

Linked Data City - Visualization of Linked Enterprise Data

Joachim Baumeister^{1,2}, Sebastian Furth¹, Lea Roth¹ and Volker Belli¹

¹ denkbares GmbH, Friedrich-Bergius-Ring 15, 97076 Würzburg
² University of Würzburg, Am Hubland, 97074 Würzburg

Abstract. A generic technique for the visualization of hierarchical structures is introduced. The actual visualization is not only defined by the underlying data but also the application of domain-driven metrics. The paper shows two use cases for the analysis of linked enterprise data in the domain of technical service information systems.

1 Introduction

In the age of digitalization and automation of industries, many companies are consolidating their business information systems and product meta-data, such as ERP, CRM, file directories, and extranet data. In many cases, not all elements of these information resources are accessible to all relevant users. The intransparent access hinders effective work processes and often threatens business success. Therefore, the primary goal of many ICT projects is the linkage of the existing information silos into an integrated information infrastructure. Here, semantic technologies, and especially linked data models, are a successful enabler for building such *knowledge warehouses* mediating the information silos. Linked Enterprise Data [1] transfers the ideas and technologies of linked data [2] into the much more restricted world of business and enterprises. Standard semantic languages, such as RDF and SPARQL, are used to represent the core entities of the enterprise. Useful de-facto standard vocabularies for the enterprise usage already exist, see for instance SKOS [14] and GoodRelations [8]. Within a semantic infrastructure, all information resources are uniformly and semantically accessible by the user and novel services. In consequence, a number of advanced applications with business added value become possible [13]:

- Semantic enterprise search
- Semantic B2B portal with standardized data exchange
- Semantic assistants
- Automated data quality and curation processes

During the migration from the existing information structure to linked enterprise data, existing information sources need to be linked with semantic concepts. Here, a toolbox of core technologies ranging from Natural Language Processing/Information Extraction to Information Retrieval methods [4] is employed.

In Figure 1 the semantification process of enterprise data is depicted [5]. Each step of the process includes a detailed analysis:

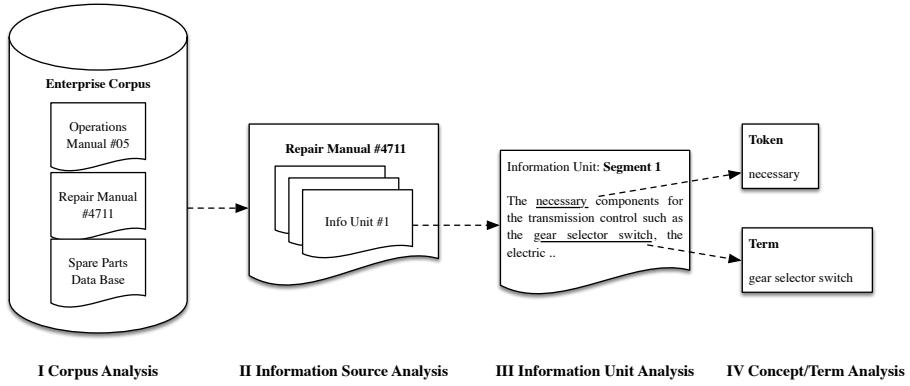


Fig. 1. Simplified process for the semantification of enterprise information sources.

I Corpus Analysis The existing data is collected and described. All relevant information systems are analyzed with respect to the included information sources.

II Information Source Analysis The included information sources are analyzed in more detail, for instance, with respect to the size and number of elements in the corpus.

III Information Unit Analysis Information sources are composed of information units, i.e., segments of a source that are not dividable anymore. Number and distribution of segments for information sources are relevant factors.

IV Concept/Term Analysis The use of ontological concepts in the information units is analyzed and improved by automated methods.

