Z39.50 and multi-national/multi-lingual environments

Makx Dekkers, Pica (Netherlands)

As a result of the possibilities offered by network technologies of the past decades, exchange of information via computer has become an every-day phenomenon in today's world. However there are problems in multi-national and multi-lingual environments that are not always obvious in the U.S. These problems are often related to cultural differences: national bodies are responsible for national rules for bibliographic descriptions; more fundamentally, different countries use different languages with different character sets and sorting rules. Other problems include: different rules for conversion from 8- to 7-bit ASCII for indexing dependent on country and language; difficulties translating system messages, variety of formats, and differences in cataloguing rules between countries.

An illustrative example of the latter problem is the treatment of multi-volume publications: under some rules these are catalogued as one single record with repeated elements, under other rules they are described as separate entities with relations between them. Another example is the use of standard phrases in national language within the cataloguing rules, such as for title changes for journals (in Dutch cataloguing the description would contain the phrase "Voortgezet als:"). International exchange of such bibliographic descriptions would ideally involve automatic translation; however, this is not done in practice.

For searching, a major problem is that standardised keyword lists are usually defined in national language and subject code systems are also agreed in a national context. All these different language and country related rules and practices cause incompatibilities that are difficult or sometimes impossible to overcome.

A number of problems are associated with differences in character sets. Many scripts are used in the world and most existing library systems are unable to handle them all. Transcription rules or character set conversions sometimes lose information as are not always reversible.

In sorting the situation is even more complicated. Where the same character set is used in two languages, sorting order might be different. In some languages "o-umlaut" is sorted as "oe", in others it might appear at the end of the alphabet. Even when the same language is used in two countries, there might be differences in sorting order of names: in Belgium a personal name of "Van Dam" will appear under "V", in the Netherlands under "D".

Solutions

Solving these problems requires the introduction of negotiations, on-the-fly conversions, and powerful explanation techniques. Negotiation aims at establishing a mutually agreed environment allowing exchanged data to be converted from one format to another. If that is impossible, the user should receive information to explain why the result is not as expected and to suggest alternative actions.

In the latest published version of the Z39.50 standard, Z39.50-1995, mechanisms are incorporated to negotiate the use of character sets as well as language. This is a big improvement compared to the 1992 version of the Z39.50 standard. Z39.50 now supports multi-lingual systems. Character sets that can be used are ISO 10646 and ISO 2022, or mutually agreed private character sets. A client/server pair may agree to the languages to be used for server message (including diagnostics) intended for display to a user.

For data formats, it is clear that national or local rules will prevail to determine how information is stored in databases. A positive development in some European projects is that implementors are trying to build table-driven, public domain toolkits, both for format conversions, as well as character set conversions. Although 100% accuracy in conversion cannot be achieved, this might help in broadening the scope of Z39.50 interoperability.

In areas where negotiation or conversions cannot solve the problems, the use of the Explain facilities defined in the Z39.50 standard provide the solution. This facility is probably the most powerful feature of Z39.50. Through Explain, the user is given information to understand better what goes on behind the scenes and to allow him to make sense of the results of certain actions. Fortunately, all messages in Explain have been designed for multi-lingual environments.

In conclusion, Z39.50 provides a very useful tool for information retrieval but it is clear that differences in language and culture have an impact on its scope and usefulness in international contexts. Internationalisation of the standard has solved some of the problems. Hopefully, through the implementation and use of Explain some of the others can be explained to users. The aim should be to make it possible to provide services to a wide international audience, respecting the multitude of cultures and languages in the world.