

Introducing an Authoring Tool for a new Generation of Law Dictionaries at Beck Publishers

Susanne Hauptmann, Wolfgang Zenk

Verlag C.H.Beck

Wilhelmstr. 9

80801 München / Germany

susanne.hauptmann@beck.de

Acolada GmbH

Lilientronstr. 13

90455 Nürnberg / Germany

w.zenk@acolada.de

Abstract

This paper describes the introduction of an authoring tool at Beck Publishers for their new generation of legal, commercial and political dictionaries. It first gives an overview over the goals - the editorial, the economical and the technical aspects - that shall be met with the introduction of a new authoring tool. Then the background of the tool used is given in more detail, the data modelling and the methodology used. The tool and the editing approach is then contrasted to other approaches in lexicographical projects. Finally some examples are given to give an impression how authors actually work with the tool and the results that they may produce by themselves.

1. Introduction

The motivation when introducing a new authoring tool is not the technical aspect of using a state-of-the-art tool. The motivation comes from a careful analysis of current working conditions, the contribution from authors, the editorial support authors get from the publisher, time to market considerations and finally but foremost the future planning of new projects, markets to target and improved production workflows which are closely linked to economical considerations.

In a large law publisher like C.H.Beck with a backlist of several thousand titles dictionaries naturally do not play the predominant role. The targets however are much the same as for other reference books: support of both traditional and new media, more productions in less time and no dependency on one typesetter / typesetting system. Every dictionary production needs to fulfill economic conditions set forth. With regard to these requirements single source publishing, i.e. one data source - multiple target formats / multiple media has almost become a must. Single source and cross media production are closely linked to the buzzword XML and can be best realized with database driven systems.

Beneath the publisher's requirements also the abilities and the demands of its authors need to be taken into account. Dictionary authors are experts in their subject, and they are experts in dictionary usage. All authors at Beck publishers work as external authors, i.e. they

hand in a manuscript whether it be in written or in an electronic format. From their own dictionary using practise they have strong ties to representation. But authors would rarely be experts in data structuring. The ability to work on structured data, the guidance to do this and to maintain layout at the same time has been one of the key considerations. Tools introduction defines responsibilities: working on the content is the author's responsibility whereas the publisher is responsible for a corporate style and the layout.

With UniTerm Pro, C.H.Beck and Acolada have chosen a tool which supports three key requirements:

- the author's support: ease of use with structured data
- close connection to representation / layout as the most important media output
- the media independent data format.

Another intention with the tool usage is that the growing size of a dictionary increases also the author's motivation so that a limited wordlist quickly develops into a full-fledged dictionary.

Project requirements and motives

There are hard factors (requirements) and soft factors (motives) for the introduction of a new authoring tool. Soft factors focus more on psychological aspects, support and guidance of the publisher's authors. On a general level aspects soft factors need to be treated separately from technical and economical aspects.

Soft factors: requirements and motives with regard to C.H. Beck dictionary authors

- Easier recruitment of new authors
- Motivate authors "to write" - the layout interface immediately shows the success
- Create stronger ties to C.H.Beck publishers
- Easier transition to new authors

Hard factors: technical and economical aspects

- Setup of editorial guidelines which had not been defined before
- Quality assurance through comparable contents
- Constant quality secures the brand name
- Possibility to import data sets which may already be present in another format (make use of "by-products")
- Production into different media from only one platform (single-sourcing)
- Cost saving in typesetting

Among motives and requirements for authors there are of course objective factors like usability, easy training on the tool and a much easier and smoother transition if an author is to be replaced and a dictionary title handed over to a new author. But the tool usage also has a strong psychological component: providing an editing tool and training on it creates additional ties and a closer author-to-publisher relationship. More specifically the tool shall animate authors to write. This is achieved by a template based data capturing on the one side (introduce additional information to describe a term) and via an immediate control of success, i.e. the author has the possibility to "produce" the final result at any stage by himself. The author thus will become his own editor because incorrect input can be immediately detected and corrected in the source data. This will cut down revision and

proofreading cycles. It releases editors and streamlines typesetting. The possibility to control the volume of the dictionary produced so far at any time and to check at the same time the number of new and the number of revised entries provides very precise information to prepare print and electronic production, marketing material, etc.

Most importantly on the economical / technical side are editorial guidelines which are now applied to the whole range of dictionaries. The authoring tool equally allows to build up dictionary lines: horizontal lines by applying the tool / the same dictionary structures for different language combinations and vertical lines by re-using the dictionary data for different editions and for different target groups. The editorial guidelines in combination with structured data management thus assure consistency, quality and add to the C.H.Beck brand.

