Knowledge Processing and Advanced Application Scenarios With the Content Factor Method

Claus-Peter Rückemann
Westfälische Wilhelms-Universität Münster (WWU),
Leibniz Universität Hannover,
North-German Supercomputing Alliance (HLRN), Germany
Email: ruckema@uni-muenster.de

***Abstract*—This paper presents the developments and results on knowledge processing for advanced application scenarios. The processing and discovery is based on the new Content Factor (CONTFACT) methodology used for data description and analysis. The Content Factor method can be applied to arbitrary data and content and it can be adopted for many purposes. Normed factors and variants can also support data analysis and knowledge discovery. This paper presents the algorithm, introduces into the norming of Content Factors, and discusses advanced examples, practical case studies, and implementations based on long-term knowledge resources, which are continuously in development. The Content Factor can be used with huge structured and even unstructured data resources, allows an automation, and can therefore also be used for long-term multi-disciplinary knowledge. The methodology is used for advanced processing and also enables methods like data rhythm analysis and characterisation. It can be integrated with complementary methodology, e.g., classification and allows the application of advanced computing methods. The goal of this research is to create new practical processing algorithms based on the general and flexible Content Factor methodology and develop advanced processing components.**

***Keywords*–Data-centric Knowledge Processing; Content Factor (CONTFACT) method; Data Rhythm Analysis; Universal Decimal Classification; Advanced Computing.**

I. INTRODUCTION

The application of the Content Factor method has created new flexible means for the enhancement of knowledge resources and for knowledge discovery processes. This extended research is based on the results from multi-disciplinary projects enhancement of knowledge resources and discovery by computation of Content Factors. The fundaments of the new Content Factor method were presented at the INFOCOMP 2016 conference in Valencia, Spain [1].

This research presents complex use cases for knowledge processing and advanced application scenarios in context with the computation of Content Factors and discusses the results. Information systems handling unstructured as well as structured information are lacking means for data description and analysis, which is data-centric and can be applied in flexible ways. In the late nineteen nineties, the concept of in-text documentation balancing has been introduced with the knowledge resources in the LX Project. Creating knowledge resources means creating, collecting, documenting, and analysing data and information. This can include digital objects, e.g., factual data, process information, and executable programs, as well as realia objects. Long-term means decades because knowledge is not isolated, neither in space nor time. All the more, knowledge does have a multi-disciplinary context. Data [2] and data specialists [3] are becoming increasingly important. Data repositories are core means [4] for long-term knowledge and are discussed to be a core field of activities [5].

Therefore, after integration knowledge should not disintegrate, instead it should be documented, preserved, and analysed in context. The extent increases with growing collections, which requires advanced processing and computing. Especially the complexity is a driving force, e.g., in depth, in width, and considering that parts of the content and context may be continuously in development. Therefore, the applied methods cannot be limited to certain algorithms and tools. Instead there are complementary sets of methods.

The methodology of computing factors [6] and patterns [7] being representative for a certain part of content was considered significant for knowledge resources and referred material. Fundamentally, a knowledge representation is surrogate. It enables an entity to determine consequences without forcing an action. For the development of these resources a definition-supported, sortable documentation-code balancing was created and implemented.

The Content Factor (CONTFACT) method advances this concept and integrates a definition-supported sortable documentation-code balancing and a universal applicability. The Content Factor method is focussing on documentation and analysis. The Content Factor can contain a digital 'construction plan' or a significant part of digital objects, like sequenced DeoxyriboNucleic Acid (DNA) does for biological objects [8]. Here, a construction plan is what is decided to be a significant sequence of elements, which may, e.g., be sorted or unsorted. Furthermore, high level methods, e.g., "rhythm matching", can be based on methods like the Content Factor.

This paper is organised as follows. Section II summarises the state-of-the-art and motivation, Sections III and IV introduce the Content Factor method and an example for the application principle. Section V shows basic Content Factor examples, explains flags, definition sets, and norming. Sections VI and VII introduce the background and provide the results from 8 application scenarios and implementations. Section VIII discusses aspects of processing and computation. Sections XI and X present and evaluation and main results, summarise the lessons learned, conclusions and future work.

II. State-of-the-art and motivation

Most content and context documentation and knowledge discovery efforts are based on data and knowledge entities. Knowledge is created from a subjective combination of different attainments, which are selected, compared and balanced against each other, which are transformed, interpreted, and used in reasoning, also to infer further knowledge. Therefore, not all the knowledge can be explicitly formalised.

Classification has proven to be a valuable tool for long-term and complex information management, e.g., for environmental information systems [9]. Conceptual knowledge is also a complement for data and content missing conceptual documentation, e.g., for data based on ontologies used with dynamical and autonomous systems [10].

Growing content resources means huge amounts of data, requirements for creating and further developing advanced services, and increasing the quality of data and services. With growing content resources content balancing and valuation is getting more and more important.

Knowledge and content are multi- and inter-disciplinary long-term targets and values [11]. In practice, powerful and secure information technology can support knowledge-based works and values. Computing goes along with methodologies, technological means, and devices applicable for universal automatic manipulation and processing of data and information. Computing is a practical tool and has well defined purposes and goals.

Most measures, e.g., similarity, distance and vector measures, are only secondary means [12], which cannot cope with complex knowledge. Evaluation metrics are very limited, and so are the connections resulting from co-occurences in given texts, e.g., even with Natural Language Processing (NLP), or clustering results in granular text segments [13].

Evaluation can be based on word semantic relatedness, datasets and evaluation measures, e.g., the WordSimilarity 353 dataset (EN-WS353) for English texts [14]. The development of Big Data amounts and complexity up to this point show that processing power is not the sole solution [15]. Advanced long-term knowledge management and analytics are on the rise.

Value of data is an increasingly important issue, especially when long-term knowledge creation is required, e.g., knowledge loss due to departing personnel [16]. Current information models are not able to really quantify the value of information. Due to this fact one of the most important assets [17], the information, is often left out [18]. Today a full understanding of the value of information is lacking. For example, free Open Access contributions can bear much higher information values than contributions from commercial publishers or providers.

For countless application scenarios the entities have to be documented, described, selected, analysed, and interpreted. Standard means like statistics and regular expression search methods are basic tools used for these purposes.

Anyhow, these means are not data-centric, they are volatile methods, delivering non-persistent attributes with minimal descriptive features. The basic methods only count, the result is a number. Numbers can be easily handled but in their solelity such means are quite limited in their descriptiveness and expressiveness.

Therefore, many data and information handling systems create numbers of individual tools, e.g., for creating abstracts, generating keywords, and computing statistics based on the data. Such means and their implementations are either very basic or they are very individual. Open Access data represents value, which must not be underestimated for the development of knowledge resources and Open Access can provide new facilities [19] but it also provides challenges [20].

The pool of tools requires new and additional methods of more universal and data-centric character – for structured and unstructured data.

New methods should not be restricted to certain types of data objects or content and they should be flexibly usable in combination and integration with existing methods and generally applicable to existing knowledge resources and referenced data. New methods should allow an abstraction, e.g., for the choice of definitions as well as for defined items.

III. The Content Factor

The fundamental method of the Basic Content Factor (BCF), κ_{B} – "Kappa-B" –, and the Normed Basic Content Factor (NBCF), $\overline{\kappa}_{\mathrm{B}}$, can be described by simple mathematical notations. For any elements o_i in an object o, holds

$$o_i \in o\,. \tag{1}$$

The organisation of an object is not limited, e.g., a reference can be defined an element. For κ_{B} of an object o, with elements o_i and the count function c, holds

$$\kappa_{\mathrm{B}}(o_i) = c(o_i)\,. \tag{2}$$

For $\overline{\kappa}_{\mathrm{B}}$ of an object o, for all elements n, with the count function c, holds

$$\overline{\kappa}_{\mathrm{B}}(o_i) = \frac{c(o_i)}{\displaystyle\sum_{i=1}^{n} c(o_i)}\,. \tag{3}$$

All normed κ for the elements o_i of an object o sum up to 1 for each object:

$$\sum_{i=1}^{n} \overline{\kappa}_{\mathrm{B}}(o_i) = 1\,. \tag{4}$$

For a mathematical representation counting can be described by a set o and finding a result n, establishing a one to one correspondence of the set with the set of 'numbers' $1, 2, 3, \ldots, n$. It can be shown by mathematical induction that no bijection can exist between $1, 2, 3, \ldots, n$ and $1, 2, 3, \ldots, m$ unless $n = m$. A set can consist of subsets. The method can, e.g., be applied to disjoint subsets, too. It should be noted that counting can also be done using fuzzy sets [21].

IV. Abstract Application example

The methodology can be used with any object, independent if realia objects or digital objects. Nevertheless, for ease of understanding the examples presented here are mostly considering text and data processing. Elements can be any part of the content, e.g., equations, images, text strings, and words.

In the following example, "letters" are used for demonstrating the application. Given is an object with the sample content of 10 elements:

$$\mathtt{A\ T\ A\ H\ C\ T\ O\ A\ R\ Z} \tag{5}$$

For this example it is suggested that A and Z are relevant for documentation and analysis. The relevant elements, AAAZ, in an object of these 10 elements for element A means 3/10 normed so the full notation is

$$\mathtt{AAAZ}/10 \text{ with } \overline{\kappa}_{\mathrm{B}}(\mathtt{A}) = 3/10 \text{ and } \overline{\kappa}_{\mathrm{B}}(\mathtt{Z}) = 1/10\,. \tag{6}$$

In consequence, the summed value for AAAZ/10 is

$$\overline{\kappa}_{\mathrm{B}}(\mathtt{A,Z}) = 4/10\,. \tag{7}$$

AAAZ in an object of 20 elements, for element A means 3/20 normed, which shows that it is relatively less often in this object. 3/22 for element A for this object would mean this object or an instance in a different development stage, e.g., at a different time or in a different element context. The notation

$$\{i_1\}, \{i_2\}, \{i_3\}, \ldots, \{i_n\}/n \tag{8}$$

of available elements holds the respective selection where $\{i_1\}, \{i_2\}, \{i_3\}, \ldots, \{i_n\}$ refers to the definitions of element groups. Elements can have the same labels respectively values. From this example it is easy to see that the method can be applied independent from a content structure.