During the process, the exploration and visualization of the data is of core importance [11]. Since much of the data is generated during the process, the manual inspection and evaluation of the results need to be supported. Visualization methods help to understand existing deficiencies and motivate further process steps. In this paper, we introduce the generic visualization method *Linked Data City* that effectively supports the exploration and analysis of linked enterprise data during the semantification phase. The visualization of linked data differs from general ontology visualization methods [3, 6, 9], since usually linked data models exploit less relational structure but tend to be larger by orders-of-magnitude. We emphasize that the method is general usable for hierachic structures in a way that it can be deployed very easily in different scenarios.

The rest of the paper is organized as follows: First we introduce the core components of a linked data city, namely *buildings* and (nested) *districts*. Then, we describe the current implementation of the approach and demonstrate the usefulness of the visualization method by examples. In the conclusions we show promising steps for further work.

2 Linked Data Cities

The metaphor for visualizing linked data as a city is inspired by the work of Wettel [15]. In the original approach, the code of software applications is visualized as structures of a city. Figure 2 shows an example of a city visualization. Classes of software code are represented as *buildings* and code packages are defining the *districts* of a city. Special properties of classes and packages are communicated via color and size of the artifacts. Later this idea was adapted for the visualization of the test coverage after the evaluation of knowledge bases [7], where knowledge base elements are represented as buildings.

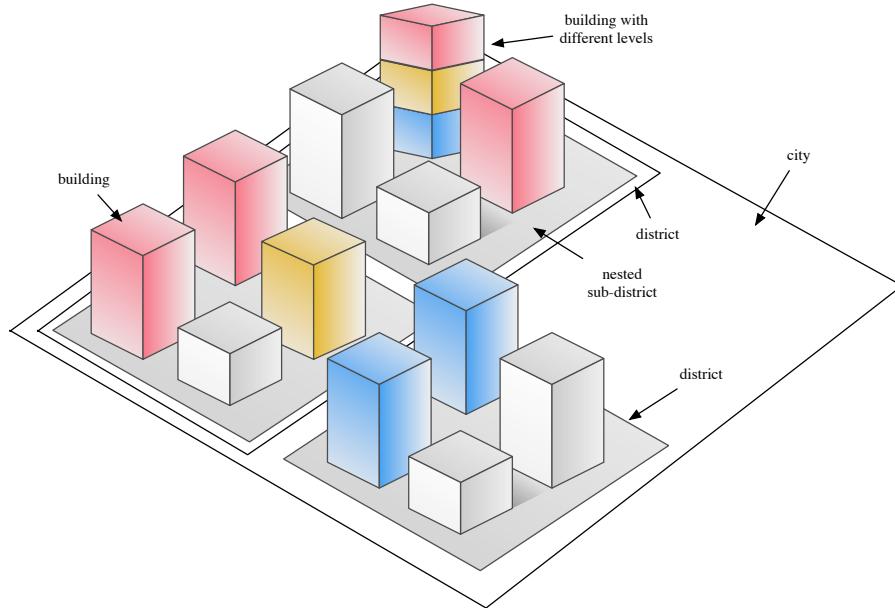


Fig. 2. Building, districts, and sub-districts of a data city.

Complex artifacts with part-of or hierarchical relations can be naturally visualized as a city: As we see in Figure 2, core elements of the artifacts are usually depicted as *buildings* of the city, that are grouped within *districts*. For deeper part-of or hierarchy relations the districts can be nested in *sub-districts*. The size of districts correlates with the number and sizes of the included buildings, whereas the size and color of the buildings represent specific performance indicators that are defined by the current analysis query. The specific configuration of buildings, districts, sizes, and colors is called *data city metric*. At the top of the figure we also see that buildings can have different levels. Different levels of a building are commonly used to include different metric attributes in the visualization.

For humans it is easy to understand the city metaphor. Like in the real world, local areas of the city are represented in districts and often the size of houses corresponds to their weight. For large cities (real and artificial) the user is familiar in incrementally explore the particular districts and buildings. For this reason, an advanced visualization application need to allow for the interactive exploration by drill-down and roll-up operations within the city.