Conversion routines equally provide the inclusion of material that an author may already have collected. Once imported the data is maintained in a unicode XML database. From this source the data is then processed automatically and provided for the relevant target media (keyword: single source publishing).

The new generation of dictionaries...

What marks the new generation of juridical dictionaries? One could almost describe it as a paradigm change: away from annotated wordlists towards an explanatory dictionary. The new generation of Beck dictionaries takes over a number of characteristics from general language lexicography: more elaborate micro structure, phraseology, contexts, and explanations. Particularly in the context of bilingual juridical dictionaries describing different legal systems, so called info boxes are used to explain different juridical concepts, comment on false friends and give additional comments about interrelations between different concepts. The traditionally strong focus on the number of headwords is pushed into the background to make room for a more pedagogical and target group oriented approach.

In the authoring tool this is emphasized by entry templates that always offer a number of data categories to fill in. This design creates awareness and invites authors to contribute more than mere word lists.

How does the approach differ from approaches in other lexicographic projects?

With UniTerm Pro, Beck Publishers tries to bridge the gap between a structured, XML driven editing and a layout driven editing. Keywords in this approach are: reduced complexity in XML editing and full layout support via an integrated print production environment.

Reduced XML complexity: building structures and substructures in XML rapidly becomes a highly complex tasks in DTD design, author training and editing. This complexity and the introduction of standard XML editors can be managed in in-house projects but becomes much more difficult when working with external authors.

To better characterize the UniTerm Pro authoring tool the following matrix is used to list other approaches for structured dictionary editing:

Tool	Characterization
XML Editor, e.g. XmetaL	<ul style="list-style-type: none"> • Structured data editing • Requires in-depth knowledge of XML structures • Can support templates • Ideally to be used in connection with a content management system
Word editing (with stylesheet)	<ul style="list-style-type: none"> • Environment familiar to most authors • Customisation required in order to assure valid structures • Tendency to concentrate on layout rather than on content
Terminology Management System (=TMS)	<ul style="list-style-type: none"> • Concept oriented formats rather than lexicographic formats • Template based structure editing • Can create simple dictionaries but has its limitations in dictionary processing

Table 1: Tools for structured data capturing

From the underlying concept, UniTerm Pro is closest to a terminology management system (=TMS) since it is a database driven system. All standard functionalities of a database system like indexing, generation of lists, various search strategies, etc. can thus be performed. Distinguishing features to a TMS are its sorting module, its lexicography and its production component. The UniTerm Pro system can be seen in the tradition of the Siemens TEAM System (TEAM = Terminologie Erfassungs- und Auswertungsmethode) which has been widely described in the 70ies / 80ies literature as a model for computer based multilingual terminology. TEAM ran on mainframe computers. It provided a limited set of data fields per language that were processed in a batch mode to produce a typesetting format. This format was handed over to typesetting systems. The dictionary production process was thus reduced from about six months to only a few weeks.

In a similar way as the TEAM system, UniTerm Pro dissolves complex nesting structure into a set of individual entries (one for each sub-entry). The structure of these database entries is easier to understand and to edit. The author establishes/ preserves the connection between entry headword and sub-headwords, i.e. he defines in which entry nest(s) a database entry is to be displayed. One database entry may be assigned to multiple dictionary entries, e.g. a phrasal entry may be assigned to two search entries in the dictionary. Based on this principle the data is processed by subsequent modules that are equally integrated in the authoring tool. Other concepts that UniTerm Pro picked up from the TEAM system and which have been extended for a broader coverage.

Consistency in sorting. *Numerous sorting rules which can be applied to a dictionary. The sorting of sub-headwords within an entry follows predefined rules. Both alphabetical sorting*

and nest-alphabetical sorting are rule-driven and are carried out automatically. So authors do not have to care about intra-nest sorting which has always been a source of error.

Consistency in Layout. Lexicography and production are closely interrelated. Some keywords on these modules: lexicographical markup, automatic tilding and entry nesting, definition of multiple headwords, automatic generation of reference entries, e.g. for abbreviations, synonyms, link control, media specific data processing.

UniTerm Pro produces a generic / flexible typesetting format that can be used with common typesetting engines (e.g. 3B2, Miles Oasys, PageOne). The production module integrated in the authoring tool processes the typesetting format further to generate a RTF (= rich text format which is supported by standard word processors) proof. The RTF proof is the author's preview of the actual dictionary. It differs only in small details (pagination, column titles, hyphenation) from the results of a professional typesetting engine. Authors can thus decide at any point of time in the editing process to generate a dictionary preview for the complete dictionary or for a selection thereof (e.g. one entry letter, a selection of subject fields, etc.).

