INFOCOMP 2020

The Tenth International Conference on Advanced Communications and Computation

ISBN: 978-1-61208-807-5

September 27th – October 1st, 2020

INFOCOMP 2020 Editors

Claus-Peter Rückemann, Westfälische Wilhelms-Universität Münster (WWU); DIMF, Germany, LUH, Germany

INFOCOMP 2020

Foreword

The Tenth International Conference on Advanced Communications and Computation (INFOCOMP 2020), held between September 27 – October 1st, 2020, continued a series of events dedicated to advanced communications and computing aspects, covering academic and industrial achievements and visions.

The diversity of semantics of data, context gathering and processing led to complex mechanisms for applications requiring special communication and computation support in terms of volume of data, processing speed, context variety, etc. The new computation paradigms and communications technologies are now driven by the needs for fast processing and requirements from data-intensive applications and domain-oriented applications (medicine, geo-informatics, climatology, remote learning, education, large scale digital libraries, social networks, etc.). Mobility, ubiquity, multicast, multi-access networks, data centers, cloud computing are now forming the spectrum of de factor approaches in response to the diversity of user demands and applications. In parallel, measurements control and management (self-management) of such environments evolved to deal with new complex situations.

We take here the opportunity to warmly thank all the members of the INFOCOMP 2020 Technical Program Committee, as well as the numerous reviewers. The creation of this event would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to INFOCOMP 2020.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the INFOCOMP 2020 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that INFOCOMP 2020 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the areas of communications and computations.

.

INFOCOMP 2020 Chairs:

INFOCOMP 2020 Publicity Chair
Mar Parra, Universitat Politecnica de Valencia, Spain

INFOCOMP 2020
COMMITTEE

INFOCOMP 2020 Publicity Chair

Mar Parra, Universitat Politecnica de Valencia, Spain

INFOCOMP 2020 Technical Program Committee

Vicki H. Allan, Utah State University, USA
Ghada Almashaqbeh, NuCypher, USA
Daniel Andresen, Kansas State University, USA
Marc Baaden, CNRS, France
Raymond Bair, Argonne National Laboratory / University of Chicago, USA
Jacob Balma, Hewlett Packard Enterprise Company, USA
Bernhard Bandow, GWDG - Göttingen, Germany
Christine Bassem, Wellesley College, USA
Raoudha Ben Djemaa, MIRACL, Sfax, Tunisia
Abbas Bradai, University of Poitiers, France
Stephanie Brink, Lawrence Livermore National Laboratory, USA
Hans-Joachim Bungartz, Technische Universität München (TUM) - Garching, Germany
Paolo Burgio, University of Modena and Reggio Emilia, Italy
Xiao-Chuan Cai, University of Colorado Boulder, USA
Nicola Calabretta, Eindhoven University of Technology, Netherlands
Enrico Casella, University of Kentucky, USA
Jian Chang, Bournemouth University, UK
Jieyang Chen, Oak Ridge National Laboratory, USA
Albert M. K. Cheng, University of Houston, USA
Enrique Chirivella Pérez, Universitat de Valencia, Spain
Noelia Correia, Center for Electronics Opto-Electronics and Telecommunications (CEOT) | University of Algarve, Portugal
Kei Davis, Los Alamos National Laboratory, USA
Tiziano De Matteis, ETH Zurich, Switzerland
Daniele De Sensi, ETH Zurich, Switzerland
Iman Faraji, Nvidia Inc., Canada
Josué Feliu, Universitat Politècnica de València, Spain
Francesco Fraternali, University of California, San Diego, USA
Hans-Hermann Frese, Gesellschaft für Informatik e.V., Germany
Steffen Frey, Visualization Research Center - University of Stuttgart, Germany
Marco Furini, University of Modena and Reggio Emilia, Italy
Jason Ge, Snark AI Inc, USA
Alfred Geiger, T-Systems Solutions for Research GmbH, Germany
Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany
Franca Giannini, IMATI-CNR, Italy
Barbara Guidi, University of Pisa, Italy
Önder Gürcan, CEA LIST, France
Enrique Hernández Orallo, Universidad Politécnica de Valencia, Spain
Gonzalo Hernandez, CCTVal - USM & STII, Chile

