orthodontics, UKSH, Kiel University, 24105 Kiel, Germany.  
E-mail address: orthuber@kfo-zmk.uni-kiel.de

DV: UL plus number sequence

At this "UL" is identifier and efficient global pointer [2] to the online definition of the following number sequence. The online definition contains all information about the numbers' binary form and their domain. It can be very detailed, also with optional multimedia content, and it specifies the necessary distance function for similarity comparison and high resolution search of DVs [6]. In contrast to this the defined DVs contain only minimal metadata and are designed for maximal efficiency, because DVs typically are transported for information exchange. Due to the (global) online definition and uniform identification of DVs by UL, all features which have been defined online by the users also are globally searchable [3]. The online definition can be used additionally for other important applications, e.g. for worldwide uniform specification of scientific data and of data for global machine learning (AI), or for definition of programming interfaces for global programming. However, this simple and efficient technical opportunity has not been used since decades!

Even if today it is new to exactly define information online using a "set of possibilities" or "domain", this is not difficult. Sets are defined again and again e.g. in science. Also the value sets of databases are domains. However, we don't want increasing patchwork with local databases. We want globally defined and searchable data. Of course we should define "sets of possibilities" resp. "domains" of data globally - and therefore online - for searchable global information. The enabled user defined data search is very important, e.g. for decision support in medicine and for comparison of scientific data. Such data search is (in comparison to text search) also a new precise language independent alternative for navigation of increasing e-commerce sales which worldwide amounted meanwhile to trillions of US dollars [7]. This clarifies also the commercial importance of precise data search. To prepare the infrastructure for this, initial step is an attractive online presence where users can get "own" groups of ULs [2] for global definition of information which is searchable according to criteria which are regarded as relevant. There may be multiple such online presences with online definitions. However, we should stay in contact to coordinate the common standard for the machine readable online definitions and DVs. Optimal applicability and technical efficiency are common aims.

---

### 3. Conclusion

Globally searchable data can be realized in an efficient way. Any piece of digital information is a number sequence. It can be defined online. The DV data structure "UL plus number sequence" is uniformly identified and defined, because "UL" is an efficient global pointer to the unique machine readable online definition of the number sequence. Therefore the DV data structure is globally searchable according to criteria which can be defined online by the users. Because this principle is language independently and generally applicable at the beginning of the definition of digital information, it is far reaching. The optimization of online definitions and DVs and their application is a very promising future research area.

---

### Compliance with ethical standards

#### *Acknowledgments*

Thanks to those around me for the good working atmosphere. The research is self-financed.

#### *Disclosure of conflict of interest*

The author declares that he has no conflict of interest.

---

### References

- [1] Orthuber W. Information Is Selection—A Review of Basics Shows Substantial Potential for Improvement of Digital Information Representation. *International Journal of Environmental Research and Public Health*. 2020; 17(8): 2975.
- [2] Orthuber W. How to make medical information comparable and searchable. *Digital Medicine*. 2020; 6(1): 1-8.
- [3] Orthuber W. Global predefinition of digital information. *Digital Medicine*. 2018; 4(4): 148-156.
- [4] Orthuber W. Online definition of comparable and searchable medical information. *Digital Medicine*. 2018; 4(2): 77-83.

- [5] Orthuber W, Hasselbring W. Proposal for a new basic information carrier on the Internet: URL plus number sequence. In Proceedings of the 15th International Conference on the World Wide Web, Mannheim, Germany. 2016: 279–284.
- [6] Zezula P, Amato G, Dohnal V, Batko M. Similarity search: the metric space approach (Vol. 32). Springer Science & Business Media. 2006.
- [7] Purnamasari AM, Pah CEA, Yoga MDI, Girsang AS, Isa SM. Business Intelligent in an E-Commerce Industry. In IOP Conference Series: Materials Science and Engineering. 2019; 598(1): 012085.