3 Implementation

The presented visualization technique is implemented as a JavaScript library. That way, it can be easily integrated into (web-based) knowledge engineering tools but also runs as a stand-alone tool in a web browser. The definition of the city itself is represented as a JSON document. The following example shows books of a technical documentation that are represented as districts. Sections of a book are represented as buildings contained in the district.

```
{
  "Label": "Linked Data City 0815",
  "Districts": [
    {
      "Label": "Technical Documentation",
      "color": "#297B48"
      "Districts": [
        {
          "Label": "Repair Manual #4711",
          "color": "#848484#"
          "Buildings": [
            {
              "Label": "Section #4711.1",
              "color": "#0F2A65"
              "depth": 2,
              "height": 1,
              "width": 2
            },
            {
              "Label": ["Section #4711.2a",
                        "Section #4711.2b"],
              "color": ["#0F2A65", "#AF5C0B"],
              "height": [1, 3],
              "depth": 2,
              "width": 2
            },
            ...
          ]
        }
      ]
    }
  ]
}
```

The definition of the city is straight-forward, as we can see that districts can be nested in other districts. A leaf district contains a collection of buildings in

a corresponding element “Buildings”. For the district “Repair Manual #4711” two buildings are shown. The first building “Section #4711.1” is an example for a simple building having only one level included. The second building shows two labels, colors and heights representing the two levels of the building.

4 Case Study: Metrics for Linked Enterprise Data in Technical Service Information Systems

The architecture of a city is defined by the applied *data city metric*, i.e., the definition of colors, sizes, and levels of the buildings and districts. In this section, we demonstrate the approach by two use cases.

The presented metrics were used in the context of an industrial semantification project, where information sources of technical service information were analyzed. Here, buildings of a city represent a special kind of elements of the linked enterprise data. In enterprise systems the data refers to a domain-specific ontology. For instance, machinery builders typically align their data (documentation, parts, 3D models, etc.) to an ontology of products, components and functions, cf. [10, 12]. Enterprise information resources—document sections, parts, and wiring diagrams—are usually annotated by one or more instances of this hierarchy, for example a repair paragraph is annotated by the involved components and influenced functions.

In the following we present two basic metrics that investigate (a) the use of the product structure within the available information resources and (b) the availability of annotations in the information resources.

Use Case: Usage of Product Structure (UPS)

The *product structure* of an enterprise defines how products are organized in different levels. This organization includes multiple hierarchies for representing the relation of components and parts, but also for the functions of the product.

The primary subject of the UPS analysis is the actual use of the elements defined in the product structure. The use of the elements corresponds to annotations done with these elements included in the data of the investigated information systems. The metric is applied to find out how well the product structure is used in current enterprise information.

In the visualization, leaf elements of the product structure are represented by buildings and upper elements of the product structure are represented as wrapping districts. The height of the buildings correlates with the number of uses within all considered information sources. Higher buildings are thus used more often.

At the left side of Figure 3, a zoomed building representing the component “Engine block” is shown. The building itself is located in the district with the name “Engine”. We see that the building consists of multiple levels specializing the location of occurrences. Here, the element was used most often in the resource “doc#1” and the resource “3d#3”.

