

Human-Friendly, Machine-Readable: Coreon MKS for Visual Curation of Semantic Content and Linked Data

Alena Vasilevich, Michael Wetzel

Coreon GmbH, Rungestrasse 20, 10179 Berlin, Germany

Abstract

In this work we introduce Coreon Multilingual Knowledge System, a visual tool for concept-based data modeling and curation. Coreon environment is built on the visual paradigm; its goal is to offer a comprehensible and user-friendly solution for domain experts, well-versed in concept modelling dialects, as well as ad-hoc users, who are new to curation of semantic content. We describe the mechanism of the tool, which, being machine-readable, is powered by the language-agnostic knowledge graph, capably embedding the non-deterministic phenomena of the human language.

Keywords

Visual data curation tools, semantic knowledge management, terminology management

1. Introduction

A picture is worth a thousand words. With all compelling stories data can tell us, no means would be able to convey them as eloquently as visualization. Visual exploration revolutionizes how users navigate and inspect multifaceted, metadata-rich resources[1, 2]. Visual approach empowers data maintainers[3] and enables easier control over large collections of structured semantically-rich content.

Manually curated by domain experts, semantic resources carry a big value in the modern data world. Creation of such resources per se is a laborious and time-consuming process, whereas available software, meant to facilitate conceptual modeling, tends to lack a user-friendly, visual UI and often has a steep learning curve (quirks of various formats and standards aside). In this paper, we present a tool that addresses these challenges and delivers a clean visual concept-based modeling environment, accessible for ad-hoc users.

2. Coreon MKS: Digital Curation Tool with a Visual Aptitude

Coreon Multilingual Knowledge System (MKS)¹ is a visual web-based tool that aims to commoditize the curation and maintenance of structured data[4]. Having witnessed domain experts'

Qurator 2022: 3rd Conference on Digital Curation Technologies, September 19-23, 2022, Berlin, Germany

✉ alena@coreon.com (A. Vasilevich); michael@coreon.com (M. Wetzel)

🌐 <https://www.coreon.com/> (M. Wetzel)

🆔 0000-0002-9769-1885 (A. Vasilevich)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

📄 CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://www.coreon.com/>

frustration with cumbersome software, our goal is to reduce instrumental barriers and improve working experience for maintainers, who may struggle with obscure data-modeling tools, standards, format intricacies and their limitations[5].

Unlike scrolling through endless flat lists of entries, a dynamic and context-illustrating navigation allows maintainers to scale up and down within the data resource, zooming in on specific parts or branches that require their attention.

With the visual paradigm and model adaptability as its core features, Coreon MKS offers the user the following key functionalities:

- a collaborative environment for easy data curation, accessible for domain experts who had no previous experience with ontology-editing or knowledge-management software. In Coreon, users can create, combine, and edit terminologies, ontologies, taxonomies, controlled vocabularies, and graph-based data collections via familiar *drag-and-drop* actions, shortcutting extensive training and studying of software manuals;
- a customizable browser-based interface, suitable for ad-hoc users and domain experts, for internal and external contributors. The system reacts *on the user's role*, hiding pre-configured elements as well as selected properties, should they be flagged as *visible only for power users*, whereas maintainers have the rest of interaction elements at their disposal;
- supporting a variety of established formats: SKOS, TBX, RDF in all relevant syntax flavours (Turtle, N3, JSON-LD, etc.);
- validation of data against custom-established constraints and rules;
- data transparency – MKS tracks all record changes, tracing the complete evolution of the data which is often crucial for institutions and highly regulated industries;
- seamless interaction and smooth integration with third-party systems and services via RESTful Web API or SPARQL² protocol that relies on the well-pronounced LLOD standards;
- adherence to the Linked Data principles, supporting interoperability and re-usability of language resources [6];
- deployed as light-weight web-based SaaS, it needs no software installation, patches, and configuration updates;
- Single-Sign-On mechanism enables simple integration into enterprise authentication environments, based on OASIS' SAML2 standard.