Mert Hidayetoglu, University of Illinois at Urbana-Champaign, USA
Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek, Hannover, Germany
Thomas Hupperich, University of Münster, Germany
Mohamed Assem Ibrahim, William & Mary, USA
Sergio Ilarri, University of Zaragoza, Spain
Eugene B. John, The University of Texas at San Antonio, USA
Ram Srivatsa Kannan, Uber Technologies Inc., USA
Izabela Karsznia, University of Warsaw, Poland
Alexander Kipp, Robert Bosch GmbH, Germany
Felix Klapper, Leibniz Universität Hannover, Germany
Zlatinka Kovacheva, Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, Sofia, Bulgaria
Manfred Krafczyk, Institute for Computational Modeling in Civil Engineering (iRMB) - TU Braunschweig, Germany
Nane Kratzke, Lübeck University of Applied Sciences, Germany
Sonal Kumari, Samsung Research India-Bangalore (SRI-B), India
Julian M. Kunkel, University of Reading, UK
Stephen Leak, NERSC User Engagement, USA
Yiu-Wing Leung, Hong Kong Baptist University, Kowloon Tong, Hong Kong
Guanpeng Li, University of Illinois Urbana Champaign, USA
Shigang Li, ETH Zurich, Switzerland
Yanting Li, City University of Hong Kong, Hong Kong
Jinwei Liu, Florida A&M University, USA
Hui Lu, SUNY Binghamton, USA
Sandeep Madireddy, Argonne National Laboratory, USA
Adnan Mahmood, Macquarie University, Australia / Telecommunications Software & Systems Group, WIT, Republic of Ireland
Antonio Martí-Campoy, Universitat Politècnica de València, Spain
Artis Mednis, Institute of Electronics and Computer Science, Latvia
Roderick Melnik, MS2Discovery Interdisciplinary Research Institute | Wilfrid Laurier University (WLU), Canada
Mariofanna Milanova, University of ArkansasLittle Rock, USA
Behzad Mirkhanzadeh, University of Texas at Dallas, USA
Victor Mitrana, Polytechnic University of Madrid, Spain
Sébastien Monnet, Savoie Mont Blanc University (USMB), France
Hans-Günther Müller, HPE, Germany
Duc Manh Nguyen, University of Ulsan, Korea
Alex Norta, Tallinn University (TLU), Estonia
Krzysztof Okarma, West Pomeranian University of Technology in Szczecin, Poland
Giuseppe Patane’, CNR-IMATI, Genova, Italy
Han Qiu, Telecom Paris, Paris, France
Francesco Quaglia, Università di Roma "Tor Vergata", Italy
Danda B. Rawat, Howard University, USA
Carlos Reaño, Queen's University Belfast, UK
Ustijana Rechkoska-Shikoska, University for Information Science and Technology "St. Paul the Apostle" - Ohrid, Republic of Macedonia
Yenumula B Reddy, Grambling State University, USA
Theresa-Marie Rhyne, Visualization Consultant, Durham, USA

Copyright Information

Table of Contents

The Impact of Information Science Accompanied Structural Information on Computation of Knowledge Pattern Matching and Processing: A Prehistory, Archaeology, Natural Sciences, and Humanities Conceptual Integration Perspective 1
Claus-Peter Ruckemann

The Impact of Information Science Accompanied Structural Information on Computation of Knowledge Pattern Matching and Processing

A Prehistory, Archaeology, Natural Sciences, and Humanities Conceptual Integration Perspective

Claus-Peter Rückemann
Westfälische Wilhelms-Universität Münster (WWU), Germany;
Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany;
Leibniz Universität Hannover, Germany
Email: ruckema@uni-muenster.de

Abstract—**This paper delivers new results from the long-term information science research on creation and deployment of structure and structural information for problem solving. The paper presents the new methodological base and methodology focussing on structure-based fusion solutions and implementations for computational scenarios, especially advanced knowledge-centric mining. Information science interrelations and knowledge complements used for pattern matching and processing are fundaments for mastering these challenges. Facilities for combination of complementary and descriptive knowledge of meaning and intrinsic object properties are essential means. The research is illustrated by advanced practical implementations of knowledge pattern matching, including challenges of computation and processing of multi-disciplinary and multi-lingual knowledge object entities and resources. The goal of this fundamental research is to create structure-based methods for efficient problem solving. Case studies implement the method to consistently integrate knowledge context from prehistory and archaeology disciplines with knowledge in natural sciences and humanities. Previously unpublished insights are available with this publication.**

Keywords–Structure-based methods; Information Science; Computation and Processing; Knowledge Complements; UDC.

I. INTRODUCTION

In the last decades, it has become common practice to tackle challenges regarding knowledge and related content solely with procedural approaches, besides the fact that creation processes, handling, and management may allow more effective and efficient measures in context of computation, processing, analysis, and long-term development or resources. Common ways of implementing procedural approaches as plain technical solutions are often neither effective nor efficient. In addition, such approaches often lack long-term adaptability and scalability.

Procedure-based approaches are largely not addressing the knowledge and information content. Structure-based approaches can fill the gap. This motivation is supported by the experience that inefficiencies of procedural approaches regarding their creation, development, and execution can often be avoided by focussing on structure. Compared to procedure-based approaches, structure-based approaches are in general substantially different. Structure and formatting contain valuable information and closely correspond with logic, which should not be lost in many cases, e.g., this is especially the case for any sustainable long-term knowledge. That means, it is not intended to convert structures or to change formatting of resources. For information science and a universal knowledge context, meaning knowledge complements including conceptual knowledge, we also require a consistent, advanced definition of structure. It is important to understand what structure and form mean in theory and practice. The fundaments will be delivered before practical implementations are discussed.

The presented methodology addresses the shortcomings of common procedure-based approaches. The new structure-based fusion methodology and method implementations presented here are created as general purpose implementations, which can be realised in any Turing complete machine programming language, supporting creation and development of resources as well as computation, processing, and analysis. For the illustrative case studies, these approaches are used for dealing with various aspects of knowledge management and knowledge mining in information science context.

Knowledge resources' structures are commonly set and have proven long-term flexibility and sustainability. They cover content, context, consistency, and sustainability features for millions of information content, references, and object entities for long periods of time. Therefore, solutions for procedural components have to be found, which do provide comparably defiant long-term flexibility and sustainability. In addition, the procedural components require a very high level of knowledge-centricity and may need to exploit resource features, which are commonly not paid attention. When working with long-term resources, structural information of resources and entities and their organisation has shown to provide a high potential. Structural information also promises to achieve a high level of integration. Therefore, it might seem obvious to consider structure-based approaches for advanced and challenging tasks.

