

Workshop 4

Cross lingual applications

Report prepared by Patrice Landry

Participants:

Badr, Aly
Hakala, Juha
Hoppenbrouwers, Jeroen
Kangas, Sinikka
Landry, Patrice
Murray, Neil
Saloranta, Pertti
Walther, Frédéric

This workshop investigated how multilingual access can be achieved. In using as example how could case laws or community law documents could be accessed through a multilingual access, it was decided that a "Roadmap for multilingual applications" be established.

Roadmap for multilingual applications:

Basic conditions: Mostly technical problems

- Automation system (management and access software?)
- Script conversion and display (UNICODE?)

1. Recognition of requirements: Identify the access needs of the intended users

- Identify existing solutions, possible resources
- Announce concept / project so that others do not start duplicating the work, or can warn you that you are about to start duplicating the work.
- Assign work based on commitment

2. Demarcation of conceptual borders

- Conceptual heritage (ex.: Civil law)

3. Demarcation of linguistic borders

- Vocabulary and syntax: controlled vocabularies are different from natural languages and also between different controlled vocabularies.

Steps 2 and 3: identify differences and similarities and organize (classify) the family of languages and conceptual borders (ex. Laws from different cultures and languages).

Milestone 1: This is the end of the first stage: Resources can be evaluated and planned. The work can be organised. Decisions can be taken (what to do).

4. Recognition of resources and assigning points 2 and 3 to them: Resources are targeted to a particular cell that represents a language and / or culture. Resources can be defined as: system, data and human.

Work matrix

	French	German	Br. English	U.S. English	Classification concept
Law 1					
Law 2					
Law 3					

5. The appropriate resource (machine, data, human) makes the bridges (links) between two cells. This is a two way process. In principle, we need two specialists to validate both ends of a link. There is a preferred order in deploying resources: automation, existing data, before human. The final validation should be carried out by human experts.

6. Bridges are used to establish more bridges: once a bridge is done, it leads to another bridge, the more bridges we have, the easier it is to add new bridges (network effect).

Steps 5 and 6 constitute the editorial process of the multilingual applications process, it will run until either money or time runs out. It will never be finished, just paused

Milestone (step 2): System delivery: It is entirely possible to deliver the system with the editorial loop still running. There is no need to issue releases. Consider feedback into the editorial loop, implicit (tracing) or explicit (participation).

7. Bridges are used to learn about differences and equivalents. This is the user search process. This step should loop into the editorial process to take in consideration feedbacks from users.

8. Consolidation and publication of core concepts network with the aim of connecting to other semantic domains using standard generally recognized syntax (Semantic Web).