3. Data Model

MKS is a semantic knowledge repository, with a language-independent knowledge graph as its backbone (see Figures 1, 2). Conceptually, the solution brings the best of Knowledge Organization Systems (KOS) and Terminology Management Systems (TMS) together. MKS leverages the lean structure of SKOS³ yet does not neglect natural language phenomena, bringing along the expressiveness of ISO's TermBase eXchange (TBX)⁴.

²<https://www.w3.org/TR/rdf-sparql-query/>

³<https://www.w3.org/2004/02/skos/>

⁴<https://www.tbxinfo.net/>

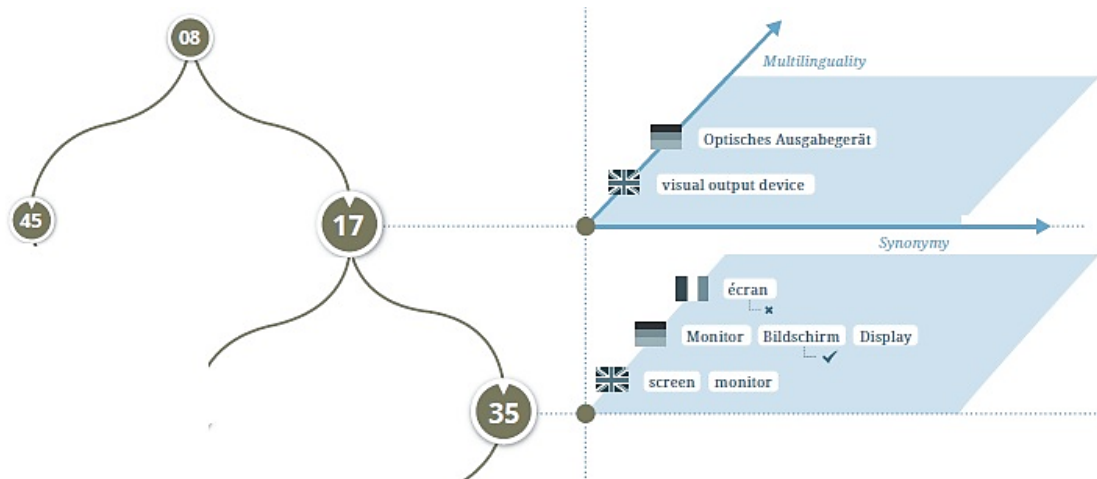


Figure 1: Linking a language-agnostic knowledge graph with richly annotated multilingual terminology.

Coreon is based on an open data model: the user does not have to fit rigid types of complex conceptual models but rather gets maximum flexibility in defining and configuring properties to describe the world. The model is therefore schema-less and easily adaptable for any metadata requirements. Data maintainers have different data types at their disposal to best define a concept and its terms (see Figure 3). In production environments, Coreon repositories encounter up to 100 custom properties to fully illustrate a domain.

4. Catering for Both Knowledge and Language

In a repository, all players receive unique identifiers – individual persistent IDs that unambiguously locate any given element, regardless of whether it is a concept, term, property or a concept relation. In essence, two kinds of information, i.e. knowledge via the semantically linked concepts and language via the terms, are modelled and stored separately from each other while being linked through unique *concept* identifiers.

From the knowledge perspective, the linking in MKS is performed **not** at the *term* but at the *concept* level. This approach makes concept maps completely independent from the terminological information. By linking entries this way, it becomes possible to model knowledge for phenomena that reflect the non-deterministic nature of the human language, such as word sense ambiguity, synonymy, and multilingualism. Each concept can be populated with unlimited descriptive and documentary information, in dozens of languages. Linking *per concept* also ensures smooth maintenance of relations without additional data clutter: relation edges are independent from labels, terms and their variants, and other metadata.