The rest of this paper is organised as follows. Section II gives an overview of previous work and deployed components. Section III introduces to structure and addressing. Section IV presents the methodology and resulting method and realisation. Section V presents knowledge ranges and a computational footprint. Section VI summarises conclusions and future work.

II. PREVIOUS WORK, COMPONENTS, AND RESOURCES

The fundaments of terminology and understanding knowledge are layed out by Aristotle being an essential part of 'Ethics' [1]. Information sciences can very much benefit from Aristotle's fundaments and a knowledge-centric approach [2] but for building holistic and sustainable solutions, supporting a modern definition of knowledge [3], they need to go beyond the available technology-based approaches and hypothesis [4] as analysed in Platon's Phaidon.

For the implementation of case studies, the modules are built by support of a number of major components and resources, which can be used for a wide range of applications, e.g., creation of resources and extraction of entities. Here, it is essential to regard the most important fundaments of structural information of language and respective content.

The concept of meaning differs from the concept of signification. Semantic and syntactic structures do not suffice to determine the discursive meaning of an expression [5]. Discourse means a way of speaking. On the one hand, grammatically correct phrases may lack discursive meaning. On the other hand, grammatically incorrect sentences may be discursively meaningful. Knowledge and meaning are closely tied with intrinsic and extrinsic properties. Therefore, understanding of intrinsic and extrinsic properties of entities is significant for any context. This is nevertheless true for any case of natural language, especially considering langage, langue, and parole [6], especially when interpretation [7] and meaning [8] should be considered, especially regarding cognition and insight [9].

The Universal Decimal Classification (UDC) [10] is the world's foremost document indexing language in the form of a multi-lingual classification scheme covering all fields of knowledge and constitutes a sophisticated indexing and retrieval tool. The UDC is designed for subject description and indexing of content of information resources irrespective of the carrier, form, format, and language. UDC is an analytico-synthetic and facetted classification. UDC schedules are organised as a coherent system of knowledge with associative relationships and references between concepts and related fields. UDC-based references in this publication are taken from the multi-lingual UDC summary [10] released by the UDC Consortium under a Creative Commons license [11]. Facets can be created with any auxiliary tables.

A means of choice to achieve overall efficient realisations even for complex scenarios is to use the principles of Superordinate Knowledge, integrating arbitrary knowledge. The core assembly elements of Superordinate Knowledge are methodology, implementation, and realisation [12]. Comprehensive focussed subsets of conceptual knowledge can also provide excellent modular and standardised complements for information systems component implementations, e.g., for environmental information management and computation [13]. The presented implementations strictly follow the fundamental methodological algorithm base of the the Conceptual Knowledge Pattern Matching (CKPM) methodology [14] providing and accessing knowledge object patterns based on the Superordinate Knowledge Methodology, which allows systematical use and thorough processing. Core eager beaver procedure- and structure-based implementation components, grep and join, are written in C, as commonly known. Module examples are employing Perl Compatible Regular Expressions (PCRE) [15] syntax for specifying common string patterns. This is independent from the procedural realisation using Shell and Perl [16] for component wrapping purposes with case studies.

III. STRUCTURE AND INFORMATION

Structure is an organisation of interrelated entities in a material or non-material object or system. Structure is essential in logic as it carries unique information. The more, we have to recognise the differences of structure and form. The case of text is a good example. The structure of a text consists of the particular text units and their context, in order to make the text coherent. The form of a text is the arrangement of the text units, which commonly has to follow predefined rules.

A. Structure systematics, meaning, levels

As meaning can be understood in context of langage, langue, and parole, the available rules of structure and form should be used. If whatever non linguistic, artistic expression is primary target then different structure and form could be used. Anyway any linguistic parole context should be aware of the specific conditions: Academic use should be aware of the specific academic context. Commercial use should be aware of the specific commercial context. Marketing use should be aware of the specific marketing context. There are rarely reasonable compromises fitting diametrical approaches to form equally well. Consequences of these fundaments are, especially: Structure is not dependent on physical, non physical, analogous, digital or comparable being and properties. Structure is not dependent on same/uniform ways of structuring. Structure is not intrinsic to a certain scale of information. Structure should not imagined to be dependent of a location or dependent of management. Instead, it is more likely to yield consistent results when we follow a methodology regarding the systematics of structure. In information science and context of knowledge resources we can consider three major levels:

- Object entity structure (entities as part of an object).
- Object structure.
- Supra-object structure (e.g., complexity or inconsistency introduced by application or service scenario).

The methodological approach is beneficial when expressing disciplinary views and targeting purposes as it can

- help to create a consistent common understanding of structure,
- address responsibility,
- help to assure logical and consistent development and management of structure, etc.

B. Structure and means of addressing

There are merely higher and lower facility levels of how structures can be addressed, which result from structure levels. For example, structure can be addressed by:

- Logic.
- Names.
- References.
- Address labels.
- Pointers.
- Fuzzy methods.
- Phonetic methods. ...

For example, 'non-structures' can be addressed by:

- Locality.
- Source.
- Context.
- Logic.
- Attributes.
- Size.
- Quantity. ...

