PILLS: Addressing the challenge of localisation in global pharma publishing

Whether you call them consumers, customers or patients - users of pharmaceutical products increasingly demand direct access to information. At the same time the globalisation of the pharma industry - and of strategies for marketing specific products - creates a much wider range of audiences for product information. Just as pharma companies begin to get control of their communication with home markets, they are required to replicate their communication strategies across cultures and languages in the global market, particularly on the Web.

Recognizing the significant challenge of multilingual publishing for global pharma companies, the European Commission has funded a research project to explore the potential for authoring pharma content using a concept-based authoring tool that will enable publishers to generate product information in multiple languages. In a joint project between Berlitz GlobalNET, the University of Brighton in the UK, and the University of Freiburg in Germany, PILLS is developing a prototype tool to enable pharma companies to author product information that can be published simultaneously in multiple languages, for multiple formats, in multiple styles, all from a single source.

A matter of "style"

Researchers at the Information Technology Research Institute (ITRI), part of the University of Brighton, realized that Web publishing has brought a major innovation to the management of content; for the first time it is not only desirable, but actually necessary for information owners to create and maintain all their content electronically. This creates the opportunity to exploit the well-established principle in electronic publishing that "content" and "style" should be managed separately. Using XML (eXtended Markup Language), it is possible to take the same content elements - a list of contraindications, for example - and publish them in one format for a patient leaflet, another for a consumer Web site, and yet another for a technical audience such as physicians or pharmacists.

In an XML environment, the "style" of a document is defined quite separately from the "content". This structured approach to content publishing obviously has many advantages over the creation and maintenance of multiple individual forms of the same document. But it only addresses layout and does not handle more subtle aspects of style, such as the choice of vocabulary or the complexity of syntax. In pharma publishing these stylistic issues can be critical, where for example the use of a technical term is required for professionals, while a non-technical term for the same concept should be used for consumers.

From this perspective, pharma publishers are perhaps unique in that they are required by regulation to publish specific types of documents, but the nature of these documents makes it difficult to exploit the economies of scale of re-usable information elements. Information about drugs must be published for regulators, physicians, pharmacists, purchasing agents, patients, customers, consumers, etc. - in different stylistic registers and formats, using different terminology and vocabulary - and increasingly, in multiple languages as well. The financial and administrative burden will become intolerable as the globalisation of the industry advances.
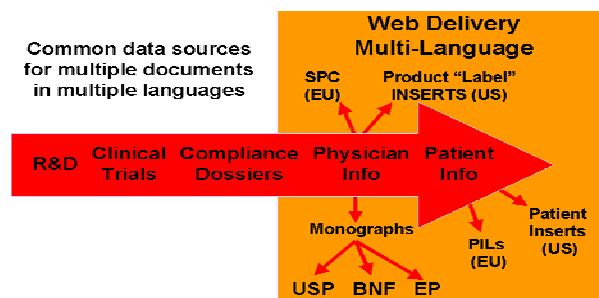
Pharma Publishing Info Flows:
The unexploited opportunity to leverage content by re-using common sources of data and information

WYSIWYM©

The ITRI team at Brighton first addressed the question of pharma publishing by developing an editor designed to produce Patient Information Leaflets (PILs) with more subtle and flexible publishing options. The editor, called WYSIWYM (What You See Is What You Meant) is an intelligent menu-driven system which allows the author to define meaning, style and presentation features in order to generate suitable text in multiple languages.

In Germany, the Universitätsklinikum at the University of Freiburg are developing a pharmaceutical knowledge base or "domain model" by extracting and editing data from the Unified Medical Language System (UMLS), a set of vocabularies and language resources supplied by the US National Library of Medicine. These basic concepts are being encoded and integrated with ITRI's Natural Language Generation (NLG) technology, which uses a formal language (a kind of encoding), rather than a natural language such as English or French, to represent the information content of a document, Each "formula" has only one, precise meaning but can be represented textually using different vocabulary depending on the style required or the language of the reader.

The PILLS project will demonstrate the feasibility of combining a domain-specific knowledge base with the WYSIWYM© Editor to enable multilingual text generation in the pharmaceutical domain. The application will be designed with the potential to support various types of pharma publishers, including product developers and manufacturers, health portal publishers and healthcare e-marketplaces.

What's next for PILLS

The PILLS project, which continues through the end of 2001, will:

- Validate the need for multilingual authoring and publishing, and the feasibility of a knowledge-based approach in the medical/pharmaceutical domain using the WYSIWYM Editor

- Assess the viability and cost-effectiveness of multilingual authoring as a localisation strategy, as an alternative to translation

- Document the pharmaceutical publishing lifecycle, the requirements for localisation, the impact of regulation and harmonisation, and opportunities for content reusability in the domain

- Evaluate the impact multilingual authoring could have on the digital delivery of information in the healthcare field (availability to consumers, standardisation of content, potential for supporting new business models)

- Develop a prototype system to illustrate and verify the feasibility of the multilingual publishing concept

Berlitz GlobalNET is responsible for market research, and validation of results with representatives from the industry who will be invited to test-drive the PILLS prototype. Any one interested in participating should contact the project coordinator: Rose Lockwood, Director of Research, Berlitz GlobalNET: rose.lockwood@equipe.co.uk.

Multilingual Publishing with PILLS
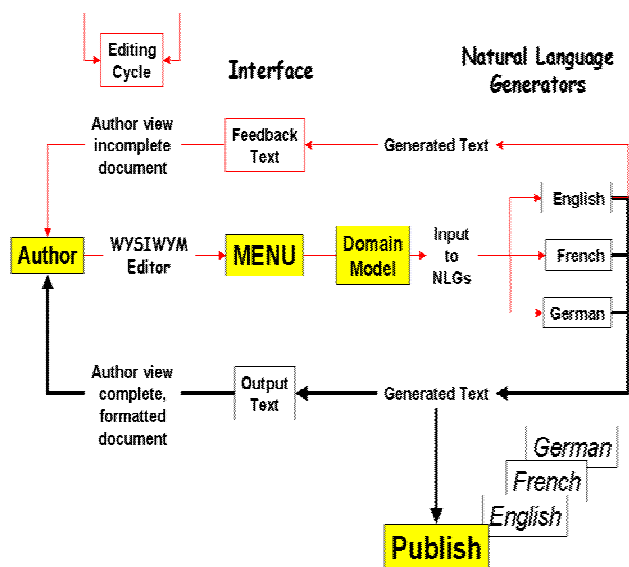
1. The author creates a document, e.g. a Patient Information Leaflet using point-and-click selections offered by the editor from the underlying domain model.

2. As each required element is selected this triggers a link to the natural language generators that produce the text in a natural language, e.g. French.

3. The "Feedback Text" allows the author to check that the content being generated is correct for the purposes of the document.

4. When all required elements have been specified the author can see the completed document as "Output Text".

5. Output Text can be published to multiple formats using XML tagging routines or other publishing standards.

6. Once a correct Output Text has been created, PILLS can generate the content to any language for which a Natural Language Generator exists.