V. PRACTICAL CONTENT FACTOR EXAMPLES

The following examples (Figures 1, 2, 4, 3, 5) show valid notations of the Normed Basic Content Factor $\overline{\kappa}_{\mathrm{B}}$, which were taken from the LX Foundation Scientific Resources [22]. The LX Project is a long-term multi-disciplinary project to create universal knowledge resources. Application components can be efficiently created to use the resources, e.g., from the Geo Exploration and Information (GEXI) project. Any kind of data can be integrated. Data is collected in original, authentic form, structure, and content but data can also be integrated in modified form. Creation and development are driven by multifold activities, e.g., by workgroups and campaigns. A major goal is to create data that can be used by workgroups for their required purposes without limiting long-term data to applications cases for a specific scenario. The usage includes a targeted documentation and analysis. For the workgroups, the Content Factor has shown to be beneficial with documentation and analysis. There are countless fields to use the method, which certainly depend on the requirements of the workgroups. For the majority of use cases, especially, selecting objects and comparing content have been focus applications. With these knowledge resources multi-disciplinary knowledge is documented over long time intervals. The resources are currently already developed for more than 25 years. A general and portable structure was used for the representation.

```
CONTFACT:20150101:MS:{A}{A}{G}{G}{G}/2900
CONTFACT:20150101:M:{A}:=Archaeology|Archeology
CONTFACT:20150101:M:{G}:=Geophysics
```

Figure 1. NBCF $\overline{\kappa}_{\mathrm{B}}$ for an object, core notation including the normed CONTFACT and definitions, braced style.

The Content Factor can hold the core, the definitions, and additional information. The core is the specification of κ_{B} or $\overline{\kappa}_{\mathrm{B}}$. Definitions are assignments used for the elements of objects, specified for use in the core.

Here, the core entry shows an International Standards Organisation (ISO) date or optional date-time code field, a flag, and the CONTFACT core. The definitions hold a date-time code field, flag, and CONTFACT definitions or definitions sets as shown here. Definition sets are groups of definitions for a certain Content Factor. The following examples show how the definition sets work.

```
CONTFACT:20150101:MS:AAG/89
CONTFACT:20150101:M:A:=Archaeology|Archeology
CONTFACT:20150101:M:G:=Geophysics
```

Figure 2. NBCF $\overline{\kappa}_{\mathrm{B}}$ for an object, core notation including the normed CONTFACT and definitions, non-braced style.

```
CONTFACT:20150101:MU:A{Geophysics}{Geology}/89
CONTFACT:20150101:M:A:=Archaeology|Archeology
CONTFACT:20150101:M:{Geophysics}:=Geophysics|
  Seismology|Volcanology
CONTFACT:20150101:M:{Geology}:=Geology|
  Palaeontology
```

Figure 3. NBCF $\overline{\kappa}_{\mathrm{B}}$ for an object, core notation including the normed CONTFACT and definitions, mixed style.

```
CONTFACT:20150101:MU:{Archaeology}{Geophysics}/120
CONTFACT:20150101:M:Archaeology:=Archaeology|
  Archeology
CONTFACT:20150101:M:Geophysics:=Geophysics
```

Figure 4. NBCF $\overline{\kappa}_{\mathrm{B}}$ for an object, core notation including the normed CONTFACT and definitions, multi-character non-braced style.

```
CONTFACT:20150101:MU:vvvvaSsC/70
CONTFACT:20150101:M:v:=volcano
CONTFACT:20150101:M:a:=archaeology
CONTFACT:20150101:M:S:=Solfatara
CONTFACT:20150101:M:s:=supervolcano
CONTFACT:20150101:M:C:=Flegrei
```

Figure 5. NBCF $\overline{\kappa}_{\mathrm{B}}$ for an object from a natural sciences collection, multi-case non-braced style.

Definitions can, e.g., be valid in braced, non-braced, and mixed style. Left values can have different labels, e.g., uppercase, lowercase, and mixed style can be valid. Figure 6 shows an example using Universal Decimal Classification (UDC) notation definitions.

```
CONTFACT:20150101:MS:{UDC:55}{UDC:55}/210
CONTFACT:20150101:M:{UDC:55}:=Earth Sciences. Geological
  sciences
```

Figure 6. NBCF $\overline{\kappa}_{\mathrm{B}}$ for an object from a natural sciences collection, UDC notation definitions, braced style.

Conceptual knowledge like UDC can be considered in many ways, e.g., via classification and via description.

A. Flags

Content Factors can be associated with certain qualities. Sample flags, which are used with core, definition, and additional entries are given in Table I.

TABLE I. SAMPLE FLAGS USED WITH CONTFACT ENTRIES.

Purpose	*Flag*	*Meaning*
Content Factor quality	U	Unsorted
	S	Sorted
Content Factor source	M	Manual
	A	Automated
	H	Hybrid

The CONTFACT core entries can have various qualities, e.g., unsorted (U) or sorted (S). Unsorted means in the order in which they appear in the respective object. Sorted means in a different sort order, which may also be specified. CONTFACT entries can result from various workflows and procedures, e.g., they can be created on manual base (M) or on automated base (A). If nothing else is specified the flag refers to the way object entries were created. Content Factor quality refers to core entries, source also refers to the definitions and information.

The Content Factor method provides the specified instructions. The required features with an implementation can, e.g., implicitly require large numbers of comparisons, resulting in highly computationally intensive workflows on certain architectures. It is the choice of the user to weighten between the benefits and the computational efforts, and potentially to provide suitable environments.

B. Definition sets

Definition sets for object elements can be created and used very flexibly, e.g., word or string definitions. Therefore, a reasonable set of elements can be defined for the respective purpose, especially:

- Definition sets can contain appropriate material, e.g., text or classification.
- Groups of elements can be created.
- Contributing elements can be subsummarised.
- Definition sets can be kept persistent and volatile.
- Definition set elements can be weighted, e.g., by parameterisation of context-sensitive code growth.
- Context sensitive definition sets can be referenced with data objects.
- Content can be described with multiple, complementary definition sets.
- Any part of the content can be defined as elements.

The Content Factors can be computed for any object, e.g., for text and other parts of content. Nevertheless, the above definition sets for normed factors are intended to be used with one type of elements.

C. Normed application

$\overline{\kappa}_{\mathrm{B}}$ is a normed quantity. Norming is a mathematical procedure, by which the interesting quantity (e.g., vector, operator, function) is modified by multiplication in a way that after the norming the application of respective functionals delivers 1. The respective $\overline{\kappa}_{\mathrm{B}}$ Content Factor can be used to create a weighting on objects, e.g., multiplying the number of elements with the respective factor value.

VI. VALUE AND APPRECIATION

The value of objects and collections, e.g., regarding libraries [23], is matter of discussion [24]. Nevertheless, bibliometrics is a very disputable practice with highly questionable results from content point of view and relevance.

Whereas some data is of high scientific value it may currently have less or no economic value [25]. Studies on data genomics has delivered a lot of information [26] on the related aspects.

It is interesting to see that on the other hand the form of the content is associated with resulting citations, e.g., more figures may lead to more citations [27]. However, visual information in scientific literature [28] is only one small aspect, it may also have some value.

The demand for better information and reference services is obvious for scientific knowledge, however in rare cases the question if separate services [29] may be required is still asked [30]. A large implementation, which cannot recognise the value of data and knowledge in huge heterogeneous data sources is surely neither a viable solution nor a desirable state. Basic definitions for "data-centric" and "Big Data" in this context are emphasizing the value:

"The term data-centric refers to a focus, in which data is most relevant in context with a purpose. Data structuring, data shaping, and long-term aspects are important concerns. Data-centricity concentrates on data-based content and is beneficial for information and knowledge and for emphasizing their value. Technical implementations need to consider distributed data, non-distributed data, and data locality and enable advanced data handling and analysis. Implementations should support separating data from technical implementations as far as possible." [31].

"The term Big Data refers to data of size and/or complexity at the upper limit of what is currently feasible to be handled with storage and computing installations. Big Data can be structured and unstructured. Data use with associated application scenarios can be categorised by volume, velocity, variability, vitality, veracity, value, etc. Driving forces in context with Big Data are advanced data analysis and insight. Disciplines have to define their 'currency' when advancing from Big Data to Value Data." [31].

The long-term creation and development of knowledge values as well as next generation services require additional and improved features and new algorithms for taking advantage of high quality knowledge resources and increasing the quality of results.

VII. Application Scenarios and Implementations

The implementation has been created for the primary use with knowledge resources' objects (`lxcontfact`). This means handling of any related content, e.g., documentation, keywords, classification, transliterations, and references. The respective objects were addressed as Content Factor Object (CFO) (standard file extension `.cfo`) and the definition sets as Content Factor Definition (CFD) (standard file extension `.cfd`).

A. Case study: Computing complementation and properties

The following case, consisting of a sequence of short examples shows a knowledge resources object (Figure 7), and three pairs of complementary CONTFACT definition sets and the according $\overline{\kappa}_{\mathrm{B}}$ computed for the knowledge resources object and respective definition sets (Figures 8 and 9; 10 and 11; 12 and 13).

```
object A      %-GP%-XX%---: object A            [A, B, C, D, O]:
              %-GP%-EN%---:                     A B C D O
              %-GP%-EN%---:                     A B C D O
              %-GP%-EN%---:                     A B C D O
              %-GP%-EN%---:                     A B C D O
              %-GP%-EN%---:                     A B C D O
```

Figure 7. Artificial knowledge resources object (LX Resources, excerpt).