Substantial differences of properties and facilities of different levels of structure and non-structure do have multi-fold origin and reason, especially:

- Structure is associated with different formalisation levels and respective consequences.
- Less complementary knowledge realisation, less potential, e.g., for logic.
- Less structure, less potential for approaches.
- Intrinsic and extrinsic properties are not interchangeable.
- Higher levels of structure mostly include tools usable for lower levels.

- Low level structures are limited to low level tools and soft criteria, e.g., statistics and heuristics.
- Potential from quality is different than from quantity.

With that background we should be aware that lower structure levels can only be addressed on higher formalisation levels, independent of the fact that structure may either be not available or not recognised. Substantial deficits of lower level structured data cannot be compensated by tools. In consequence, structure is and especially reflects knowledge (complements of factual, conceptual, procedural, metacognitive, ...), context, experience, persistence, reusability, sustainability, value, and formalisation (including abstraction and reduction).

In result, it is structure that means features and facilities.

IV. Methodology and Implementation

The implementation strictly follows the fundamental methodological algorithm base, summarised in the following passages.

A. Methodological algorithm base and resulting method

The structure-based fusion methodology targets on supporting efficient problem solving, providing and accessing knowledge and knowledge object patterns. The methodology for creating structure-based fusion methods can be summarised on high level by:

- Pre-processing of structures.
- Option routines, for optional steps preparing fusion.
- Structure-based fusion of knowledge complements.
- Post-processing of structures.

This methodology is contrasting to plain procedure-based approaches. Here, structures are adapted for solving problems mostly allowing minimising procedure-based efforts and gaining higher efficiency and performance regarding creation and realisation of solutions. A method based on the structure-based fusion, implemented for knowledge mining context does consist of the basic steps:

- Pre-processing of knowledge mining structures.
- Options' routine, e.g., used for prioritisation and sorting. May contain procedures outside the range of balanced pre- and post-processing.
- Structure-based fusion of object structures, objects represented by knowledge complements.
- Post-processing of fusion result structures.

Besides the strictly structure-based fundament for the steps, realisations can use the whole gamut working for knowledge complements, e.g., comparisons, generic and standard component implementations, and knowledge complements and identification. Table I shows the method implementation and realisation of a structure-based fusion.

TABLE I. METHOD IMPLEMENTATION AND REALISATION: STEPS OF STRUCTURE-BASED FUSION.

Method Implementation Steps	*Realisation Example*
Outer pre-processing	individual, out of scope here
Input	**standard input, echo**
Inner pre-processing of structures	**perl**
Options' routines, prioritisation & sorting	**perl & sort**
Structure-based fusion	**join**
Inner post-processing of structures	**perl**
Output	**standard output**
Outer post-processing	individual, out of scope here

Any input and output can be intermediate, part of a complex mining process. Pre- and post-processing are handling the input and output for the options' routines and consecutive fusion, the central steps. For this implementation and demonstration processing is done via Perl, options' routines via Perl and sort, and the fusion via join. The illustrations of the realisations use inline GNU Bourne-Again SHell syntax, I/O and naming features for ease of demonstration.

B. Structure-based realisation: Simple case result

An instructive, simple case example implementation is a comparison and filter process of groups of arbitrary numbers of objects and object entities, which each can be of arbitrary volume and length. A procedural approach would create a procedure handling the structure and form of the entries as they appear and create and call a grep function for each of all the target entries or patterns. In advanced knowledge mining and processing, we often have to deal with sequences of steps creating intermediate results from previous results, all of which may need to be compared, sorted, filtered and so on. In that context the following scenario works as a basic example.

1) At a certain stage in a mining process we may have two groups of different knowledge object line entities.
2) We have to find only those various different string entries contained in one group and list those of the entries, which are also contained in the other group and produce combined object entities containing the content of respective entries from both groups.
3) We have to create an appropriate method and realisation, which, ideally, works for arbitrary numbers of objects with different sizes and content and which is flexible and knowledge-centric.

So, how can such 'search, comparison, filter, and sort' be realised for large numbers of objects, avoiding to call a routine or thread thousands or hundreds of thousand times per intermediate step and deploying structural information instead? Figure 1 shows a structure-based solution. Its realisation is a self-contained regular shell script containing object groups and solution for ease of reconstructing the train of thought.

```
#
# Structure-based fusion sample -- (c) CPR, 2019, 2020
#
cnta="Natural Sciences collection entry 10:05:34 Volcano
Natural Sciences collection entry entities 10:05:35
 Soufriere
Media attachment entry 10:06:34 Soufriere Photo
Addendum entry 10:05:30 References
Object entry compendium 10:06:37 comments"
cntb="Object entry 10:05:35 delivered
Excavation slide 10:05:34 updated
Object documentation update 10:06:34 request
System service no date
Object entity mining request 10:06:37 researcher id
 DF98_007
Object collection status 10:05:28 no resources reference
Object entity documentation request 10:05:30 user id
 database"
export cnta
export cntb
join -1 1 -2 1 <(echo "$cnta"|perl -pe 's/^(.*?)
 ([0-9][0-9]:[0-9][0-9]:[0-9][0-9])(.*?)$/$2 1_BEFORE{$1}
  1_AFTER{$3}/'|sort) <(echo "$cntb"|perl -pe 's/^(.*?)
 ([0-9][0-9]:[0-9][0-9]:[0-9][0-9])(.*?)$/$2 2_BEFORE{$1}
  2_AFTER{$3}/'|sort)
exit
```

Figure 1. Structure-based fusion solution, working with arbitrary one-lined object entities (excerpt), overcoming thousands of grep calls.

