DOI 10.14622/Advances_49_2023_04

# TaxoCatalog: costless workflow integration of an expert system environment to personalize product catalogs in omni-channel context

*Heiko Angermann*

Hochschule Darmstadt – University of Applied Sciences, Faculty of Design
E-mail: heiko.angermann@h-da.de

**Short abstract**

Taxonomies are a formal method to semantically structure information using hierarchical ordered concepts. Those play a crucial role in omni-channel retailing to publish product catalogs across media channels, i.e. digital (portal-, paper-based) and print media (paper-based). Portal-based media use taxonomies to structure product-related content by concepts to facilitate customers navigating through the e-commerce site. That are the product categories. Paper-based media require taxonomies for automatic layout setting using data-driven publishing software. To achieve consistency, all media use the identical product taxonomy as basis. When additionally using recommender systems, products are published individually on the e-commerce site. Paper-based media, in contrast, show content regardless of dynamic preferences. Limited personalization takes place using sub-catalogs to only show products of, or specific categories. However, the resulted information loss may result in sales loss, as preferences change. An approach to produce paper-based catalogs without information loss is not available. This is, as individual layout setting and printing is uninteresting, if automating the required processes cannot be achieved. With TaxoCatalog, an expert system environment is presented, which can automatically process customers' preferences for producing individual product catalogs without information and automation loss. Using TaxoCatalog, a semantic personalization is computed, before the layout is set. The first component of TaxoCatalog is semantically analyzing the initial product taxonomy using natural language processing (NLP). The second component is analyzing customers' preferences to create a personalized taxonomy based on the NLP outcome. The third component is used to scale the personalized taxonomy, depending on the desired volume of the catalog. The fourth component is used to perform the rule-based layout setting using standard data-driven publishing software. All decision processes are achieved using background knowledge to allow a cost-less integration into standard workflows. A comprehensive evaluation using two public and one private database provided by a German retailer underline TaxoCatalogs' efficiency.

**Keywords:** data driven publishing, expert systems, content flow, premedia, workflows

## 1. Introduction

Today, products are advertised on different media channels (Hänninen, Kwan and Mitronen, 2021). Digital media channels in the form of internally and externally operated e-commerce portals, digital media in the form of paper-based documents, and paper-based print media products can be used (Gao, Melero and Sese, 2020). All have own benefits, but when using e-commerce, retailers can derive customers' preferences in real-time using recommender systems (Chen, et al., 2022; Gil-Gomez, et al., 2020). As a result, products can be shown individually to customers based on its order probability (Behera, et al., 2020). For example, suitable products are shown at the bottom of product sites, known as cross- and up-selling (Norvell, Kumar and Contractor, 2018). Paper-based digital product catalogs, in contrast, show static content, even if its appearance is dynamic. Here, formats like EPUB, or (interactive) PDF are used. Printed product catalogs, also show static content, as individualization affects de-automation of processes (Hoffmann-Walbeck, 2022). Usually, the specific format PDF/X is used for printing. Minimal individualization is enabled for paper-based product catalogs using proven methods of print-on-demand, variable data printing, sub-catalogs, and programmatic printing.

Using print-on-demand, the catalog is produced based on customers' interaction, e.g. an order. Using variable data printing, single pages of the catalog are individualized, e.g. an individual name (Lin, 2006). Using sub-catalogs, catalogs containing only a selection of products or categories are produced, e.g. for a season (Angermann and Ramzan, 2016). Using programmatic printing, a combination of the before-mentioned techniques and paradigms is achieved, which is possible since the advent of non-impact digital printing. A typical use-case is: the customer interacts on an e-commerce site, the preference is analyzed, the media is produced and shipped to the customer. For example, when a shopping cart is not ordered, the customer receives a mailing showing the products or alternatives, and a promotion code. Hereby, programmatic printing is not limited to print media, or to a specific range of print media products. The media can also be sent digitally.

Of course, programmatic printing is a major step forward, as it enables sharing knowledge across channels, e.g. customers' preferences. However, the decisive factor is that produced media shows a certain selection of products based on preferences. This leads to an information loss, as a large amount of products is not shown. This in turn, leads to a potential loss of sales. Sending the standard main catalog or a predefined sub-catalog instead, has no added value either. A discount could be provided here as well, but the rest of the content would be regardless of preferences.