Here, the algorithm can count in object entry name (right "object A") and label, keywords (in brackets), and object documentation (lower right block).

```
% (c) LX-Project, 2015, 2016
{A}:=\bA\b
{O}:=\bO\b
```

Figure 8. CONTFACT definition set 1 of 3 (LX Resources, excerpt).

The definition set defines {A} and {O}. The definitions are case sensitive for this discovery. We can compute $\overline{\kappa}_{\mathrm{B}}$ (Figure 9) according to the knowledge resources object and definition set.

```
CONTFACT:BEGIN
CONTFACT:20160117-175904:AU:{A}{A}{O}{A}{O}{A}{O}{A}{O}{A}{O}{A}{O}/32
CONTFACT:20160117-175904:AS:{A}{A}{A}{A}{A}{A}{A}{O}{O}{O}{O}{O}{O}/32
CONTFACT:20160117-175904:M:{A}:=\bA\b
CONTFACT:20160117-175904:M:{O}:=\bO\b
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSDEF=2
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSALL=32
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSMAT=13
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSCFO=.40625000
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSKWO=2
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSLAN=1
CONTFACT:20160117-175904:M:INFO:OBJECTELEMENTSOBJ=object A
CONTFACT:20160117-175904:M:INFO:OBJECTELEMENTSDCM=(c) LX-Project, 2015, 2016
CONTFACT:20160117-175904:M:INFO:OBJECTELEMENTSMTX=LX Foundation Scientific
 Resources; Object Collection
CONTFACT:20160117-175904:M:INFO:OBJECTELEMENTSAUT=Claus-Peter R\"uckemann
CONTFACT:END
```

Figure 9. NBCF $\overline{\kappa}_{\mathrm{B}}$ computed for knowledge resources object and definition set 1 (LX Resources, excerpt).

The result is shown in a line-oriented representation, each line carrying the respective date-time code for all the core, statistics, and additional information. The second complementary set (Figure 10) defines {B} and {D} with its $\overline{\kappa}_{\mathrm{B}}$ (Figure 11).

```
% (c) LX-Project, 2015, 2016
{B}:=\bB\b
{D}:=\bD\b
```

Figure 10. CONTFACT definition set 2 of 3 (LX Resources, excerpt).

```
CONTFACT:BEGIN
CONTFACT:20160117-175904:AU:{B}{D}{B}{D}{B}{D}{B}{D}{B}{D}{B}{D}/32
CONTFACT:20160117-175904:AS:{B}{B}{B}{B}{B}{B}{D}{D}{D}{D}{D}{D}/32
CONTFACT:20160117-175904:M:{B}:=\bB\b
CONTFACT:20160117-175904:M:{D}:=\bD\b
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSDEF=2
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSALL=32
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSMAT=12
CONTFACT:20160117-175904:M:STAT:OBJECTELEMENTSCFO=.37500000
...
```

Figure 11. NBCF $\overline{\kappa}_{\mathrm{B}}$ computed for knowledge resources object and definition set 2 (LX Resources, excerpt).

The third complementary set (Figure 12) defines {C}.

```
% (c) LX-Project, 2015, 2016
{C}:=\bC\b
```

Figure 12. CONTFACT definition set 3 of 3 (LX Resources, excerpt).

The resulting $\overline{\kappa}_{\mathrm{B}}$ is shown in the excerpt (Figure 13).

```
CONTFACT:BEGIN
CONTFACT:20160117-175905:AU:{C}{C}{C}{C}{C}{C}/32
CONTFACT:20160117-175905:AS:{C}{C}{C}{C}{C}{C}/32
CONTFACT:20160117-175905:M:{C}:=\bC\b
CONTFACT:20160117-175905:M:STAT:OBJECTELEMENTSDEF=1
CONTFACT:20160117-175905:M:STAT:OBJECTELEMENTSALL=32
CONTFACT:20160117-175905:M:STAT:OBJECTELEMENTSMAT=6
CONTFACT:20160117-175905:M:STAT:OBJECTELEMENTSCFO=.18750000
...
```

Figure 13. NBCF $\overline{\kappa}_{\mathrm{B}}$ computed for knowledge resources object and definition set 3 (LX Resources, excerpt).

The sum of all elements considered for $\overline{\kappa}_{\mathrm{B}}$ by the respective CONTFACT algorithm in an object is 100 percent. Here, the overall number of

- definitions is $2 + 2 + 1 = 5$,
- elements is 32 (25, 5 keywords, 2 name and label),
- matches is $13 + 12 + 6 = 31$.

The sum of the aggregated $\overline{\kappa}_{\mathrm{B}}$ values for complementary definitions and all relevant elements results in

$$0.40625000 + 0.37500000 + 0.18750000 + 1/32 = 1$$

This also means the used definitions completely cover the elements in an object with their description.

B. Case study: Complex resources and discovery scenario

The data used here is based on the content and context from the knowledge resources, provided by the LX Foundation Scientific Resources [22]. The LX knowledge resources' structure and the classification references [32] based on UDC [33] are essential means for the processing workflows and evaluation of the knowledge objects and containers.

Both provide strong multi-disciplinary and multi-lingual support. For this part of the research all small unsorted excerpts of the knowledge resources objects only refer to main UDC-based classes, which for this part of the publication are taken from the Multilingual Universal Decimal Classification Summary (UDCC Publication No. 088) [34] released by the UDC Consortium under the Creative Commons Attribution Share Alike 3.0 license [35] (first release 2009, subsequent update 2012).

The excerpts (Figures 14, 15, 16), show a CFO from the knowledge resources a CFD and the computed CONTFACT.

```
Vesuvius [Volcanology, Geology, Archaeology]:
         (lat.) Mons Vesuvius.
         (ital.) Vesuvio.
         Volcano, Gulf of Naples, Italy.
         Complex volcano (compound volcano). Stratovolcano, large cone (Gran
          Cono).
         ...
         The most well known antique settlements at the Vesuvius are \lxidx{
          Pompeji}, \lxidx{Herculaneum}, and \lxidx{Stabiae}.
         s. also seismology, phlegra, Solfatara
         %%IML: keyword: volcano, Vesuvius, Campi Flegrei, phlegra, scene of
          fire, Pompeji, Herculaneum, volcanic ash, lapilli, catastrophe,
          climatology, eruption, lava, gas ejection, Carbon Dioxide
         %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(4+37+23+24)=12=14
         ...
         Object:             Volcanic material.
         Object-Type:        Realia object.
         Object-Location:    Vesuvius, Italy.
         Object-FindDate:    2013-10-00
         Object-Discoverer:  Birgit Gersbeck-Schierholz, Hannover, Germany.
         Object-Photo:       Claus-Peter Rückemann, Minden, Germany.
         %%IML: media: YES 20131000 {LXC:DETAIL--M-} {UDC:(0.034)(044)770}
          LXDATASTORAGE://...img_3824.jpg
         %%IML: UDC-Object:[551.21+55]:[911.2](37+4+23)=12
         %%IML: UDC: 551.21      :: Vulcanicity. Vulcanism. Volcanoes. Eruptive
          phenomena. Eruptions
         %%IML: UDC: 55          :: Earth Sciences. Geological sciences
         %%IML: UDC: 911.2       :: Physical geography
```

Figure 14. Knowledge resources object (geosciences collection, LX, excerpt).

Labels, language fields, and spaces were stripped. A knowledge object can contain any items required, e.g., including storing data, documentation, classification, keywords, algorithms, references, implementations, in any languages and representations, allowing support tables and algorithms. An object can also include subobjects and references [36] as shown here. Examples of application scenarios for the Content Factor method range from libraries, natural sciences and archaeology, statics, architecture, risk coverage, technology to material sciences [37].

```
% (c) LX-Project, 2009, 2015
{Ve}:=Vesuvius
{Vo}:=\b[Vv]olcano
{Po}:=Pompe[ji]i
{UDC:55}:=Geology
{UDC:volcano}:=UDC.*\b911\b.*\b55\b
```

Figure 15. CONTFACT definition set (geosciences collection, LX, excerpt).

The definition sets can contain anything required for the definitions and additional information for the respective Content Factor implementation, e.g., definitions of elements and groups as well as comments. The left side defines the element used in the Content Factor and the right side states the matching element components. Left value and right value are separated by "`:=`" for an active definition.

```
CONTFACT:BEGIN
CONTFACT:20160130-235804:AU:{Ve}{Vo}{UDC:55:geology}{Ve}{Ve}{Vo}{Vo}{Vo}{Vo}{Vo
 }{Vo}{Vo}{Ve}{Ve}{Po}{Ve}{Po}{Ve}{Ve}{Ve}{Vo}{Vo}{Vo}{Vo}{Vo}{UDC:volcano}{Vo}{
 Vo}/319
CONTFACT:20160130-235804:AS:{Po}{Po}{UDC:55:geology}{UDC:volcano}{Ve}{Ve}{Ve}{Ve
 }{Ve}{Ve}{Ve}{Ve}{Ve}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{Vo}{
 Vo}/319
CONTFACT:20160130-235804:M:{Ve}:=Vesuvius
CONTFACT:20160130-235804:M:{Vo}:=\b[Vv]olcano
CONTFACT:20160130-235804:M:{Po}:=Pompe[ji]i
CONTFACT:20160130-235804:M:{UDC:55:geology}:=Geology
CONTFACT:20160130-235804:M:{UDC:volcano}:=UDC.*\b911\b.*\b55\b
CONTFACT:20160130-235804:M:STAT:OBJECTELEMENTSDEF=5
CONTFACT:20160130-235804:M:STAT:OBJECTELEMENTSALL=319
CONTFACT:20160130-235804:M:STAT:OBJECTELEMENTSMAT=28
CONTFACT:20160130-235804:M:STAT:OBJECTELEMENTSCFO=.09180304
CONTFACT:20160130-235804:M:INFO:OBJECTELEMENTSDCM=(c) LX-Project, 2009, 2015
...
CONTFACT:END
```

Figure 16. NBCF $\overline{\kappa}_{\mathrm{B}}$ computed for knowledge resources object and definition set (geosciences collection, LX Resources, excerpt).