 ISBN: 978-1-61208-807-5

Here, two content groups of single line object instances are used for demonstration, content “a” and content “b”. As can be seen, the content groups are asymmetric regarding object instances, content, and context aspects. For convenience of demonstration an excerpt of the contents is embedded in the code and represented by the exported variables named `cnta` and `cntb`. This excerpt is doing a selection of objects by fusion of arbitrary length and arbitrary number of objects by criteria (time stamps), which are reflected by structure.

The solution for that purpose achieves that result explicitly without the use of ‘grep’ (Global Regular Expression Parser), ‘search’, or comparable procedural routine instances. In order to create a straightforward solution and to easily follow the strategy, the steps are implemented using 5 external calls. Figure 2 shows the result of the realisation (Figure 1).

```
10:05:30 1_BEFORE{Addendum entry} 1_AFTER{ References} 2_BEFORE{Object entity
 documentation request} 2_AFTER{ user id database}
10:05:34 1_BEFORE{Natural Sciences collection entry} 1_AFTER{ Volcano} 2_BEFORE{
 Excavation slide} 2_AFTER{ updated}
10:05:35 1_BEFORE{Natural Sciences collection entry entities} 1_AFTER{ Soufriere
 } 2_BEFORE{Object entry} 2_AFTER{ delivered}
10:06:34 1_BEFORE{Media attachment entry} 1_AFTER{ Soufriere Photo} 2_BEFORE{
 Object documentation update} 2_AFTER{ request}
10:06:37 1_BEFORE{Object entry compendium} 1_AFTER{ comments} 2_BEFORE{Object
 entity mining request} 2_AFTER{ researcher id DF98_007}
```

Figure 2. Result of the structure-based approach, working with arbitrary one-lined object entities (excerpt).

The result reflects the target task based on structure-based fusion. The solution includes a sort of resulting object entities by respective string entries and appropriate marking of content from the groups for illustration.

C. Structure-based realisation: Multi-line case result

A different kind of complexity is what we commonly face in context of knowledge resources, same task and still with arbitrary length and arbitrary number of objects and entities with multi-line formatting (Figure 3) to be preserved.

```
Nisyros                 [Volcanology, Geology]:
                        Volcano, Type: Strato volcano, Island.
                        Status: Historical, Summit Elevation: 698\UD{m}. ...
                        VNUM: 0102-05=, ...,
                        Craters: ..., ...
                        %%IML: UDC:[550.3],[930.85],[911.2]
                        %%IML: media:...{UDC:[550.3+551.21],[911.2](4+38+23)}...jpg
                        Stefanos Crater, Nisyros, Greece.
                        LATLON: 36.578345,27.1680696
                        %%IML: GoogleMapsLocation: https://www.google.com/...@36
                         .578345,27.1680696,337m/...
                        Little Polyvotis Crater, Nisyros, Greece.
                        LATLON: 36.5834105,27.1660736 ...
```

Figure 3. Knowledge resources’ object (‘Nisyros’): Multi-line formatting, conceptual knowledge, media object entities, and georeferences (excerpt).

Focus task is to find only those arbitrary object instances, which appear in one content context and also in another content context and to combine the data of those instances in a result object instance. It is preferable if the realisation allows a multi-object fusion, meaning more than one object in a process. A common procedure realisation would, e.g., have to call a ‘grep’ function (especially a Global Regular Expression Parser) for every of the thousands of object instances in one context for searches in another context

Figure 4 shows a more efficient, structure-based realisation for such objects. As with the previous example realisation above, its realisation is presented as a self-contained regular shell script for ease of demonstration. The excerpt fully confirms with a standard shell and Perl syntax and features and is compact. In order to create a straightforward solution and to easily follow the strategy, the steps are implemented using 10 external calls, which could still be further reduced. As can be seen, these calls already include formatting cleanup with pre- and post-processing, too. The solution targets contexts for larger numbers of multi-line, multi-entity object instances (thousands or hundreds of thousands). As common, results should be considered intermediate for complex knowledge mining procedures.

```
#
# Structure-based fusion mining -- (c) CPR, 2019, 2020
#
cnta="Nirgal                  [Etymology]:
                        ...
Nisyros                 [Archaeology, Geology, Volcanology]:
                        Island, Volcano, Greece, Dodecanese Islands. ...
                        The island provides unique archaeological remains, esp. ...
                        History and mythology of the island and volcano are ...
                        %%IML: UDC:[902],[930.85],[911.2]"63"(4+38+23+24)=14
Nisyros_archive         [Archaeology]:
                        Media, Archaeology Digital Object Archive (ADOA). ...
NMR                     [Archaeology]:
                        ..."
cntb="Niggli                  [Petrography, Mineralogy]:
                        ...
Nisyros                 [Volcanology, Geology]:
                        Volcano, Type: Strato volcano, Island, Greece.
                        Status: Historical, Summit Elevation: 698\UD{m}. ...
                        VNUM: 0102-05=, ..., Craters: ..., ...
                        %%IML: UDC:[550.3],[930.85],[911.2]
                        %%IML: media:...{UDC:[550.3+551.21],[911.2](4+38+23)}...jpg
                        Stefanos Crater, Nisyros, Greece.
                        LATLON: 36.578345,27.1680696
                        %%IML: GoogleMapsLocation: https://www.google.com/...@36
 .578345,27.1680696,337m/...
                        Little Polyvotis Crater, Nisyros, Greece.
                        LATLON: 36.5834105,27.1660736 ...
Nisyros_archive         [Volcanology]:
                        Media, Geosciences Digital Object Archive (GDOA). ...
NLBR                    [Platetectonics, Volcanology]:
                        ...
N-MORB                  [Platetectonics, Volcanology]:
                        ..."
export cnta
export cntb
join -1 1 -2 1 -t" " \
  <(echo "$cnta"|\
    perl -pe 's/^(\S)/TMPBOL$1/;s/$/TMPEOL/;s/\n//;'|perl -pe 's/TMPBOL//;s/
 TMPBOL/\n/g'|\
    perl -pe 's/^(.*?)(\[.*?\]:)(.*)$/$1$2$3/'|sort -k 1b,1)\
  <(echo "$cntb"| \
    perl -pe 's/^(\S)/TMPBOL$1/;s/$/TMPEOL/;s/\n//;'|perl -pe 's/TMPBOL//;s/
 TMPBOL/\n/g'|\
    perl -pe 's/^(.*?)(\[.*?\]:)(.*)$/$1$2$3/'|sort -k 1b,1)|\
  perl -pe 's/TMPEOL$//g;s/TMPEOL/\n/g;'
exit
```