To provide a fully-automatic solution for producing paper-based product catalogs without information loss, the expert system environment TaxoCatalog is presented. The main difference compared to existing works is that the structure of the catalog is semantically personalized based on the underlying product taxonomy, before the layout setting is performed. Four independent components that are based on using background knowledge, are part of TaxoCatalog. All have in common that the initial taxonomy can remain. The first component is aimed to semantically analyze correspondences of the concepts being part of the taxonomy, i.e. the product categories. The second component is analyzing the customers' individual preferences based on the order history. The third component is automatically modifying the taxonomy based on the computed preferences and correspondences, as well as the desired level of personalization. The fourth component is finally using the personalized taxonomy to automatically generate the personalized product catalog using standard data-driven publishing software. In summary, TaxoCatalog provides three contributions to the field of data-driven publishing and the printing industry:

- TaxoCatalog provides the first omni-channel application that is capable of using complex knowledge for producing paper-based product catalogs. The approach further advances the interlinking of portal-based digital media, and paper-based digital and print media.
- TaxoCatalog provides the first solution for personalizing paper-based product catalogs without resulting an information loss. This is achieved by semantically personalizing the content structure based on the underlying product taxonomy, before layout setting.
- TaxoCatalog provides the first solution that is able to dynamically but still fully-automatically scale the level of desired semantic personalization.

## 2. Background

This section succinctly describes the underlying concepts necessary to develop TaxoCatalog and to understand the used method explained in the next section. Four concepts are explained briefly: omni-channel retailing, taxonomy applications, expert systems, and data-driven publishing.

### 2.1 Omni-channel retailing

The retailing industry has been significantly influenced by the digitization of media channels. Today, retailers use omni-channel retailing, which allows that all channels are supplied by one single publishing source, but the digital channels can nevertheless be networked with each other (Verhoef, Kannan and Inman, 2015). In this way, the retailer can explore knowledge about the customer's preferences in real-time to display product-related content in a personalized manner on the digital portal-based channels. To acquire those preferences, recommender systems are used. Such systems use and combine different statistical methods to measure order probability of products.

## 2.2 Taxonomy applications

Taxonomies are a formal methodology to semantically structure information using hierarchical ordered concepts (Angermann and Ramzan, 2017). The underlying concept can be described as a two-tuple $\tau = \{\varphi,\rho\}$, where $\rho$ is a set of edges for connecting partially ordered concepts of $\varphi$. Each edge represents a semantic hypernym-hyponym (is-a) relationship between the concepts. For example, a "printed book" is-a "printed product" is-a "media product". Each concept is defined by a label and the position in $\tau$. The deeper its position, the more specific the concept is. Taxonomies have many applications in information and content management (Angermann, 2017):

- In Product Information Management a taxonomy is used to hierarchically structure concepts in the form of product categories. These are used to group similar products, and to assign category-specific (e.g. attributes) and -unspecific (e.g. title) content to products.
- In Media Asset Management, the above-mentioned product categories are used to structure product-related media (e.g. images), and to assign those to the products.

In omni-channel retailing, the taxonomies are the most essential tool for publishing content across channels (Angermann, 2022). The data stored in the content management systems, is transferred via interfaces to the e-commerce portal. Same for transferring products to external marketplaces. And, the taxonomy is also required for enabling data-driven publishing. Its structure is used by the templates to automatically fill the pages with dynamic content provided using interfaces.

## 2.3 Expert systems

Expert systems are one kind of intelligent systems that aim to acquire new knowledge based on existing knowledge (Oleshchuk and Fensli, 2011). The main purpose of the systems is to derive decisions as done by a human expert (Mirmozaffari, 2019). In contrast to machine learning, the initial knowledge is acquired from domain experts, and not from training. Formally, this is known as background knowledge (BK) (Angermann and Ramzan, 2017). These can be internally or externally generated from different sources and in different formats, e.g. taxonomies, thesauri, etc.