The authoring environment

In order to make the Beck approach more transparent, an example is given from a running dictionary project for a English-German-English legal dictionary.

Figure 1 shows the authoring environment. The window provides a register window and a second window which serves both for editing and for entry previewing. The register displays different headword lists (here: English nest sorting). The user may switch between different list, i.e. headwords in the languages of the dictionary, a nesting entry view as well as other lists corresponding to indexing defined for the relevant dictionary format.

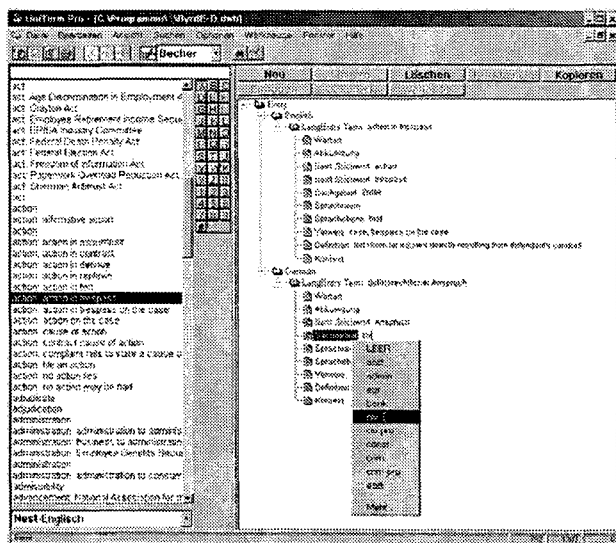


Figure 1: UniTerm Pro authoring environment

The editing window provides an entry template that the author fills in. Term and translation are mandatory fields. Other data fields are optional in an entry / sub-entry. Different types of templates may be defined for different types of entries, for different authors, etc. Here the phrase (action in trespass is further described by a subject field label, a register. It gives a definition and reverts to other entries. The phrase is to be found in the entries action and trespass.

```

action s II:ZivilR.Klage.f.,~in:assumpsit.(hist)
ZivilR.vertragsrechtlicher-Anspruch,~in:
contract ZivilR.vertragsrechtlicher-Anspruch,~
in:detinue (hist) ZivilR,~in:replevin ZivilR,~
in:tort ZivilR.deliktsrechtlicher-Anspruch,
Anspruch aus Delikt,~in:trespass (hist) ZivilR
deliktsrechtlicher-Anspruch: tort claim for
injuries directly resulting from defendant's
conduct → case, trespass on the case,~in:
trespass on the case (hist) ZivilR
deliktsrechtlicher-Anspruch: tort claim for
injuries indirectly caused by defendant's
conduct =: action on the case,~on the case
(hist) ZivilR: deliktsrechtlicher-Anspruch =:
action in trespass on the case.¶
    
```

Figure 2: RTF dictionary preview

Figure 2 shows the corresponding RTF proof for the entry shown in figure 1. Sorting - also nest alphabetic sorting as well as tilding - is done automatically. Pagination and hyphenation is done in the type-setting system used. The generation of the RTF however is done from the XML data that is provided for typesetting systems.

Not all information coded in the database is processed for the dictionary output (e.g. only the German subject field label for the English-German language direction). Dictionary databases can be used for both source-target and for target-source dictionary. With lexicographic markers, mere translations or paraphrases can be excluded from the inverted dictionary. More frequently an initial dictionary database is taken as source material to work on a second database for the target-source dictionary.

Conclusion

The authoring tool introduced at Beck publishers can be seen as a middle course between a standard XML tool and the use of a (terminology) database. It makes use of the advantages of a database (number of new / revised / deleted entries, link checking) but maintain the traditional ties they have to a print product (control structure / microstructure, number of pages).

Already this limited XML approach requires strong guidance of authors in order to train them to work on content without worrying about layout. It is the production environment, the link towards layout which convinces authors that the tool is a help rather than an obstacle

in the authoring process. Only if authors accept using the tool and if its advantages become transparent, subsequent economical and technical goals can be achieved.

References

- Hauptmann, Susanne** 2002. 'Das ideale zweisprachige Rechtswörterbuch - aus der Sicht des Benutzers und aus Sicht des Verlags' in MDÜ - *Mitteilungen für Dolmetscher und Übersetzer* 3/2002. Berlin: Bundesverband für Dolmetscher und Übersetzer.
- Zenk, Wolfgang** 2002. 'Vom Buch zum elektronischen Wörterbuch. Lexikographie aus der Sicht des Herstellers elektronischer Wörterbücher' in MDÜ - *Mitteilungen für Dolmetscher und Übersetzer* 3/2002. Berlin: Bundesverband für Dolmetscher und Übersetzer.