

MT Summit IX Tutorial Proposal

Building the Information Architecture for Controlled Authoring and Translation

In the last decade, the idea of controlled authoring (CA) is discussed within several industries at various levels, but not often implemented, and if so then with quite different success stages. Common to most of the implemented CA applications is that they are technology add-ons to already existing processes. The technologies that are employed range from natural language processing utilities to machine translation. This plug'n'play approach has caused several information handling problems and interfacing problems which led to unsatisfactory integration, disappointed and frustrated users, and the expected results, such as to improve translatability of source texts for machine translation, to cut costs and save time, and to ensure quality and consistency, were not really achieved.

In this tutorial, we analyze the situation and introduce the steps that are necessary for building a success story in controlled authoring and translation. Many examples are taken from our customers who use our product CLAT (Controlled Language Authoring Technology); among them are BMW, Boehringer, DaimlerChrysler, HDM, SEW, Siemens, and Sun Microsystems. The focus of the tutorial however is not on our product, it is on the industry-level stages of definition, realization and deployment of the concept of CA.

Aspects that are included in the presentation concern:

1. Delimiting the complexity of language in terms of grammar, vocabulary meaning and style variations with a sufficient analysis of the information in terms of user, content and context, and considering an appropriate change management of the information in an iterative process.

User – Who will use the information (human and machine)?

Content – What will the information be used to describe (repair and maintenance, end-user)?

Context – Where and how will the information be used (processes and workflows)?

2. Creation of a controlled vocabulary (CV) with sufficient knowledge of the organizational approach (thesauri, taxonomies, etc.) and the impact on other processes or the entire workflow.

What does a preferred term mean, and what are the relations to possible broader terms, narrower terms, variant terms and related terms, and what is the scope of notes that include a definition and provide a context.

For 1. and 2. it is needed to:

- analyze the language that is used,
- build the content inventory,
- examine (existing) indexing,
- understand context.

3. Setting up of review processes and testing procedures which shall avoid lack of quality and to assure quality assurance of the employed technology across releases and updates.
4. Well-defined interfaces for users (human and machine). Reuse of results (either complete analyses or marked up) in other processes (identified terms, sentence boundaries, etc.).
5. Identifying and using appropriate types of meta information. Analysis of how document markup can trigger the analysis process and related processes to permit an efficient and effective control and enrichment of the information (adding value to the information and to the processes).
6. Understanding the technology and its impact for the business to help to
 - manage and distribute the information (distributed information system),

- support document markup,
- improve navigation and search,
- allow for several customization routines including monitoring tools and feedback tools together with focus groups that elicit suggestions for refinements and extensions,
- build and foster a holistic understanding and planned development.

In this tutorial, we will argue in favor of the use of an information architecture that takes into account these aspects which eventually will challenge the full deployment of a complete CA application thereby supporting "excellence in design" and "universal usability" at various levels (user-centered, simple design vs. media-rich, cinematic).

The intended audience should be technical writers, translators, information architects and information managers, as well as decision makers on upper and middle level positions.

Timing: Half day (3 hours) or full day (6 hours)

Media needed: PC projector, white-board or clip-chart

Tutor:

Dr. Jörg Schütz is director and chief software architect of the Institute of Applied Information Sciences (IAI), an R&D organization dedicated to providing multilingual information management solutions using smart language technology combined with advanced information technology.

Jörg was one of the initiator of the German Multilint project and the European Multidoc project, both focussing on taming the language complexity and the language barriers of technical information within a controlled language setting. Meanwhile, the language technology developed in these projects is deployed successfully in different industries ranging from the automotive industry to the software industry.

Jörg is involved in several industrial projects that are concerned with the definition and setup of workflows and standards for information and knowledge management that also comprise exchange formats for lexical and terminological data including ontologies. He has published over hundred technical articles and one book on language technology and language engineering, and has lectured and consulted worldwide. His current interests are the deployment of symbolic authoring and multilingual generation in SGML/XML information and knowledge management environments. He holds a diploma in computer science and a PhD in Machine Translation (MT) and Artificial Intelligence (AI).

email: joerg@iai-sb.de

tel: +49 681 3895 132

fax: +49 681 3895 140

web: <http://www.iai-sb.de>