The left value can include braces (e.g., curly brackets) in order to support the specification and identification of the left value. The right value can include common representations of pattern specification. The result of which can be seen from the computed CONTFACT.

The example patterns follow the widely used Perl (Practical Extraction and Report Language) regular expressions [38], e.g., `\b` for word boundaries and `[...]` and multiple choices of characters at a certain position.

C. Definitions

Definitions link the elements used in an Content Factor with a certain content. The following figures show examples for a collection object (Figure 17), a related definition set (Figure 18), and a computed CONTFACT (Figure 19).

```
object A      %-GP%-XX%---: object A           [A, B, C, D, O]:
              %-GP%-EN%---:                    A B C D O
              %-GP%-EN%---:                    A B C D O
              %-GP%-EN%---:                    A B C D O
              %-GP%-EN%---:                    A B C D O
              %-GP%-EN%---:                    A B C D O
```

Figure 17. Example of single LX collection object, used with CONTFACT (LX Resources, excerpt).

```
% (c) LX-Project, 2015, 2016
{A}:=\bA\b
{Letter_B}:=\bB\b
{charC}:=\bC\b
{004}:=\bD\b
{Omega}:=\bO\b
```

Figure 18. Example of CONTFACT definitions (LX Resources, excerpt).

The definitions (braced) define single letters in this case. In this representation, the CONTFACT computation sees the right side of the object entry (right of the language flags 'EN' and 'XX').

The computed CONTFACT (Figure 19) uses the braced definitions for building the CONTFACT core.

```
CONTFACT:BEGIN
CONTFACT:20160829-094358:AU:{A}{A}{Letter_B}{charC}{004}{Omega}{A}{Letter_B}{
 charC}{004}{Omega}{A}{Letter_B}{charC}{004}{Omega}{A}{Letter_B}{charC}{004}{
 Omega}{A}{Letter_B}{charC}{004}{Omega}{A}{Letter_B}{charC}{004}{Omega}/32
CONTFACT:20160829-094358:AS:{004}{004}{004}{004}{004}{004}{A}{A}{A}{A}{A}{A}{A}{
 charC}{charC}{charC}{charC}{charC}{charC}{Letter_B}{Letter_B}{Letter_B}{
 Letter_B}{Letter_B}{Letter_B}{Omega}{Omega}{Omega}{Omega}{Omega}{Omega}/32
CONTFACT:20160829-094358:M:{A}:=\bA\b
CONTFACT:20160829-094358:M:{Letter_B}:=\bB\b
CONTFACT:20160829-094358:M:{charC}:=\bC\b
CONTFACT:20160829-094358:M:{004}:=\bD\b
CONTFACT:20160829-094358:M:{Omega}:=\bO\b
CONTFACT:20160829-094358:M:STAT:OBJECTELEMENTSDEF=5
CONTFACT:20160829-094358:M:STAT:OBJECTELEMENTSALL=32
CONTFACT:20160829-094358:M:STAT:OBJECTELEMENTSMAT=31
CONTFACT:20160829-094358:M:STAT:OBJECTELEMENTSCFO=.96875000
CONTFACT:20160829-094358:M:STAT:OBJECTELEMENTSKWO=2
CONTFACT:20160829-094358:M:STAT:OBJECTELEMENTSLAN=1
CONTFACT:20160829-094358:M:INFO:OBJECTELEMENTSOBJ=object A
CONTFACT:20160829-094358:M:INFO:OBJECTELEMENTSDCM=(c) LX-Project, 2015, 2016
CONTFACT:20160829-094358:M:INFO:OBJECTELEMENTSMTX=LX Foundation Scientific
 Resources; Object Collection
CONTFACT:20160829-094358:M:INFO:OBJECTELEMENTSAUT=Claus-Peter R\"uckemann
CONTFACT:END
```

Figure 19. Example of CONTFACT output (LX Resources, excerpt).

The left side values can be used in the core. For application purposes these values can internally be mapped or referenced to other unique values or representations like meta-levels and numbering schemes, e.g., if this practice may provide benefits for a certain implementation.

D. Case study: Rhythm matching and core sequences

As soon as Content Factors have been computed for an object the patterns can be compared with pattern of other objects. The Content Factor method allows to compare occurrences of relevant elements in objects in many ways. The following example shows the "rhythm matching" method on the basis of an object and a definition set (Figure 20), for two computed unsorted CONTFACT core sequences (Figures 21, 22).

```
% (c) LX-Project, 2009, 2015, 2016
{Am}:=\b[Aa]mphora
{Ce}:=[Cc]eramic
{Gr}:=\b[Gg]reek\b
{Pi}:=[Pp]itho[is]
{Ro}:=\b[Rr]oman\b
{Tr}:=[Tt]ransport
{Va}:=[Vv]ases
```

Figure 20. Example of CONTFACT definition set, geoscientific and archaeological resources (LX Resources, excerpt).

```
CONTFACT:20160828-215751:AU:{Am}{Gr}{Ce}{Gr}{Gr}{Ro}{Am}{
 Gr}{Am}{Am}{Va}{Am}{Ce}{Am}{Am}{Pi}{Pi}{Tr}{Tr}{Gr}{Ro}{
 Am}{Am}{Tr}{Am}{Gr}{Ro}{Am}{Tr}{Am}{Tr}{Am}{Am}{Am}{Am}{
 Tr}{Am}{Am}{Am}{Am}{Gr}{Ce}{Ce}{Tr}/474
```

Figure 21. CONTFACT rhythm matching: Computed core for same object (before modification) and definition set (LX Resources, excerpt).

```
CONTFACT:20160828-231806:AU:{Am}{Gr}{Ce}{Gr}{Gr}{Ro}{Am}{
 Gr}{Am}{Am}{Va}{Am}{Ce}{Am}{Am}{Pi}{Pi}{Tr}{Tr}{Gr}{Ro}{
 Am}{Am}{Tr}{Am}{Gr}{Ro}{Am}{Tr}{Am}{Tr}{Am}{Am}{Am}{Am}{
 Tr}{Am}{Am}{Am}{Am}{Gr}{Ce}{Ce}{Tr}{Ce}{Ce}{Tr}{Ce}{Ce}{
 Tr}{Ce}{Ce}{Ce}{Ce}{Pi}{Pi}{Am}/589
```

Figure 22. CONTFACT rhythm matching: Computed core for same object (after modification) and definition set (LX Resources, excerpt).

The comparison shows that relevant passages were appended to the object (italics font). Relevant regarding the rhythm matching means relevant from the object and definition set. Even short sequences like {Am}{Gr}{Ce} and even when sorted like {Am}{Ce}{Gr} can be relevant and significant in order to compute factors and identify and compare objects. The Content Factor method does not have built-in or intrinsic limitations specifying certain ways of further use, e.g., with comparisons and analysis.

Unsorted CONTFACT are more likely to describe objects and quality, including their internal organisation. Sorted CONTFACT tend to describe objects by their quantities, with reduced focus on their internal organisation.

Objects with larger amount of documentation maybe candidates for unsorted CONTFACT. Objects, e.g., with factual, formalised content maybe candidates for sorted CONTFACT. Combining several methods in a workflow is possible.

Anyhow, the further use of the CONTFACT core, e.g., sorting the core data for a certain comparison, is a matter of application and purpose with respective data.

E. Object Comparison

The Content Factor can be used with arbitrary data, e.g., with knowledge resources, for all objects, referenced data and information, collections, and containers. The example (Figure 23) shows an excerpt of a collection object in an arbitrary stage of creation. The excerpt contains some elements, which can be relevant regarding Content Factor and respective definitions sets.

```
Amphora [Archaeology, Etymology]:
        (greek) amphoreus = ceramic container with two handles.
        (greek) amphí = on both sides.
        (greek) phérein = carry.
        The Greco-Roman term amphora is of ancient Greek origin and has been
         developed during the Bronze Age.
        Container of a characteristic shape and size and two handles.
        Amphoras are a subgroup of antique \lxidx{vases}.
        Most amphoras are made from ceramic material, often clay.
        There are rare amphoras made from stone and metal, like bronze, silver
         or gold.
        Amphoras typically have a volume of 5--50\UD{l}, in some cases 100 or
         more litres.
        Larger containers mostly had the purpose of storage only,
        named pithos and pithoi (pl.).
        ...
        Object:            Amphora, transport.
        Object-Type:       Realia object.
        Object-Relocation: Museu d'Arqueologia de Catalunya, Barcelona, Spain.
        %%IML: media: YES 20111027 {LXC:DETAIL----} {UDC:(0.034)(460)770}
         LXDATASTORAGE://.../img_5831.jpg
        %%IML: media: YES 20111027 {LXC:DOC-------} {UDC:(0.034)(460)770}
         LXDATASTORAGE://.../img_5831.jpg
        %%IML: UDC-Object:[902+903.2+904]+738+738.8+656+(37)+(4)
        %%IML: UDC-Relocation:069.51+(4)+(460)+(23)
        %%IML: label: {MUSEUM-Material:    Ceràmica}
```

Figure 23. Example of LX collection object, matter of change, used with CONTFACT (LX Resources, excerpt).