## 2.4 Data-driven publishing

Data-Driven publishing is a rule-based technique for automatically setting the layout for paper-based documents using a layout generator software. This software provides different data interfaces to databases of different formats (e.g. SQL, XML), and a graphical interface for a designer, to set the data connection, and to create the needed master pages and templates:

- The master page defines the dimensions of the document, and of the page frame. That is the area of the paper that actually contains content (document minus margins).
- The (layout) template defines the position and size of static content to be included in a page frame using content frames, as well as the rules for formatting dynamic content.
- A data source finally provides the different content for the content frames as a data stack.

The most distinguish feature compared to earlier attempts is, that now content types are used (Enlund and Andersin, 2007). Those content types (e.g. for title, description, etc.) are also assigned with rules for positioning and sizing, but also with rules for formatting the dynamic content. Thus, the dynamic content can be independently designed for producing various documents. Both, the needed structure of the data, and the included content types, is enabled using taxonomies.

Another criteria is that the content frames of the templates can be filled automatically by the above-mentioned stack. To do so, two conditions must be met (Gundougan, 2022): separation of the (media neutral) content from the layout, metadata assignment to the content. Both is achieved using single source publishing, which is already required when providing omni-channel retailing. Most retailers fulfill all criteria and conditions by using the hierarchical database format XML. It has the advantage that the data is already structured hierarchically, and the metadata can be individually specified in the form of so-called tags (van der Vlist, 2002). In addition, the required hierarchy and metadata can be validated using the data format Doctype Definition (DTD).

## 3. TaxoCatalog

In this section, the use-case of TaxoCatalog is discussed, before the components are explained. The implementation of the program will be shown at the conference.

### 3.1 TaxoCatalog use-case

The main difference compared to existing works is that the proposed use-case is semantically personalizing the product taxonomy, before the layout is set. A current order does not necessarily have to be an authoritative interaction to start the workflow of the system (see Figure 1):

1. The initial product taxonomy is analyzed with regard to semantic correspondences.
2. The customers' preferences are analyzed based on the order history. Orders that have been placed longer in the past are viewed at a correspondingly lower rate. The expert of the retailer can dynamically define to what extent part orders are taken into account.
3. The semantic correspondences and the preferences are taken into account to compute a personalized product taxonomy. The expert can define how strongly the taxonomy is semantically personalized. Theoretically, categories (and products) can be excluded, but this is not the focus of TaxoCatalog, since it results in a loss of information as explained.
4. The personalized taxonomy defines the structure of the catalog to be layout to fill the layout template with the dynamic content, as required for data driven publishing.
5. Finally, the file is exported according to a media-specific format, e.g. EPUB, PDF, or PDF/X.

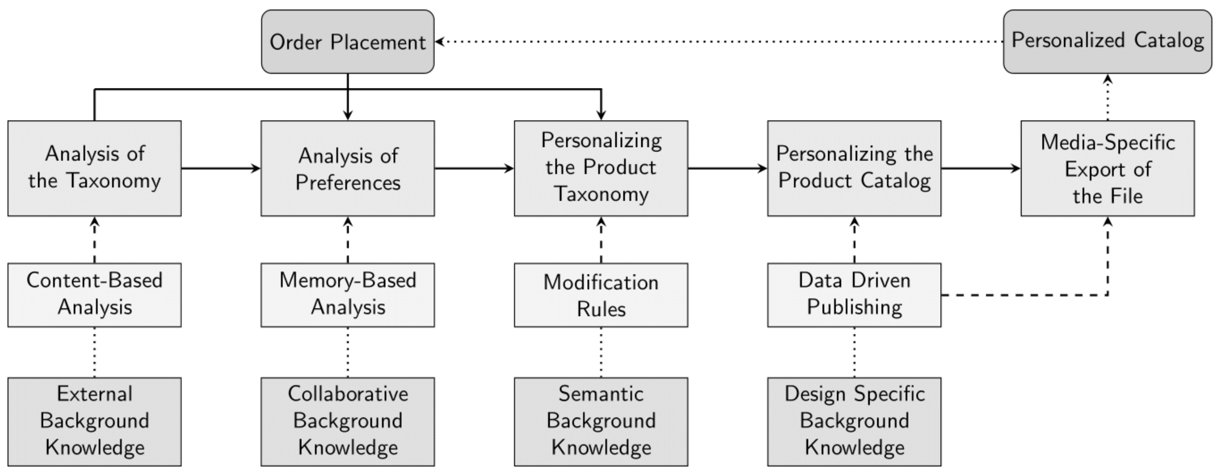


*Figure 1: Use-case and workflow of TaxoCatalog, its components, and used background knowledge*

### 3.2 TaxoCatalog components

Four components form TaxoCatalog. All decisions are automated using background knowledge to allow a cost-less integration into print media workflows (see Figure 1) (Angermann and Ramzan, 2016).

