Consistency Management in Parallel Multilingual Documents

Amit Pariyar
Department of Social Informatics, Kyoto University
amit@ai.soc.i.kyoto-u.ac.jp

ABSTRACT
One of the aspects of managing multilingual knowledge resources is managing multilingual documents shared with the multilingual communities. Managing inconsistencies in multilingual documents refers to managing unavailability of same information in all languages, missing information or part of document not translated, changes in document in one language not being propagated in other languages. In this research we propose a state transition model to define the state, action, state transition of the sentences in multilingual documents and define consistency checking with consistency rules to signal inconsistencies which is resolved with consistency handling.

Keywords: Multilingual document, inconsistencies, state transition, consistency rules.

INTRODUCTION
The multilingual documents related to safety guidelines, disaster related manuals; product user manuals, diplomatic documents are examples of some multilingual documents that often require consistent information in all its language versions. Such multilingual documents produced in several languages are the translation of some source document and are collectively called as Parallel Multilingual Document [1]. As these documents become available to the communities it becomes difficult to guarantee the consistency of information in all the language versions of the document due to modification of the documents by the respective communities.

Inconsistencies in the multilingual context is characterized by the unavailability of same information in all languages, missing information or part of the document not translated, documents not up to date in all languages, changes in the document in one language not being propagated in other languages[2]. As mostly discussed in consistency management of documents, consistency checking and consistency handling is required for signaling inconsistencies whenever it occurs in the document and resolving inconsistencies [3][4][5]. In doing so, we also want to facilitate the task of locating the inconsistent portions of the document and maintaining consistency of information in the multilingual document.

MECHANISM
We extend the logical data structure holding the exact correspondences of the segments in the parallel multilingual document presented in [1] with information about the states of the sentences. As observed in [6] the level of alignment i.e. consistency of information in the multilingual document is maintained at sentence level with parallel aligned sentences in multiple languages such that corresponding translation of sentences in multiple languages is available in all the documents. Our mechanism deals with maintaining consistent information in multiple languages at the sentence level.

State Transition Model
The state transition model can be described as a tuple: 
M= (Q,∑,δ,qo,qf), where Q = {S,T,I,M} is the set of states of the sentences. ∑= {add, delete, update, mdf_content, mdf_quality , translate, ϵ} is the set of actions. qo, qf are initial and final states of sentence.

A sentence in Source state(S) holds the latest information to be translated into other language, Translated state(T) holds the translated content from parallel aligned sentence, Intermediate state(I) if the modifications are not saved and Modified state (M) if the content modified requires specifying the modification purpose to determine if it is a new source.

The action mdf_content checks modified content with added or deleted facts; mdf_quality checks for modified content with quality improvements e.g. translation errors; add/delete for adding and deleting contents; update for saving contents; translate for translating contents. Automated action ϵ is triggered to establish the dependency relations between the source of content and translated content. δ is the state

Fig.1 State Transition diagram of a sentence
transition function $\delta(q, x) = q'$ refers to the change in state from $q$ to $q'$ for some input action $x$. Fig 1 presents the state transition diagram for a sentence.

**Consistency Checking and Handling**

Using the state transition model we define consistency rules to describe the relation between the states of the sentences in the multilingual document that must hold for consistency of information.

**Translation Consistency rules** govern the translation relation of multilingual contents in the multilingual documents and also represents the state transition for which the content needs to be translated. The rules are:

1. The multilingual content in Source state is a source for translation in multiple languages.
2. The state transition with final state as Source state resulting from the content modification (addition or deletion of facts or information) in the sentence has to be translated into corresponding languages.

**Transition Consistency rules** governs the relation between the state of sentence for state transition during modification and identifies the unstable states leading to inconsistency. The rules are:

3. Initial and final state of the sentence is either Source or Translated state.
4. A sentence as Source state in multiple language is unstable as there is only a single source of content.
5. As a sentence becomes a new source and is translated, the state of corresponding parallel aligned sentences changes to a Translated state with automated action $\epsilon$ to maintain dependency relations.
6. The sentence in Intermediate state is unstable as the content is not yet saved.
7. A Modified state is unstable and requires checking the purpose of modification.
8. The modified content with action mdf_content becomes the source and needs to be translated and the modified content with action mdf_quality do not need translation.

Consistency checking uses both translation and transition consistency rules to identify the state of the sentences that violate these rules and signal inconsistencies whenever it occurs in the document. Consistency Handling provides the resolution actions for inconsistencies.

**References**