Figure 24 shows a definition set as used with objects as with the example (Figure 23), instances of which are to be compared.

```
% (c) LX-Project, 2009, 2015, 2016
{Am}:=\b[Aa]mphora
{Ce}:=[Cc]eramic
{Gr}:=\b[Gg]reek\b
{Pi}:=[Pp]itho[is]
{Ro}:=\b[Rr]oman\b
{Tr}:=[Tt]ransport
{Va}:=[Vv]ases
```

Figure 24. Example of CONTFACT definitions, geoscientific and archaeological resources (LX Resources, excerpt).

Definition sets are used when the Content Factor is applied to objects. This definition set is used for comparing an instance of an object with an instance of the same object, which has been modified later.

Figure 25 presents the resulting Content Factor of this implementation, including $\overline{\kappa}_{\mathrm{B}}$ for this context, the core lines (lines 2–3), unsorted (U) and sorted (S), definition set lines (lines 4–10) resolving the used elements, and integrated additional information and statistics (lines 11-20).

```
1  CONTFACT:BEGIN
2  CONTFACT:20160829-123531:AU:{Am}{Gr}{Ce}{Gr}{Gr}{Ro}{Am}{Gr}{Am}{Am}{Va}{Am}{Ce
    }{Am}{Am}{Pi}{Pi}{Tr}{Tr}{Gr}{Ro}{Am}{Am}{Tr}{Am}{Gr}{Ro}{Am}{Tr}{Am}{Tr}{Am}{
    Am}{Am}{Am}{Tr}{Am}{Am}{Am}{Am}{Gr}{Ce}{Ce}{Tr}/496
3  CONTFACT:20160829-123531:AS:{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am
    }{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Ce}{Ce}{Ce}{Ce}{Gr}{Gr}{Gr}{Gr}{Gr}{Gr}{Gr}{Pi}{
    Pi}{Ro}{Ro}{Ro}{Tr}{Tr}{Tr}{Tr}{Tr}{Tr}{Tr}{Va}/496
4  CONTFACT:20160829-123531:M:{Am}:=\b[Aa]mphora
5  CONTFACT:20160829-123531:M:{Ce}:=[Cc]eramic
6  CONTFACT:20160829-123531:M:{Gr}:=\b[Gg]reek\b
7  CONTFACT:20160829-123531:M:{Pi}:=[Pp]itho[is]
8  CONTFACT:20160829-123531:M:{Ro}:=\b[Rr]oman\b
9  CONTFACT:20160829-123531:M:{Tr}:=[Tt]ransport
10 CONTFACT:20160829-123531:M:{Va}:=[Vv]ases
11 CONTFACT:20160829-123531:M:STAT:OBJECTELEMENTSDEF=7
12 CONTFACT:20160829-123531:M:STAT:OBJECTELEMENTSALL=496
13 CONTFACT:20160829-123531:M:STAT:OBJECTELEMENTSMAT=44
14 CONTFACT:20160829-123531:M:STAT:OBJECTELEMENTSCFO=.09282680
15 CONTFACT:20160829-123531:M:STAT:OBJECTELEMENTSKWO=2
16 CONTFACT:20160829-123531:M:STAT:OBJECTELEMENTSLAN=2
17 CONTFACT:20160829-123531:M:INFO:OBJECTELEMENTSOBJ=Amphora
18 CONTFACT:20160829-123531:M:INFO:OBJECTELEMENTSDCM=(c) LX-Project, 2009, 2015,
    2016
19 CONTFACT:20160829-123531:M:INFO:OBJECTELEMENTSMTX=LX Foundation Scientific
    Resources; Object Collection
20 CONTFACT:20160829-123531:M:INFO:OBJECTELEMENTSAUT=Claus-Peter R\"uckemann
21 CONTFACT:END
```

Figure 25. Computed CONTFACT, geoscientific and archaeological resources (LX Resources, excerpt).

Additional information maybe required for supporting an integration with an application scenario and practical implementation can be added very flexibly.

Figure 26 presents the resulting core lines after changes to the object.

```
CONTFACT:20160829-123532:AU:{Am}{Gr}{Ce}{Gr}{Gr}{Ro}{Am}{Gr}{Am}{Am}{Va}{Am}{Ce
 }{Am}{Am}{Pi}{Pi}{Tr}{Tr}{Gr}{Ro}{Am}{Am}{Tr}{Am}{Gr}{Ro}{Am}{Tr}{Am}{Tr}{Am}{
 Am}{Am}{Am}{Tr}{Am}{Am}{Am}{Am}{Gr}{Ce}{Ce}{Tr}{Ce}{Ce}{Tr}{Ce}{Ce}{Tr}{Ce}{Ce
 }{Ce}{Ce}{Pi}{Pi}{Am}/510
CONTFACT:20160829-123532:AS:{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am
 }{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{Ce}{Ce}{Ce}{Ce}{Ce}{Ce}{Ce}{Ce}{Ce}{Ce}{Ce}{
 Ce}{Gr}{Gr}{Gr}{Gr}{Gr}{Gr}{Gr}{Pi}{Pi}{Pi}{Pi}{Ro}{Ro}{Ro}{Tr}{Tr}{Tr}{Tr}{Tr
 }{Tr}{Tr}{Tr}{Tr}{Va}/510
```

Figure 26. Computed CONTFACT core after changes, geoscientific and archaeological resources (LX Resources, excerpt).

This method can be used to document and analyse the development of objects over time. It is possible to compare different objects or instances as well as comparing sequences and movements of sequences inside an object. In principle there is no limitation for changes, which can be considered when comparing results. Comparing results with arbitary changes can be reasonable for an application scenario.

Anyway, if one parameter changes at a time then the interpretation from a comparison is most unambiguousness.

F. *Multi-lingual Discovery and Concordances*

The Content Factor method can also be used for discovery procedures based on multi-lingual definitions (Figures 27, 28, 29). Figure 27 excerpts complementary relevant parts of the collection object (Figure 23).

The parts are relevant for this application regarding Content Factor and respective definitions sets.

```
Amphora       %-GP%-XX%---: Amphora               [Archaeology, Etymology]:
              %-GP%-EN%---:                       (greek) amphoreus = ceramic
               container with two handles.
              %-GP%-EN%---:                       (greek) amphí = on both sides.
              %-GP%-EN%---:                       (greek) phérein = carry.
              %-GP%-DE%---:                       (altgriech.) amphoreus =
               zweihenkliges Tongefäß.
              %-GP%-DE%---:                       (griech.) amphí = auf beiden
               Seiten.
              %-GP%-DE%---:                       (griech.) phérein = tragen.
...
              %-GP%-XX%---:                       catalan:   \lxidxlangeins{
               àmphora}
              %-GP%-XX%---:                       english:   \lxidxlangeins{
               amphore, amphorae / amphoras (pl.)}
              %-GP%-XX%---:                       french:    \lxidxlangeins{
               amphora}
              %-GP%-XX%---:                       german:    \lxidxlangeins{
               Amphore}
              %-GP%-XX%---:                       greek:     \lxidxlangzwei{
               amphora, amphoreas}{$\alpha\mu\varphi o\rho\epsilon\alpha\
               varsigma$}
              %-GP%-XX%---:                       italian:   \lxidxlangeins{
               anfora, anfore}
              %-GP%-XX%---:                       latin:     \lxidxlangeins{
               amphora}
              %-GP%-XX%---:                       spanish:   \lxidxlangeins{
               àmfora}
```

Figure 27. Example LX collection object, multi-lingual elements, used with CONTFACT (LX Resources, excerpt).

Regarding this case study, the excerpt contains multi-lingual entries (EN, DE) in an object as well as multi-lingual elements in the multi-lingual entries, including translations and transcriptions. Figure 28 excerpts a CONTFACT definition set, which has been generated from concordance references.

```
% (c) LX-Project, 2009, 2015, 2016
{Am}:=[Aa]mphora
{AA}:=[Àà]mphora
{Ae}:=[Aa]mphore
{An}:=[Aa]nfor[ae]
{Af}:=[Àà]mfora
```

Figure 28. Example of CONTFACT definitions, generated from concordance references (LX Resources, excerpt).

In this case different representations of the same term are defined. The resulting core will contain the different distinctable occurences and supports a complex analysis. Figure 29 excerpts the resulting Content Factor core lines.

```
CONTFACT:20160101-220551:AU:{Am}{Ae}{Ae}{Am}{Am}{Ae}{Am}{Am}{Am}{Ae}{Am}{Ae}{Ae
 }{Am}{Am}{Ae}{Am}{Ae}{Am}{Ae}{Ae}{Am}{Am}{Am}{Am}{Am}{Am}{Am}{AA}{Ae}{Am}{Am}{
 Am}{Ae}{Am}{Ae}{An}{An}{Am}{Af}{Ae}{Am}/488
```

Figure 29. Computed CONTFACT core only containing multi-lingual/concordances information (LX Resources, excerpt).

The resulting Content Factor allows to document and analyse multi-lingual entries as well as concordances in many ways, e.g., all the data, dedicated entries or translations. The method also allows to create relations from the context and deduct relevances.

G. *Concordances Discovery*

Knowledge processing can benefit from creating concordances with the conceptual knowledge [39] as well as concordances can be used with advanced association processing [40].

The Content Factor works with classification references the same way as with patterns and definitions. The application of concordances for the use with Content Factors is therefore comparable but introduces additional complexity at the level of evaluating concordances.