The first component is aimed to semantically analyze correspondences of concepts, i.e. the product categories of the initial product taxonomy. Its computation results a four-tuple $\delta$ for all possible combinations of $\varphi$: $\delta = (\varphi_1, \varphi_2, \rho, \varepsilon)$, where $\rho$ states if two concepts $\varphi_{1,2}$ sharing the same is-a relationship are semantically similar or not. This is resulted through a computed value $\varepsilon$ (between 0 and 1) and its comparison with a dynamic threshold. The computation of $\delta$ is performed using content-based natural language processing techniques. To increase the quality of the computation, external background knowledge is used as suggested in relevant literature (Angermann, Pervez and Ramzan, 2017).

The second component is analyzing the customers' individual preferences based on the order history. Here, a memory-based approach is used. This means that past orders are taken into account, as well as collaboratively acquired background knowledge, i.e. the analysis of other customers (Angermann and Ramzan, 2016). Its computation results a two tuple $\omega = (\varphi_{1,2}, \beta)$, where $\varphi_{1,2}$ are two semantically similar concepts, and $\beta$ states the level of preference.

The third component is modifying the taxonomy based on $\delta$ and $\omega$ using different taxonomic operations, e.g. combining concepts and extending concepts. Each modification is performed in such a way that the semantics are adapted, but not corrupted. If a concept is extended, e.g., by further concepts, then this gets another label if necessary. This is particularly important, otherwise the customer may not be able to interpret the categories correctly. Since the new taxonomy was created on the basis of the initial taxonomy, the products can be assigned to the new categories without major hurdles. And, the expert can further define the semantic volume of the personalized taxonomy. This is performed using rules stored as background knowledge.

The fourth component is finally using the modified taxonomy to automatically generate the personalized product catalog using data-driven publishing.

## 4. Results

The efficiency of TaxoCatalog was evaluated with regard to a wide range of criteria and databases (see Table 1). The results for the two most relevant are summarized in this section. These are the *F*-Measure of the recommender system, and the semantic flexibility of personalizing taxonomies.

*Table 1: Characteristics and parameters of the databases used for experimental results*

| Criteria | Adventureworks | Northwind | Festool |
|---|---|---|---|
| *N* Customers | 700 | 93 | 500 |
| *N* Orders | 31 464 | 829 | 1 400 |
| *N* Products | 320 | 77 | 118 |
| *N* Concepts | 42 | 31 | 53 |

The *F*-Measure score states the decision quality of the system to analyze customers' preferences, compared to a human expert. A score of 1 means that in all test cases, the right decision has been made. The average *F*-Measure score of TaxoCatalog is 0.93 for all three databases: 0.92, 0.88, and 0.98. It is particularly noteworthy that the good score is similar for all three databases. Thus, an equally good result can be achieved for a variety of products and publications. The Semantic Flexibility is measured to verify the flexibility when using the taxonomy as main instrument for personalizing the catalog. Two scores are measured. Firstly, the reduction flexibility is computed by identifying the maximum possible decrease of the taxonomy. Secondly, the enlargement efficiency is computed to verify a maximum possible decrease of the taxonomy. The average reduction flexibility score is 48.10 %. The average enlargement flexibility score is similar: 51.26 %. This states that the taxonomy provides a great solution in terms of semantic flexibility.

## 4. Conclusions

This work presented TaxoCatalog, the first taxonomy driven expert system environment for automatically personalizing paper-based product catalogs. Compared to related works, TaxoCatalog does the personalization without information and automation loss. As future work, the content types will be taken into account to also individualize the included graphic design.

### Acknowledgements

Thanks to the Festool GmbH & Co. KG and the InBetween Deutschland GmbH for providing data and software for the presented research.

### References

Angermann, H. and Ramzan, N., 2016. TaxoPublish: towards a solution to automatically personalize taxonomies in e-catalogs. *Expert Systems with Applications*, 66, pp. 76–94. https://doi.org/10.1016/j.eswa.2016.08.058.