From the language perspective, Coreon’s concept model captures the following aspects:

- multi-directionality: terms in all languages that belong to the same concept are all stored in one record. Instead of maintaining several databases per language, a user can simply

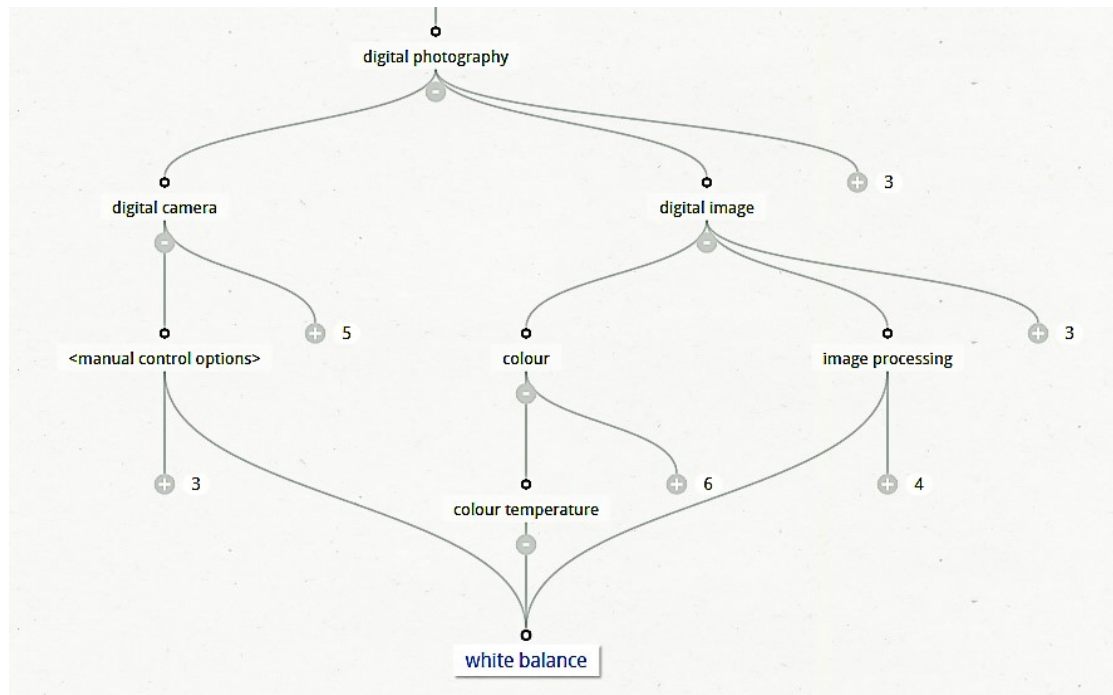


Figure 2: Graph-based data model supporting multiple inheritance.

change the desired language with a mouse-click;

- unlimited amount of terms: there is no need to maintain links between terms, neither is there *one synonym* field. Terms become synonyms as they are simply stored within the same concept; there is no limit on the amount of terms/synonyms for a concept;
- term autonomy: each term can carry a full set of descriptive metadata information. This means that all terms in all languages can be exhaustively described; different terms can become **preferred** to support different contexts, e.g., in one situation the full form *European Central Bank*, in another – its acronym *ECB*.

In contrast with TMSs, the power of Coreon knowledge graph gives users control over large amounts of data. Related concepts (e.g. *flat screen*, *LCD screen*, and *TFT display*) are not just listed under **F**, **L**, and **T** letters, which are alphabetically remote. Rather, they are semantically linked as parent-child concepts and rendered in proximity to each other on the concept map.

5. UI Components

A demo read-only MKS containing European Union's multilingual thesaurus (EUROVOC)⁵ can be accessed and explored at <https://www.coreon.com/dashboard/eurovoc>.

Coreon users interact with the environment through four main UI components:

⁵<http://europa.eu/eurovoc/>

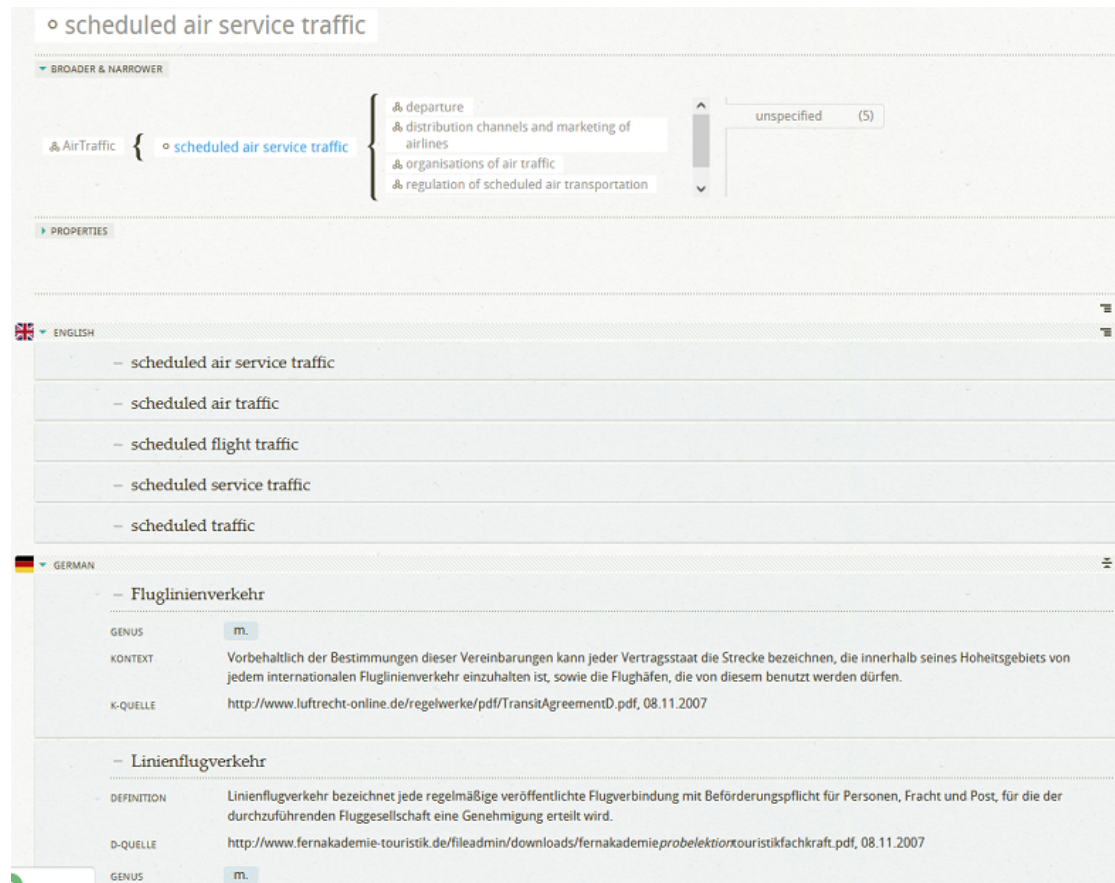


Figure 3: Concept view with synonyms and metadata.

- Concept View (Figure 3), displays all the information of one selected concept, in all languages, together with all metadata properties;
- Concept Map (Figure 4), displays the inter-concept relations in a tree-like hierarchy for easy and structured navigation and supports multiple inheritance;
- Term List, alphabetic, lexical access to all terms;
- Search field, to directly search and locate records in the concept map.

Coreon UI renders not only the hierarchical relations (explicitly shown in the concept-map mode in Figure 4) but also associated relations. To avoid visualization challenges triggered by a potential graph complexity, associated relations are clustered per relation type and currently are displayed for one concept at a time.