The differences in classification and concordances are resulting from the different level of detail in the collections and containers as well as in different potential of the various classification schemes to describe certain knowledge as can be seen from the different depth of classification. In integration, together the concordances can create valuable references in depth and width to complementary classification schemes and knowledge classified with different classification.

The term concordance is not only used in the simple traditional meaning. Instead, the organisation is that of a meta-concordances concept. That results from the use of universal meta-classification, which in turn is used to classify and integrate classifications. The samples include simple classifications from UDC, Mathematics Subject Classification (MSC) [41], Library of Congress Classification (LCC) [42], and Physics and Astronomy Classification Scheme (PACS) [43].

The Universal Classified Classification (UCC) entries contain several classifications. The UCC blocks provide concordances across the classification schemes. The object classification is associated with the items associated with the object whereas the container classification is associated with the container, which means it refers to all objects in the containers.

Figure 30 excerpts a definition set based on UCC entries.

```
% (c) LX-Project, 2009, 2015, 2016
{UCC:}:=
{UCC:UDC2012:}:=UDC2012:551.21
{UCC:UDC2012:}:=UDC2012:551
{UCC:UDC2012:}:=UDC2012:902/908
{UCC:MSC2010:}:=MSC2010:86,86A17,86A60
{UCC:LCC:}:=LCC:QE521-545
{UCC:LCC:}:=LCC:QE1-996.5
{UCC:LCC:}:=LCC:QC801-809
{UCC:LCC:}:=LCC:CC1-960,CB3-482
{UCC:PACS2010:}:=PACS2010:91.40.-k
{UCC:PACS2010:}:=PACS2010:91.65.-n,91.
```

Figure 30. Concordances information: UCC (LX Resources, excerpt).

In general, the typification for taking advantage of concordances can consider all the according levels spanned by the classification trees. In practice, organising concordances discovery means to care for the individual typecasting, mapping, and referencing with the implementation.

H. Element Groups

The algorithm can be used with discovery procedures using definitions based on element groups (Figures 31, 32, 33).

```
object or     %-GP%-EN%---: object or           [Alternatives]:
              %-GP%-EN%---:                     Archaeology, Archeology.
              %-GP%-DE%---:                     Archäologie.
              %-GP%-EN%---:                     Underwaterarchaeology,
               Underwaterarcheology.
              %-GP%-DE%---:                     Unterwasserarchäologie.
              %-GP%-EN%---:                     archaeology, archeology.
              %-GP%-DE%---:                     ...archäologie.
```

Figure 31. Example LX collection object for computing Content Factors including element groups (LX Resources, excerpt).

This example (Figure 31) defines a collection object with several main lines. The lines contain terms composed in two languages, with and without umlauts, and using upper case and lower case. A definition set containing an element group delivering several hits is given in Figure 32.

```
% (c) LX-Project, 2009, 2015, 2016
{Boundary_A}:=\b[Aa]rchaeology\b|\b[Aa]rcheology\b|\b[Aa]rchäologie\b
```

Figure 32. Example definition set for computing Content Factors including element groups (LX Resources, excerpt).

The definition set defines an element group of terms with and without umlauts, all choosing lower case and upper case terms with word boundaries.

```
CONTFACT:BEGIN
CONTFACT:20160829-220828:AU:{Boundary_A}{Boundary_A}{Boundary_A}{Boundary_A}{
 Boundary_A}{Boundary_A}/12
CONTFACT:20160829-220828:AS:{Boundary_A}{Boundary_A}{Boundary_A}{Boundary_A}{
 Boundary_A}{Boundary_A}/12
CONTFACT:20160829-220828:M:{Boundary_A}:=\b[Aa]rchaeology\b|\b[Aa]rcheology\b|\b
 [Aa]rchäologie\b
CONTFACT:20160829-220828:M:STAT:OBJECTELEMENTSDEF=1
CONTFACT:20160829-220828:M:STAT:OBJECTELEMENTSALL=12
CONTFACT:20160829-220828:M:STAT:OBJECTELEMENTSMAT=6
CONTFACT:20160829-220828:M:STAT:OBJECTELEMENTSCFO=.50000000
CONTFACT:20160829-220828:M:STAT:OBJECTELEMENTSKWO=1
CONTFACT:20160829-220828:M:STAT:OBJECTELEMENTSLAN=2
CONTFACT:20160829-220828:M:INFO:OBJECTELEMENTSOBJ=object or
CONTFACT:20160829-220828:M:INFO:OBJECTELEMENTSDCM=(c) LX-Project, 2009, 2015,
 2016
CONTFACT:20160829-220828:M:INFO:OBJECTELEMENTSMTX=LX Foundation Scientific
 Resources; Object Collection
CONTFACT:20160829-220828:M:INFO:OBJECTELEMENTSAUT=Claus-Peter R\"uckemann
CONTFACT:END
```

Figure 33. Example CONTFACT output including element groups (LX Resources, excerpt).

This results in one definition and six matches from twelve elements for the CONTFACT: The definitions define groups of alternative element representation subsummarised in the same element group. The subsummarisation may be created for specific purposes, e.g., for different writing for a certain term.

In a Perl notation alternatives are separated with pipe symbols (`|`). The right side value is used accordingly for counting. The two commented examples in the definition set show using lower and upper case specification for letter and defining word boundaries.

In principle, the definitions are subject of the respective application scenario and creator. Anyhow, it is a good practice to think about the sort order, e.g., to consider more special/-conditions first. In a Content Factor implementation this can mean to use a sort key, a priority or simply place the respective groups on top.

Here, the definitions can include substring alternatives, boundary delimited first-letter case insensitive alternatives, and first-letter case insensitive substring alternatives.

With element groups the alternatives are counted for the respective element group. The implementation of the Content Factor has to make sure to handle the alternatives and the counting appropriately.

VIII. Processing and Computation

It is advantageous if algorithms used with arbitrary content can be adopted for different infrastructure and data-locality, e.g., with different computing, network, and storage resources. This is especially helpful when data quantities are large. Therefore, scalability, modularisation, and dynamical use as well as parallelisation and persistence of individual stages of computation should be handled in flexible ways.

A. Scalability, modularisation, and dynamical use

The algorithms can be used for single objects as well as for large collections and containers, containing millions of entries each. Not only simulations but more and more Big Data analysis is conducted using High Performance Computing. Therefore, data-centric models are implemented expanding the traditional compute-centric model for an integrated approach [44]. In addition to the data-centric knowledge resources, the Content Factor computation routines allow a modularised and dynamical use.

The parts required for an implementation computing a Content Factor can be modularised, which means that not only a Content Factor computation can be implemented as a module but even core, definitions, and additional parts can be computed by separate modules.

Sequences of routine calls can be used in order to modularise complex workflows. The sequence of routine calls used for examples in this case study shows the principle and modular application of respective functions (Figure 34). The modules create an entity for the implemented Content Factor (`contfactbegin` to `contfactend`). They include labels, date, unsorted elements and so on as well as statistics and additional information.

The possibility to modularise the routine calls even within the Content Factor provides the features increased flexibility and scalability, which can be used for individual implementations optimised for distributed and non-distributed Big Data.

```
contfactbegin

contfact
contfactdate
contfacttype
contfactelementsu
contfactref
contfactsum

contfact
contfactdate
contfacttypes
contfactelementss
contfactref
contfactsum

contfactdef

contfact
contfactdate
contfacttypestat
contfact_stat_def_lab
contfact_stat_def

contfact
contfactdate
contfacttypestat
contfact_stat_all_lab
contfact_stat_all

contfact
contfactdate
contfacttypestat
contfact_stat_mat_u_lab
contfact_stat_mat_u

contfact
contfactdate
contfacttypestat
contfact_stat_cfo_lab
contfact_stat_cfo

contfact
contfactdate
contfacttypestat
contfact_stat_kwo_lab
contfact_stat_kwo

contfact
contfactdate
contfacttypestat
contfact_stat_lan_lab
contfact_stat_lan

contfact
contfactdate
contfacttypeinfo
contfact_info_obj_lab
contfact_info_obj

contfact
contfactdate
contfacttypeinfo
contfact_info_dcm_lab
contfact_info_dcm

contfact
contfactdate
contfacttypeinfo
contfact_info_mtx_lab
contfact_info_mtx

contfact
contfactdate
contfacttypeinfo
contfact_info_aut_lab
contfact_info_aut

contfactend



...
```

Figure 34. Sequence of modular high-level CONTFACT routines for `lxcontfact` implementation (LX Resources, excerpt).

In this case atomised modules are used to create entries. The module calls are grouped by their purpose for creating certain entries. In the example one single Content Factor with additional information is created. For example, after the `contfactbegin`, the `contfact`, `contfactdate` up to `contfactsum` create an entry with date / timestamp, type specification, specification of unsorted elements, reference specification (/), and sum. The next block adds a sorted entry to the Content Factor. The contfactdef calculates and adds the definitions used with the above entries. The following blocks add additional information and statistics, e.g., statistics on the number of elements or information on the referred object in the knowledge resources. This means any core entries, statistics and so on can be computed with individual implementations if required.

Application scenarios may allow to compute Content Factors for many objects in parallel. Content Factors can be computed dynamically as well as in batch mode or "pre-computed". Content Factors can be kept volatile as well as persistent. Everything can be considered a set, e.g., an object, a collection, and a container. Content Factors can be computed for arbitrary data, e.g., objects, collections, and containers. A consistent implementation delivers a Content Factor for a collection, which is the sum of the Content Factors computed for the objects contained in the collection. Therefore, an implementation can scale from single on the fly objects to millions of objects, which may also associated with pre-computed Content Factors.

B. Parallelisation and persistence

There is a number of modules supporting computation based on persistent data, e.g., in collections and containers. The architecture allows task parallel implementations for multiple instances as well as highly parallel implementations for core routines.