Angermann, H., 2017. *Manager's guide to SharePoint Server 2016: tutorials, solutions, and best practices*. Berkeley: Apress/Springer.

Angermann, H. and Ramzan, N., 2017. *Taxonomy matching using background knowledge*. Cham: Springer.

Angermann, H., Pervez, Z. and Ramzan, N., 2017. *Taxo-Semantics*: assessing similarity between multi-word expressions for extending e-catalogs. *Decision Support Systems*, 98, pp. 10–25. https://doi.org/10.1016/j.dss.2017.04.001.

Angermann, H., 2022. TaxoMulti: rule-based expert system to customize product taxonomies for multi-channel e-commerce. *SN Computer Science*, 3(2): 177. https://doi.org/10.1007/s42979-022-01070-8.

Behera, R. K., Gunasekaran, A., Gupta, S., Kamboj, S. and Bala, P. K., 2020. Personalized digital marketing recommender engine. *Journal of Retailing and Consumer Services*, 53: 101799. https://doi.org/10.1016/j.jretconser.2019.03.026.

Chen, Y., Li, M., Song, J., Ma, X., Jiang, Y., Wu, S. and Chen G.L., 2022. A study of cross-border E-commerce research trends: based on knowledge mapping and literature analysis. *Frontiers in Psychology*, 13: 1009216. https://doi.org/10.3389/fpsyg.2022.1009216.

Enlund, N. and Andersin, H.E., 2007. The early days of computer aided newspaper production systems. In: J. Impagliazzo, T. Järvi and P Paju, eds. *History of Nordic Computing 2: Second IFIP WG 9.7 Conference – HiNC 2*. Turku, Finland, 21–23 August 2007. Springer, pp. 238–249.

Gao, L.-X., Melero, I. and Sese, F.J., 2020. Multichannel integration along the customer journey: a systematic review and research agenda. *The Service Industries Journal*, 40(15-16), pp. 1087–1118. https://doi.org/10.1080/02642069.2019.1652600.

Gil-Gomez, H., Guerola-Navarro, V., Oltra-Badenes, R. and Lozano-Quilis, J.A., 2020. Customer relationship management: digital transformation and sustainable business model innovation. *Economic Research-Ekonomska Istraživanja*, 33(1), pp. 2733–2750. https://doi.org/10.1080/1331677X.2019.1676283

Hänninen, M., Kwan, S.K. and Mitronen, L., 2021. From the store to omnichannel retail: looking back over three decades of research. *The International Review of Retail, Distribution and Consumer Research*, 31(1), pp. 1–35. https://doi.org/10.1080/09593969.2020.1833961.

Hoffmann-Walbeck, T., 2022. *Workflow automation: basic concepts of workflow automation in the graphic industry*. Cham: Springer.

Lin, X., 2006. Active layout engine: algorithms and applications in variable data printing. *Computer-Aided Design*, 38(5), pp. 444–456. https://doi.org/10.1016/j.cad.2005.11.006.

Mirmozaffari, M., 2019. Developing an expert system for diagnosing liver diseases. *European Journal of Engineering and Technology Research*, 4(3), pp. 1–5. https://doi.org/10.24018/ejeng.2019.4.3.1168.

Norvell, T., Kumar, P. and Contractor, S. (2018). Assessing the customer-based impact of up-selling versus down-selling. *Cornell Hospitality Quarterly*, 59(3), pp. 215–227. https://doi.org/10.1177/1938965518762836.

Oleshchuk, V. and Fensli, R., 2011. Remote patient monitoring within a future 5G infrastructure. *Wireless Personal Communications*, 57(3), pp. 431–439. https://doi.org/10.1007/s11277-010-0078-5.

van der Vlist, E., 2002. *XML schema: the W3C's object-oriented descriptions for XML*. Sebastopol, CA: O'Reilly Media.

Verhoef, P.C., Kannan, P.K. and Inman, J.J., 2015. From multi-channel retailing to omni-channel retailing: introduction to the special issue on multi-channel retailing. *Journal of Retailing*, 91(2), pp. 174–181. https://doi.org/10.1016/j.jretai.2015.02.005.