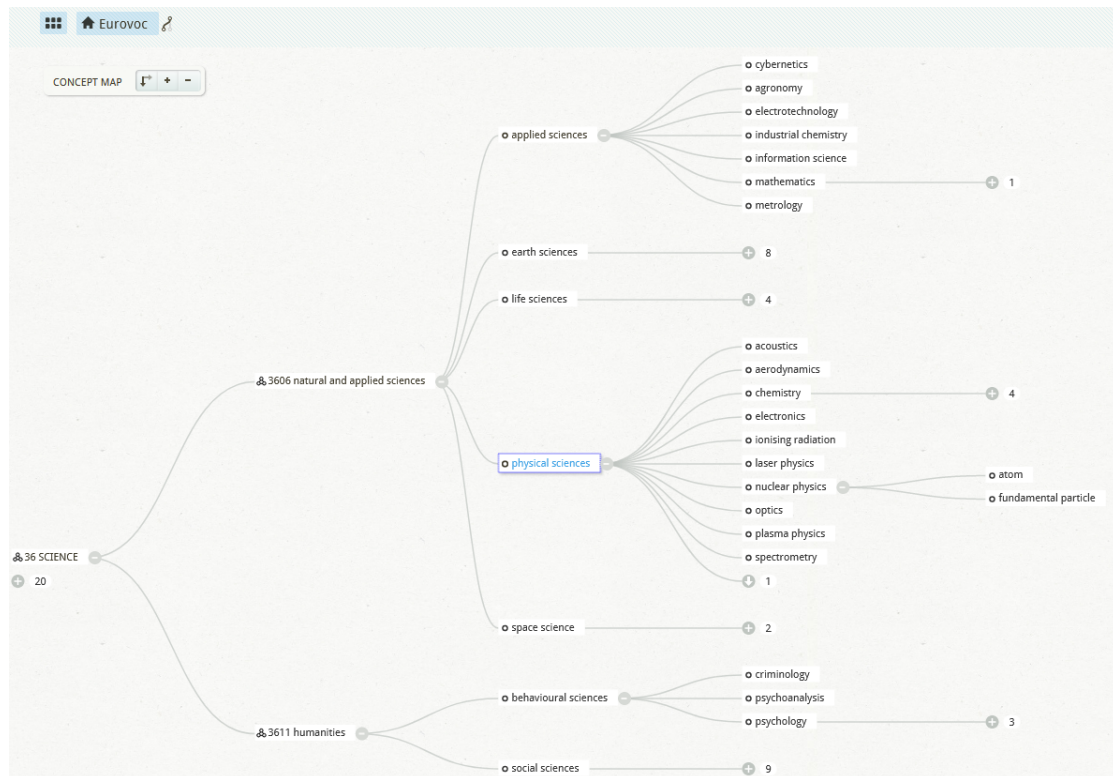


Figure 4: Concept Map in Coreon UI.

6. Conclusion

We believe that Coreon MKS removes instrumental barriers for conceptual data modeling, becoming a helping hand for domain experts, data maintainers, and other more *and* less tech-savvy stakeholders. MKS ensures that its users can focus on their domain-modeling tasks, without additional burden of format intricacies and extensive help from IT specialists.

Aside from being the tool for visual data maintenance, MKS ensures a smooth cooperation between domain experts and other professionals, interested in the final outcome of the data modeling activities. Being machine-readable, Coreon MKS promotes resource re-usability and integration into solutions that often benefit from injections of the structured data (e.g. development of virtual assistants, optimization of business processes, fine-tuning of machine translation and machine-learning models).

References

- [1] D. A. Keim, Visual exploration of large data sets, Communications of the ACM 44 (2001) 38–44.
- [2] N. Elmqvist, P. Dragicevic, J.-D. Fekete, Rolling the dice: Multidimensional visual exploration

- using scatterplot matrix navigation, *IEEE transactions on Visualization and Computer Graphics* 14 (2008) 1539–1148.
- [3] N. Tang, E. Wu, G. Li, Towards democratizing relational data visualization, in: *Proceedings of the 2019 International Conference on Management of Data*, 2019, pp. 2025–2030.
- [4] M. Wetzel, Multilinguale taxonomien mit coreon. wissens- und sprachmanagement in einer lösung, *Rechte, Rendite, Ressourcen. Wirtschaftliche Aspekte des Terminologiemanagements* 14 (2014) 41–51.
- [5] W. Ziegler, Metadaten für intelligenten content, *Intelligente Information: Schriften zur Technischen Kommunikation* 22 (2017) 51–66.
- [6] A. Vasilevich, M. Wetzel, Multilingual knowledge systems as linguistic linked open data for european language grid, in: S. Carvalho, R. R. Souza (Eds.), *Proceedings of the Workshops and Tutorials held at LDK 2021 co-located with the 3rd Language, Data and Knowledge Conference (LDK 2021)*, CEUR Workshop Proceedings, CEUR-WS.org, Zaragoza, Spain, 2021, pp. 126–134. URL: http://ceur-ws.org/Vol-3064/02_Alana_Vasilevich.pdf.