Figure 4. Structure-based fusion solution, working with arbitrary multi-line knowledge resources’ object entities (excerpt), overcoming grep.

With the above scenario the realisation should be fully logical and self explanatory. The easiest way to comprehend the implementation principle is to follow the sequence of instructions and reproduce step by step, directly to the result. The realised solution should be reasonably flexible and robust. Figure 5 shows the result of the realisation (Figure 4).

```
Nisyros                 [Archaeology, Geology, Volcanology]:
                        Island, Volcano, Greece, Dodecanese Islands. ...
                        The island provides unique archaeological remains, esp. ...
                        History and mythology of the island and volcano are ...
                        %%IML: UDC:[902],[930.85],[911.2]63(4+38+23+24)=14
              [Volcanology, Geology]:
                        Volcano, Type: Strato volcano, Island, Greece.
                        Status: Historical, Summit Elevation: 698\UD{m}. ...
                        VNUM: 0102-05=, ..., Craters: ..., ...
                        %%IML: UDC:[550.3],[930.85],[911.2]
                        %%IML: media:...{UDC:[550.3+551.21],[911.2](4+38+23)}...jpg
                        Stefanos Crater, Nisyros, Greece.
                        LATLON: 36.578345,27.1680696
                        %%IML: GoogleMapsLocation: https://www.google.com/...@36
                         .578345,27.1680696,337m/...
                        Little Polyvotis Crater, Nisyros, Greece.
                        LATLON: 36.5834105,27.1660736 ...
Nisyros_archive         [Archaeology]:
                        Media, Archaeology Digital Object Archive (ADOA). ...
      [Volcanology]:
                        Media, Geosciences Digital Object Archive (GDOA). ...
```

Figure 5. Result of the structure-based approach, working with arbitrary multi-line object entities, result shows the multi-object fusion (excerpt).

Objects (`Nisyros` and `Nisyros_archive`) have correctly been identified (criteria name string) and unified. Entities have been preserved and conceptual knowledge of instances have been combined in a unique object instance each.

Even the indentation of the resulting content reflects the operations and is exactly preserved. Besides valuable knowledge content, conceptual knowledge integration is considered a marker regarding the range of knowledge and efficiency of a solution when working with knowledge resources. In this case the knowledge resources use UDC references with millions of object instances and entities. Universally consistent conceptual knowledge is based on UDC references for demonstration, spanning the main tables [17] shown in Table II.

TABLE II. CONCEPTUAL KNOWLEDGE PATTERN MATCHING: IMPL. UDC REFERENCES, MAIN TABLE.

Code / Sign Ref.	*Verbal Description (EN)*
UDC:0	Science and Knowledge. Organization. Computer Science. Information. Documentation. Librarianship. Institutions. Publications
UDC:1	Philosophy. Psychology
UDC:2	Religion. Theology
UDC:3	Social Sciences
UDC:5	Mathematics. Natural Sciences
UDC:6	Applied Sciences. Medicine, Technology
UDC:7	The Arts. Entertainment. Sport
UDC:8	Linguistics. Literature
UDC:9	Geography. Biography. History

Natural sciences related conceptual knowledge pattern entities are created based on UDC references [18] of mathematics and natural sciences. An excerpt of the implementation is shown in Table III.

TABLE III. CONCEPTUAL KNOWLEDGE PATTERN MATCHING: IMPL. UDC REFERENCES OF MATHEMATICS AND NATURAL SCIENCES (EXCERPT).

Code / Sign Ref.	*Verbal Description (EN)*
UDC:51	Mathematics
UDC:52	Astronomy. Astrophysics. Space research. Geodesy
UDC:53	Physics
UDC:54	Chemistry. Crystallography. Mineralogy
UDC:55	Earth Sciences. Geological sciences
UDC:550.3	Geophysics
UDC:551	General geology. Meteorology. Climatology.
UDC:551.21	Vulcanicity. Vulcanism. Volcanoes. Eruptive phenomena. Eruptions
UDC:551.24	Geotectonics
UDC:56	Palaeontology
UDC:57	Biological sciences in general
UDC:58	Botany
UDC:59	Zoology

Conceptual knowledge pattern entities of archaeology and history are created based on UDC references [19] of geography, biography, history (Table IV).

