Working with terminology at Microsoft

Licia Corbolante

"Working with terminology at Microsoft" is a description of the presentation given at tcworld in November 2008 and available in the conference website.

The presentation provides an overview of terminology management at Microsoft[1], describing how it evolved from simple isolated glossaries to its current multilingual, concept-oriented database. It covers the roles and goals of English and target language terminologists in the Microsoft Language Excellence group and illustrates the terminology management workflow within the product development cycle. Terminology-related tools and resources made available to worldwide communities are also described.

Localization is the translation and adaptation for a local market of a software program originally developed in a different market and in a different language, usually English. Terminology management, defined as the investigation, documentation, and consistent reuse of terms and their associated concepts, is a key factor in a successful localization process. In Microsoft, terminology management is owned by Language Excellence, a group of language specialists whose goal is to drive terminology consistency and language quality across Microsoft products worldwide and contribute to efficiencies in product development and the localization life cycle. Language Excellence employs in-house terminologists for English and for ten target languages and retains vendors for the remaining target languages, up to 100 for some products; a team of linguistic engineers develops and maintains terminology-related and language automation tools.

Terminology management has changed considerably over the years, both in the methods used and in the points in the product life cycle in which terminology management takes place. When localization was still in its infancy, basic lists of terms were compiled at a relatively advanced stage of the localization process, often by individual translators and without any involvement from development teams. Over the years such lists evolved into glossaries with definitions, usually created at the beginning of the localization process; glossaries were generally product-specific and not always shared across different product teams. Such reactive, term-oriented approaches, however, had several disadvantages: each product group created its own terminology, different formats were used, there were few mechanisms for sharing terminology, and terminology inconsistencies across products were not unheard of.

---

[1] as in place up to the first quarter of 2009

In the terminology management model currently used by Microsoft, most new terminology is proactively identified before localization starts and made immediately available to localization teams and other users in a multilingual, concept-oriented terminology database. Such proactive process requires fewer resources, is less time-consuming than a reactive process and can lead to considerable cost savings. Additionally, it reduces potential terminology inconsistencies and it can help achieve higher standardization in the target language and, as a result, a better end user experience.

Several players are involved in the terminology management workflow of Microsoft products. Early in the product life cycle, English terminologists work with developers, writers and editors in the product development teams to organize term mining and to identify new concepts and the terms associated to them. Since each product team works independently, English terminologists verify that the terms proposed for a specific product are not duplicates to already existing terminology and concepts. Definitions are then finalized, entries are created in the terminology database and terms are ready for other languages to get involved. Target terminologists start their work by analyzing concepts and source terms to establish which of their aspects should be prioritized in the research process – a different focus might be required depending on whether the term is a neologism or an established term, whether it is product-specific or generic, which other terms are related to it within its concept system, etc. Terminologists decide on the most suitable approach for their language, carry out the necessary research, evaluate the results, if needed with additional input from a network of subject matter experts, and decide on the best fit. Once the final target terms are in the terminology database, localization can start.

Terminology management and language automation tools are used in each stage of the workflow, for example, initial term capture is performed with term mining tools. TermStudio, a concept-oriented terminology database, is the main terminology management tool used at Microsoft; each entry in it represents a concept and lists the different lexical realizations (synonyms, acronyms and abbreviated forms) that the concept can take in the source language and in the target languages each product is localized into. Hierarchical and associative relationships between concepts can be made explicit; each terminological entry contains relevant metadata, such as part of speech, number, gender, product name(s) and version, term type, term usage note, context, visual context, etc. The data in TermStudio are complemented by a database that contains all software strings of current versions of Microsoft products. Other tools are also used, e.g. to detect terminology inconsistencies in software strings and to validate additional terminology that was not captured upfront but found during the localization process.

Community engagement is an additional aspect of terminology work which has acquired great relevance. In early 2008 Microsoft made language-related tools and terminology resources

available to worldwide communities by launching the [Microsoft Language Portal](#), a multilingual site where visitors can search terminology in over 90 languages, download style guides for more than 50 languages, provide feedback on terms used in Microsoft products, find articles on specific aspects of terminology work at Microsoft and access terminology blogs in ten different languages. The terminology search tool provides access to over 15,000 TermStudio entries and up to two million product software strings for each of the main target languages. In addition, Microsoft regularly runs [Microsoft Terminology Community Forum](#) initiatives to give users the opportunity to review and provide feedback on selected Microsoft terms and help Microsoft ensure terminology closely reflects the local culture and is accepted by the communities who use the products.

In summary, over several years Microsoft has developed a set of tools and processes to manage terminology proactively in a growing number of languages; more recently, Microsoft has also recognized the importance of making that terminology available to worldwide communities and providing a way to give feedback on it. The current concept-oriented terminology approach has proven to have clear benefits in terms of efficiency, consistency, and quality, but there is still work to do. The Language Excellence team will continue to improve its terminology management system and share information about it in the Microsoft Language Portal, and welcomes feedback from language professionals who share its passion for IT terminology.