Applications are decollators, which extract objects from collection and containers and compute object based Content Factors. Other applications are slicers and atomisers, which cut data, e.g., objects, into slices or atoms, e.g., lines or strings, for which Content Factors can be computed. Examples in context with the above application scenarios are collection decollators, container decollators, collection slicers, container slicers, collection atomisers, container atomisers, formatting modules, computing modules for (intermediate) result matrix requests.

Content Factor data can easily be kept and handled on persistent as well as on dynamical base. The algorithms and workflows allow the flexible organisation of data locality, e.g., central locations and with compute units, e.g., in groups or containers.

IX. Evaluation

The presented application scenarios and according implementations have shown that many different cases targeting on knowledge processing can benefit from data description and analysis with the Content Factor method.

The case studies showed that the formal description can be implemented very flexibly and successful (`lxcontfact`). Content Factors can be computed for any type of data. The Content Factor is not limited to text processing or even NLP, term-frequencies, and statistics. It has been successfully used with long term knowledge resources and with unstructured and dynamical data. The Content Factor method can describe arbitrary data in a unique form and supports data analysis and knowledge discovery in many ways, e.g., complex data comparison and tracking of relevant changes.

Definition sets can support various use cases. Examples were given from handling single characters to string elements. Definitions can be kept with the Content Factor, together with additional Content Factor data, e.g., statistics and documentation. Any of this Content Factor information has been successfully used to analyse data objects from different sources. The computation of Content Factors is non invasive, the results can be created dynamically and persistent. Content Factors can be automatically computed for elements and groups of large data resources. The integration with data and knowledge resources can be kept non invasive to least invasive, depending on the desired purposes. Knowledge objects, e.g., in collections and containers, can carry and refer to complementary information and knowledge, especially Content Factor information, which can be integrated with workflows, e.g., for discovery processes.

The implementation is as far data-centric as possible. Data and technical implementations can be separated and the created knowledge resources and technical components comply to the above criteria.

The benefits and usability may depend on the field of application and the individual goals. The evaluation refers to the case context presented, which allows a wide range of freedom and flexibility. The benefits for the knowledge resources are additional means for documentation of objects. In detail, the benefits for the example workflows were improved data-mining pipelines, due to additional features for comparisons of objects, integrating developing knowledge resources, and creating and developing knowledge resources.

In practice, the computation of Content Factors has revealed significant benefits for the creation and analysis of large numbers of objects and for the flexibility and available features for building workflows, e.g., when based on long-term knowledge objects. In addition, creators, authors, and users of knowledge and content have additional means to express their views and valuation of objects and groups of objects. From the computational point of view, the computation of Content Factors can help minimise the recurrent computing demands for data.

X. Conclusion

This paper introduced a methodology for data description and analysis, the Content Factor (CONTFACT) method and presented the developments and results on knowledge processing algorithms and discovery for advanced application scenarios.

The paper presents the formal description and examples, a successful implementation, and a practical case study. It has been shown that the Content Factor is data-centric and can describe and analyse arbitrary data and content, structured and unstructured. Data-centricity is even emphasized due to the fact that the Content Factor can be seamlessly integrated with the data. The data locality is most flexible and allows an efficient use of different computing, storage, and communication architectures.

The method can be adopted for many purposes. The Content Factor method has been successfully applied for knowledge processing and analysis with long-term knowledge resources, for knowledge discovery, and with variable data for system operation analysis. It enables to specify a wide range of precision and fuzziness for data description and analysis and also enables methods like data rhythm analysis and characterisation, can be integrated with complementary methodologies, e.g., classifications, concordances, and references.

Therefore, the method allows weighting data regarding significance, promoting the value of data. The method supports the use of advanced computing methods for computation and analysis with the implementation. The computation and processing can be automated and used with huge and even unstructured data resources. The methodology allows an integrated use with complementary methodologies, e.g., with conceptual knowledge like UDC.

It will be interesting to see further various Content Factor implementations for individual applications, e.g., dynamical classification and concordances. Future work concentrates on high level applications and implementations for advanced analysis and automation.

Acknowledgements

We are grateful to the "Knowledge in Motion" (KiM) long-term project, Unabhängiges Deutsches Institut für Multidisziplinäre Forschung (DIMF), for partially funding this implementation, case study, and publication under grants D2014F1P04518 and D2014F2P04518 and to its senior scientific members, especially to Dr. Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover, to Dipl.-Biol. Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, and to Dipl.-Ing. Martin Hofmeister, Hannover, for fruitful discussion, inspiration, practical multi-disciplinary case studies, and the analysis of advanced concepts. We are grateful to Dipl.-Ing. Hans-Günther Müller, Cray, for his work on flexible practical solutions to architectural challenges and excellent technical support. We are grateful to all national and international partners in the Geo Exploration and Information cooperations for their constructive and trans-disciplinary support. We thank the Science and High Performance Supercomputing Centre (SHPSC) for long-term support of collaborative research since 1997, including the GEXI developments and case studies.

References

[1] C.-P. Rückemann, "Enhancement of Knowledge Resources and Discovery by Computation of Content Factors," in Proceedings of The Sixth International Conference on Advanced Communications and Computation (INFOCOMP 2016), May 22–26, 2016, Valencia, Spain. XPS Press, 2016, Rückemann, C.-P., Pankowska, M. (eds.), pages 24–31, ISSN: 2308-3484, ISBN-13: 978-1-61208-478-7, ISBN-13: 978-1-61208-061-1 (CDROM), TMDL: infocomp_2016_2_30_60047, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2016_2_30_60047 [accessed: 2016-06-18], URL: http://www.thinkmind.org/index.php?view=article&articleid=infocomp_2016_2_30_60047 [accessed: 2016-06-18].

[2] T. Koltay, "Data literacy for researchers and data librarians," Journal of Librarianship and Information Science, 2015, pp. 1–12, Preprint, DOI: 10.1177/0961000615616450.

[3] E. König, "From Information Specialist to Data Specialist, (German: Vom Informationsspezialisten zum Datenspezialisten)," library essentials, LE_Informationsdienst, March 2016, 2016, pp. 8–11, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-03-20].

[4] R. Uzwyshyn, "Research Data Repositories: The What, When, Why, and How," Computers in Libraries, vol. 36, no. 3, Apr. 2016, pp. 11–14, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-03-20].

[5] E. König, "Research Data Repositories - A new Field of Activities, (in German: Forschungsdaten-Repositorien als ein neues Betätigungsfeld)," library essentials, LE_Informationsdienst, Jun. 2016, 2016, pp. 11–14, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-03-20].

[6] C.-P. Rückemann, "Advanced Content Balancing and Valuation: The Content Factor (CONTFACT)," Knowledge in Motion Long-term Project, Unabhängiges Deutsches Institut für Multidisziplinäre Forschung (DIMF), Germany; Westfälische Wilhelms-Universität Münster, Münster, 2009, Project Technical Report.

[7] C.-P. Rückemann, "CONTCODE – A Code for Balancing Content," Knowledge in Motion Long-term Project, Unabhängiges Deutsches Institut für Multidisziplinäre Forschung (DIMF), Germany; Westfälische Wilhelms-Universität Münster, Münster, 2009, Project Technical Report.

[8] F. Hülsmann and C.-P. Rückemann, "Content and Factor in Practice: Revealing the Content-DNA," KiM Summit, October 26, 2015, Knowledge in Motion, Hannover, Germany, 2015, Project Meeting Report.

[9] C.-P. Rückemann, "Integrated Computational and Conceptual Solutions for Complex Environmental Information Management," in The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences, Proceedings of The 13th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 23–29, 2015, Rhodes, Greece, Proceedings of the American Institute of Physics (AIP), vol. 1738. AIP Press, 2016, Simos, T. E., Tsitouras, C. (eds.), ISBN-13: 978-0-7354-1392-4, ISSN: 0094-243X (American Institute of Physics Conference Proceedings, print), ISSN: 1551-7616 (online) (eISSN), DOI: 10.1063/1.4951833.

[10] D. T. Meridou, U. Inden, C.-P. Rückemann, C. Z. Patrikakis, D.-T. I. Kaklamani, and I. S. Venieris, "Ontology-based, Multi-agent Support of Production Management," in The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences, Proceedings of The 13th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 23–29, 2015, Rhodes, Greece, Proceedings of the American Institute of Physics (AIP), vol. 1738. AIP Press, 2016, Simos, T. E., Tsitouras, C. (eds.), ISBN-13: 978-0-7354-1392-4, ISSN: 0094-243X (American Institute of Physics Conference Proceedings, print), ISSN: 1551-7616 (online) (eISSN), DOI: 10.1063/1.4951834.

[11] C.-P. Rückemann, F. Hülsmann, B. Gersbeck-Schierholz, P. Skurowski, and M. Staniszewski, Knowledge and Computing. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing, September 23, 2015, The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences, The 13th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 23–29, 2015, Rhodes, Greece, 2015.

[12] O. Lipsky and E. Porat, "Approximated Pattern Matching with the L_1, L_2 and L_∞ Metrics," in 15th International Symposium on String Processing and Information Retrieval (SPIRE 2008), November 10–12, 2008, Melbourne, Australia, ser. Lecture Notes in Computer Science (LNCS), vol. 5280. Springer, Berlin, Heidelberg, 2008, pp. 212–223, Amir, A. and Turpin, A. and Moffat, A. (eds.), ISSN: 0302-9743, ISBN: 978-3-540-89096-6, LCCN: 2008938187.