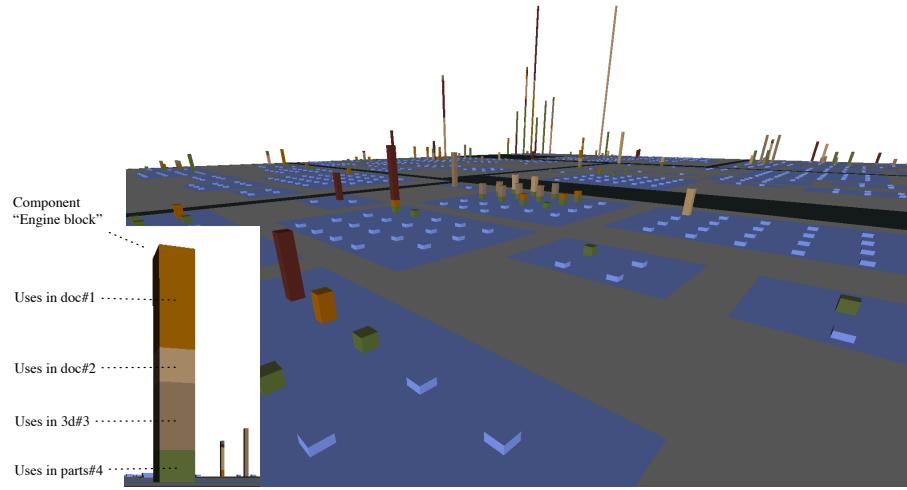


Fig. 3. Example for the usage of a product structure for a selection of technical documentation.

The visualization gives a very quick overview of actual application of a product structure. Unused areas can be easily spotted as well as elements with heavy use. Applied to a representative corpus of information resources, the visualization method points to areas in the structure that need refinement; both for lazy and frequent elements.

An interactive visualization is appropriate for very deep hierarchical structures. Then, buildings do not necessarily represent leaf-elements of the hierarchy but aggregated elements. Entering a building (e.g., by clicking on it) the building will drill-down the product structure and build a city visualization for all sub-elements contained in the aggregated element.

Use Case: Corpus Annotation Frequency (CAF)

Besides the actual use of the product structure the annotation frequency of the information resources is of prime interest. Usually, meta-data is attached to the information units to formally describe the contents. This meta-data mainly corresponds to elements of the product structure.

For the metric CAF, the city visualization is created as follows: The information sources in the corpus are represented as districts of the city, e.g., technical documentation, spare parts catalog, or FAQ data base. Sub-elements of these districts are further represented as nested sub-districts, e.g., a particular repair manual contained in the technical documentation or the spare parts catalog of a specific machine. Core information units are represented as buildings, for instance, a specific chapter of a repair manual in the technical documentation. The height of a building corresponds to its number of meta-data annotations; a

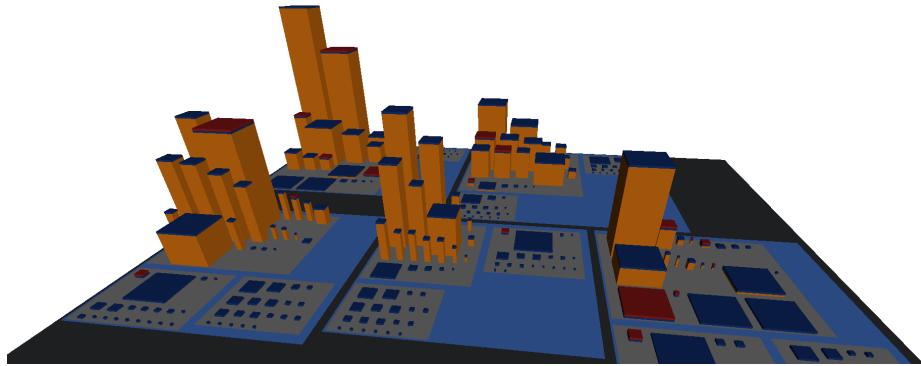


Fig. 4. Visualization of the annotations existing particular information units of a technical documentation for a specific machine.

buildings can have more than one level when different types of meta-data are included corresponding information unit. For instance, a chapter may include annotations of a component hierarchy but also of the functional hierarchy.

This visualization gives an overview of the corpus size and the existing annotations. Less annotated areas can be easily spotted but also districts (information sources, books types, etc.) with a high annotation quality. The results can help to motivate which areas of the structures need to be used much more frequently in information resources. With this knowledge, annotation initiatives (automated or manual) can be motivated and precisely planned.

5 Conclusions

Recently, more and more business information systems are transformed to linked enterprise data models. Appropriate visualization and exploration techniques support the semantification process of enterprise data. In this paper we presented Linked Data Cities, an interactive and generic method for the visualization of hierarchical structures. The actual visualization is defined by the application of a domain-specific metric. We introduced a number of metrics that showed its usefulness in an industrial semantification project.