TABLE IV. CONCEPTUAL KNOWLEDGE PATTERN MATCHING: IMPL. UDC REFERENCES OF GEOGRAPHY, BIOGRAPHY, HISTORY (EXCERPT).

Code / Sign Ref.	*Verbal Description (EN)*
UDC:902	Archaeology
UDC:903	Prehistory. Prehistoric remains, artefacts, antiquities
UDC:904	Cultural remains of historical times
UDC:908	Area studies. Study of a locality
UDC:91	Geography. Exploration of the Earth and of individual countries. Travel. Regional geography
UDC:912	Nonliterary, nontextual representations of a region
UDC:92	Biographical studies. Genealogy. Heraldry. Flags
UDC:93/94	History
UDC:94	General history

The geoscientific / prehistory / archaeology integration from the case studies and implementations for geoscientific information systems and application components is used for illustration in the next sections. The example will show a tiny subset of the comprehensive, universal conceptual knowledge used, integrating UDC:902/908 (Archaeology. Prehistory. Cultural remains. Area studies) and UDC:55 (Earth Sciences. Geological sciences) and humanities (UDC main table trees).

V. MULTI-LINE KNOWLEDGE RANGES AND COMPUTATION

As commonly we have to handle many objects, we can illustrate how efficiency and performance scale with numbers of objects. The examples use the above multi-line knowledge case, as knowledge resources' objects regulary have a high variety of content, with different object volumes and lengths. Therefore, this is more for practical experience than a benchmark. The overall number of object instances in the respective primary knowledge ranges for the resources' excerpt is shown in the UDC references' test environment (Table V).

TABLE V. PRIMARY KNOWLEDGE RANGES OF CONCEPTUAL KNOWLEDGE ENTITY REFERENCES IN THE TEST ENVIRONMENT (EXCERPT).

Knowledge Range	*Entities' Count*
UDC:9 (incl. UDC:902/904)	930,000
UDC:5 (incl. UDC:55/56)	1,700,000

The ranges can be comprehended in all details by following the publicly available online conceptual knowledge framework already discussed above. Table VI shows an implementation excerpt and computational footprint for the different approaches. The different case results were achieved on Intel® Xeon® CPU X5570 (2.933 GHz) systems under Linux.

TABLE VI. COMPUTATIONAL FOOTPRINT OF PROCEDURE-/STRUCTURE-BASED SOLUTIONS, CONCEPTUAL KNOWLEDGE REF. CASES (EXCERPT).

Knowledge Range	*Entities' Count*	*Context Calls' Count and Wall Time*			
		Procedure-based		*Structure-based*	
UDC:902	48,000	$\gg$48,000	2,440 s	10	32 s
UDC:55	54,000	$\gg$54,000	3,938 s	10	45 s
UDC:902/904	107,000	$\gg$107,000	24,775 s	10	198 s
UDC:55/56	295,000	$\gg$295,000	189,100 s	10	945 s

The values allow to rate the discussed conventional approach (max. 1,000 loosely parallel pattern matching calls 'practical') using a procedure-based solution and the structure-based approach. The two examples of the approaches to challenging mining cases are using the same range of knowledge/data content each, specified by ranges of referenced conceptual knowledge. Requirements for the consideration of wider knowledge ranges do show a major impact on the procedure-based solution, resulting in relatively larger increase of context calls and wall times. Even if more loosely parallel calls would be logically possible with a mining algorithm it is not practical to increase their number on the same machine with procedure-based solutions. The counts of object entries in the two content resources are of major impact for the efficiency differences. The context calls' count (10) for structure-based fusion is based on the above presented multi-line object solution and can be kept stable. The result of the comparison of the computational footprint is clearly in favor of the structure-based solution. This tendency even improves with increasing numbers of objects involved.

ISBN: 978-1-61208-807-5

VI. Conclusion

This research achieved the goal to create performant methods for efficient problem solving deploying the new structure-based fusion methodology. Structure-based fusion can provide a valuable, scalable option alternative to procedure-based approaches. The presented case realisation successfully considered conceptual knowledge, especially the core component of UDC references, which is most important in context of handling advanced structures for universal, multi-disciplinary, and multi-lingual knowledge for many objects.

The case implementations illustrated that even complex scenarios with computational challenges and large numbers of involved objects can be efficiently created and realised. Structure-based methods increase the means to address structure and to beneficially use structural information, which is otherwise not easily deployable by procedure-based approaches. The solutions showed the flexibility of knowledge- and data-centricity. The implementations of the methods proved being able to minimise the number of calls and threads. The methodology and efficiency in creating and adapting implementations that way can have significant impact on sustainability and consistency of long-term solutions.

The structure-based fusion solutions not just provide facilities for fast, resource efficient operation, even if not optimised as the shown realisations. They are modular, long-term sustainable, and widely programming / language implementation independent. Realisations can be easily adapted to different environments (programming languages / shells and operating systems). For the research group and partners the solutions proved adaptability and efficiency in many practical realisation, for years, new and rewritten, in context of resources development and knowledge mining and many solutions beyond.

Future research will continue creating structure-based fusion solutions for knowledge mining and day-to-day challenges.