[13] G. Ercan and I. Cicekli, "Lexical Cohesion Based Topic Modeling for Summarization," in Proceedings of The 9th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2008), February 17–23, 2008, Haifa, Israel, ser. Lecture Notes in Computer Science (LNCS), vol. 4919. Springer, Berlin, Heidelberg, 2008, pp. 582–592, Gelbukh, A. (ed.), ISSN: 0302-9743, ISBN: 978-3-540-78134-9, LCCN: 2008920439, URL: http://link.springer.com/chapter/10.1007/978-3-540-78135-6_50 [accessed: 2016-01-10].

[14] G. Szarvas, T. Zesch, and I. Gurevych, "Combining Heterogeneous Knowledge Resources," in Proceedings of The 12th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2011), February 20–26, 2011, Tokyo, Japan, ser. Lecture Notes in Computer Science (LNCS), vol. 6608 and 6609. Springer, Berlin, Heidelberg, 2011, pp. 289–303, Gelbukh, A. (ed.), ISSN: 0302-9743, ISBN: 978-3-642-19399-6, DOI: 10.1007/978-3-642-19400-9, LCCN: 2011921814, URL: http://link.springer.com/chapter/10.1007/978-3-642-19400-9_23 [accessed: 2016-01-10].

[15] A. Woodie, "Is 2016 the Beginning of the End for Big Data?" Datanami, 2016, January 5, 2016, URL: http://www.datanami.com/2016/01/05/is-2016-the-beginning-of-the-end-for-big-data/ [accessed: 2016-01-10].

[16] M. E. Jennex, "A Proposed Method for Assessing Knowledge Loss Risk with Departing Personnel," VINE: The Journal of Information and Knowledge Management Systems, vol. 44, no. 2, 2014, pp. 185–209, ISSN: 0305-5728.

[17] R. Leming, "Why is information the elephant asset? An answer to this question and a strategy for information asset management," Business Information Review, vol. 32, no. 4, 2015, pp. 212–219, ISSN: 0266-3821 (print), ISSN: 1741-6450 (online), DOI: 10.1177/0266382115616301.

[18] E. König, "The (Unknown) Value of Information (in German: Der (unbekannte) Wert von Information)," library essentials, LE_Informationsdienst, Dez. 2015 / Jan. 2016, 2015, pp. 10–14, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-08-27].

[19] E. König, "Effects of Open Access on Remote Loan and Other Information Resources, (in German: Die Auswirkungen von Open Access auf die Fernleihe und andere Informationsressourcen)," library essentials, LE_Informationsdienst, Jun. 2015, 2015, pp. 11–14, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-03-20].

[20] T. Baich, "Open access: help or hindrance to resource sharing?" Interlending & Document Supply, vol. 43, no. 2, 2015, pp. 68–75, DOI: 10.1108/ILDS-01-2015-0003.

[21] B. Kosko, "Counting with Fuzzy Sets," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 4, Jul. 1986, pp. 556–557, ISSN: 0162-8828, DOI: 10.1109/TPAMI.1986.4767822.

[22] "LX-Project," 2016, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX [accessed: 2016-06-18].

[23] C. W. Belter and N. K. Kaske, "Using Bibliometrics to Demonstrate the Value of Library Journal Collections," College & Research Libraries, vol. 77, no. 4, Jul. 2016, pp. 410–422, DOI: 10.5860/crl.77.4.410, URL: http://crl.acrl.org/content/77/4/410.full.pdf+html [accessed: 2016-08-20].

[24] E. König, "Demonstrate the Value of Libraries Using Bibliometric Analyses (German: Den Wert von Bibliotheken mittels bibliometrischer Analysen nachweisen)," library essentials, LE_Informationsdienst, Aug. 2016, 2016, pp. 11–16, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-08-20].

[25] E. König, "Most Stored Data is of no Economic Value (German: Die meisten gespeicherten Daten sind ohne wirtschaftlichen Wert)," library essentials, LE_Informationsdienst, Aug. 2016, 2016, pp. 30–32, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-08-20].

[26] "Data Genomics Index 2016," 2016, URL: http://datagenomicsproject.org/Data_Genomics_Index_2016.pdf [accessed: 2016-08-20].

[27] E. König, "More Figures = More Citations (German: Mehr Bilder = mehr Zitierungen)," library essentials, LE_Informationsdienst, Aug. 2016, 2016, pp. 34–35, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-08-20].

[28] P. Lee, J. D. West, and B. Howe, "Viziometrics: Analyzing Visual Information in the Scientific Literature," 2016, URL: http://arxiv.org/abs/1605.04951 [accessed: 2016-08-20].

[29] E. König, "Are Reference Services Still Needed in the Age of Google and Co.? (German: Werden Auskunftsdienste im Zeitalter von Google und Co. noch benötigt?)," library essentials, LE_Informationsdienst, Aug. 2016, 2016, pp. 11–14, ISSN: 2194-0126, URL: http://www.libess.de [accessed: 2016-08-20].

[30] S. P. Buss, "Do We Still Need Reference Services in the Age of Google and Wikipedia?" The Reference Librarian, vol. 57, no. 4, 2016, pp. 265–271, DOI: 10.1080/02763877.2015.1134377, URL: http://dx.doi.org/10.1080/02763877.2015.1134377 [accessed: 2016-08-20].

[31] C.-P. Rückemann, Z. Kovacheva, L. Schubert, I. Lishchuk, B. Gersbeck-Schierholz, and F. Hülsmann, Best Practice and Definitions of Data-centric and Big Data – Science, Society, Law, Industry, and Engineering. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data-centric and Big Data – Science, Society, Law, Industry, and Engineering, September 19, 2016, The Sixth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 14th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 19–25, 2016, Rhodes, Greece, 2016, URL: http://www.user.uni-hannover.de/cpr/x/publ/2016/delegatessummit2016/rueckemann_icnaam2016_summit_summary.pdf [accessed: 2016-11-06].

[32] C.-P. Rückemann, "Enabling Dynamical Use of Integrated Systems and Scientific Supercomputing Resources for Archaeological Information Systems," in Proceedings INFOCOMP 2012, Oct. 21–26, 2012, Venice, Italy, 2012, pp. 36–41, ISBN: 978-1-61208-226-4, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2012_3_10_10012 [accessed: 2016-08-28].

[33] “UDC Online,” 2015, URL: http://www.udc-hub.com/ [accessed: 2016-01-01].

[34] “Multilingual Universal Decimal Classification Summary,” 2012, UDC Consortium, 2012, Web resource, v. 1.1. The Hague: UDC Consortium (UDCC Publication No. 088), URL: http://www.udcc.org/udcsummary/php/index.php [accessed: 2016-01-01].

[35] “Creative Commons Attribution Share Alike 3.0 license,” 2012, URL: http://creativecommons.org/licenses/by-sa/3.0/ [accessed: 2016-01-01].

[36] B. F. S. Gersbeck-Schierholz, “Testimonies of Nature and History in the Napoli and Vesuvius Region, Italy,” Media Presentation, January 2014, Hannover, Germany, 2014, URL: http://www.user.uni-hannover.de/zzzzgers/bgs_volcano.html [accessed: 2016-08-27].

[37] F. Hülsmann, C.-P. Rückemann, M. Hofmeister, M. Lorenzen, O. Lau, and M. Tasche, “Application Scenarios for the Content Factor Method in Libraries, Natural Sciences and Archaeology, Statics, Architecture, Risk Coverage, Technology, and Material Sciences,” KiM Strategy Summit, March 17, 2016, Knowledge in Motion, Hannover, Germany, 2016.

[38] “The Perl Programming Language,” 2016, URL: https://www.perl.org/ [accessed: 2016-01-10].

[39] C.-P. Rückemann, “Creation of Objects and Concordances for Knowledge Processing and Advanced Computing,” in Proceedings of The Fifth International Conference on Advanced Communications and Computation (INFOCOMP 2015), June 21–26, 2015, Brussels, Belgium. XPS Press, 2015, pp. 91–98, ISSN: 2308-3484, ISBN-13: 978-1-61208-416-9, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2015_4_30_60038 [accessed: 2016-08-28].

[40] C.-P. Rückemann, “Advanced Association Processing and Computation Facilities for Geoscientific and Archaeological Knowledge Resources Components,” in Proceedings of The Eighth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2016), April 24–28, 2016, Venice, Italy. XPS Press, 2016, Rückemann, C.-P. and Doytsher, Y. (eds.), pages 69–75, ISSN: 2308-393X, ISBN-13: 978-1-61208-469-5, ISBN-13: 978-1-61208-060-4 (CDROM), TMDL: geoprocessing_2016_4_20_30144, URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2016_4_20_30144 [accessed: 2016-06-05], URL: http://www.thinkmind.org/index.php?view=article&articleid=geoprocessing_2016_4_10_30144 [accessed: 2016-06-05].

[41] “Mathematics Subject Classification (MSC2010),” 2010, URL: http://msc2010.org [accessed: 2015-02-01].

[42] Fundamentals of Library of Congress Classification, Developed by the ALCTS/CCS-PCC Task Force on Library of Congress Classification Training, 2007, Robare, L., Arakawa, S., Frank, P., and Trumble, B. (eds.), ISBN: 0-8444-1186-8 (Instructor Manual), ISBN: 0-8444-1191-4 (Trainee Manual), URL: http://www.loc.gov/catworkshop/courses/fundamentalslcc/pdf/classify-trnee-manual.pdf [accessed: 2015-02-01].

[43] “Physics and Astronomy Classification Scheme, PACS 2010 Regular Edition,” 2010, American Institute of Physics (AIP), URL: http://www.aip.org/pacs [accessed: 2015-02-01].

[44] IBM, ““Data-Centric” approach feeds cognitive computing; How to deploy a data-centric methodology for your organization,” 2016, ISSN: 2194-0126, URL: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=DCG12426USEN [accessed: 2016-06-25].