In the future we are planning to improve the simplicity of the visualization by drill-down techniques, where similar buildings are clustered in aggregated building or districts. Then, even very large system structures can be (interactively) explored. Furthermore, we are working on the automated linkage of the city structure to existing linked data. Currently, scripts are used to transfer the information into the city data notation (the shown JSON). In the future, the automated transformation by SPARQL queries could be a possible simplification of this process.

References

1. Auer, S., Bühmann, L., Dirsch, C., Erling, O., Hausenblas, M., Isele, R., Lehmann, J., Martin, M., Mendes, P.N., van Nuffelen, B., Stadler, C., Tramp, S., Williams, H.: Managing the life-cycle of linked data with the LOD2 stack. In: Proceedings of International Semantic Web Conference (ISWC 2012) (2012)
2. Berners-Lee, T.: Linked data (2009), <http://www.w3.org/DesignIssues/LinkedData.html>
3. Fluit, C., Sabou, M., van Harmelen, F.: Ontology-based information visualization. In: Visualizing the Semantic Web, pp. 36–48. Springer (2006)
4. Furth, S., Baumeister, J.: On the semantification of 5-star technical documentation. In: Proceedings of the LWA 2015 Workshops: KDML, FGWM, IR, and FGDB. pp. 264–271 (2015)
5. Furth, S., Baumeister, J.: Semantification of large corpora of technical documentation. In: Atzmueller, M., Oussena, S., Roth-Berghofer, T. (eds.) Enterprise Big Data Engineering, Analytics, and Management. IGI Global (2016), <http://www.igi-global.com/book/enterprise-big-data-engineering-analytics/145468>
6. Geroimenko, V., Chen, C. (eds.): Visualizing the Semantic Web. Springer, 2 edn. (2006)
7. Hatko, R., Baumeister, J., Puppe, F.: Coveragecity: Test coverage for clinical guidelines. In: The 8th Workshop on Knowledge Engineering and Software Engineering (KESE2012) (2012), http://ceur-ws.org/Vol-949/kese8-01_02.pdf
8. Hepp, M.: GoodRelations: An ontology for describing products and services offers on the web. In: Gangemi, A., Euzenat, J. (eds.) EKAW. Lecture Notes in Computer Science, vol. 5268, pp. 329–346. Springer (2008), <http://dblp.uni-trier.de/db/conf/ekaw/ekaw2008.html#Hepp08>
9. Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E.: Ontology visualization methods - a survey. ACM Comput. Surv. 39(4) (Nov 2007), <http://doi.acm.org/10.1145/1287620.1287621>
10. Lin, J., Fox, M.S., Bilgic, T.: A product ontology. Enterprise Integration (1997)
11. Mader, C., Martin, M., Stadler, C.: Facilitating the exploration and visualization of linked data. In: Auer, S., Bryl, V., Tramp, S. (eds.) Linked Open Data—Creating Knowledge Out of Interlinked Data, pp. 90–107. Lecture Notes in Computer Science, Springer International Publishing (2014), http://dx.doi.org/10.1007/978-3-319-09846-3_5
12. Mohammad, N.N., Amin, I.M., Othman, R.M., Asmuni, H., Hassan, R., Kasim, S.: Design and implementation of product structure ontology. Ontology-Based Applications for Enterprise Systems and Knowledge Management p. 246 (2012)
13. Oberle, D.: How ontologies benefit enterprise applications. Semantic Web 5(6), 473–491 (2014)
14. W3C: SKOS Simple Knowledge Organization System reference: <http://www.w3.org/TR/skos-reference> (August 2009)
15. Wettel, R., Lanza, M.: Visualizing software systems as cities. In: Visualizing Software for Understanding and Analysis, 2007. VISSOFT 2007. pp. 92–99 (2007)