Acknowledgements

We are grateful to the "Knowledge in Motion" (KiM) long-term project, Unabhängiges Deutsches Institut für Multidisziplinäre Forschung (DIMF), for partially funding this research, implementation, case studies, and publication under grants D2012F1P04497, D2014F2P04562, D2018F1P04962 and D2020F1P05228 and to its senior scientific members and members of the permanent commission of the science council, especially to Dr. Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover, to Dipl.-Biol. Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, to Dipl.-Ing. Martin Hofmeister, Hannover, and to Olaf Lau, Hannover, Germany, for fruitful discussion, inspiration, practical multi-disciplinary case studies, and the analysis of advanced concepts. We are grateful to Dipl.-Ing. Hans-Günther Müller, HPE, Germany, for his excellent contributions and assistance providing practical private cloud and storage solutions. We are grateful to the members of the Eastern Mediterranean research and studies campaign 2018–2020, DIMF, and all national and international partners in the Geo Exploration and Information cooperations for their constructive and transdisciplinary support. We are grateful to The Science and High Performance Supercomputing Centre (SHPSC) for long-term support. / DIMF-PIID-DF98_007.

References

[1] Aristotle, The Ethics of Aristotle, 2005, Project Gutenberg, eBook, EBook-No.: 8438, Rel. Date: Jul., 2005, Digit. Vers. of the Orig. Publ., Produced by Ted Garvin, David Widger, and the DP Team, Edition 10, URL: http://www.gutenberg.org/ebooks/8438 [accessed: 2020-01-12].

[2] L. W. Anderson and D. R. Krathwohl, Eds., A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Allyn & Bacon, Boston, MA (Pearson Education Group), USA, 2001, ISBN: 978-0801319037.

[3] 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, Sept. 23, 2015, The Fifth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), The 13th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–29, 2015, Rhodes, Greece, 2015, pp. 1–7, DOI: 10.15488/3409.

[4] Plato, Phaedo, 2008, (Written 360 B.C.E.), Translated by Benjamin Jowett, Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Plato/phaedo.html [accessed: 2020-01-12].

[5] M. Foucault, The Archaeology of Knowledge. Routledge Classics, 2002, ISBN: 978-0-415-28752-4, Translated by A. M. Sheridan Smith.

[6] F. de Saussure, Cours de linguistique générale, 1916, (title in English: Course in General Linguistics), C. Bally and A. Sechehaye (eds.).

[7] Aristotle, On Interpretation, 2008, Peri Hermeneias, Organon, (Written 350 B.C.E.), Translated by E. M. Edghill Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Aristotle/interpretation.html [accessed: 2020-06-07].

[8] C. K. Ogden and I. A. Richards, The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism, 1923.

[9] R. Gleser, Rekonstruktion der Vergangenheit: Zur methodischen Eigenart prähistorischen Erkennens. Paderborn, 2018, pp. 199–237, in: Von der Quelle zur Theorie. Vom Verhältnis zw. Objektivität und Subjektivität in den historischen Wissenschaften, ISBN: 978-3-95743-136-3.

[10] "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: 2020-01-12].

[11] "Creative Commons Attribution Share Alike 3.0 license," 2012, URL: http://creativecommons.org/licenses/by-sa/3.0/ [accessed: 2020-01-12], (first release 2009, subsequent update 2012).

[12] C.-P. Rückemann, "Superordinate Knowledge Based Comprehensive Subset of Conceptual Knowledge for Practical Mathematical-Computational Scenarios," in The Ninth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), Proceedings of The 17th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–28, 2019, Rhodes, Greece, American Inst. of Physics Conf. Proc. AIP Press, Melville, New York, USA, 2020, ISSN: 0094-243X, (to appear).

[13] C.-P. Rückemann, Sustainable Knowledge and Resources Management for Environmental Information and Computation. Business Expert Press, Manhattan, New York, USA, Mar. 2018, Ch. 3, pp. 45–88, in: Huong Ha (ed.), Climate Change Management: Special Topics in the Context of Asia, ISBN: 978-1-94784-327-1, in: Robert Sroufe (ed.), Business Expert Press Environmental and Social Sustainability for Business Advantage Collection, ISSN: 2327-333X (collection, print).

[14] C.-P. Rückemann, "From Knowledge and Meaning Towards Knowledge Pattern Matching: Processing and Developing Knowledge Objects Targeting Geoscientific Context and Georeferencing," in Proc. of The Twelfth Internat. Conf. on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2020), November 21–25, 2020, Valencia, Spain. XPS Press, Wilmington, Delaware, USA, 2020, pp. 36–41, ISSN: 2308-393X, ISBN-13: 978-1-61208-762-7, URL: http://www.thinkmind.org/index.php?view=article&articleid=geoprocessing_2020_3_10_30035 [accessed: 2020-06-07].

[15] "Perl Compatible Regular Expressions (PCRE)," 2020, URL: https://www.pcre.org/ [accessed: 2020-06-07].

[16] "The Perl Programming Language," 2020, URL: https://www.perl.org/ [accessed: 2020-06-07].

[17] "UDC Summary Linked Data, Main Tables," 2020, URL: https://udcdata.info/078887 [accessed: 2020-06-07].

[18] "UDC 5: Mathematics. Natural sciences," 2019, URL: http://udcdata.info/025403 [accessed: 2020-01-12].

[19] "UDC 9: Geography. Biography. History," 2020, URL: http://udcdata.info/068076 [accessed: 2020-06-07].