

# Consistency Analysis in Multi-language Knowledge Sharing System

Amit Pariyar, Yohei Murakami, Donghui Lin and Toru Ishida

**Abstract** Unprecedented growth in knowledge sharing among multi-language communities, both common and distinct languages, has raised the possibility of sharing inconsistent content. Though popular with traditional system, the approach to explicitly state consistency rules to avoid inconsistency is practically not suited for multi-language knowledge sharing system because of sheer complexity. Alternatively this chapter focuses on potential cause of inconsistency, cases such as *content omitted*, *content updates not propagated* and *content conflicts*. Ignoring such cases in knowledge sharing has undesirable consequences: *community bias*, *global and local inconsistency* and *regional discrepancies*. Consistency constraints from opposing knowledge sharing goals among communities is another issue. Due to which consistency policy ranges from rigid ‘one to one consistency’ to non-rigid ‘consistency where needed’. This chapter contributes with (a) process-based approach for multi-lingual content synchronization to leverage knowledge equally and (b) propagation-based approach to analyze community preferences when sharing specific content categories/geographic regions, to customize knowledge sharing; a value add-on to designing language services adhering to knowledge sharing goals.

**Keywords** Inconsistency · Multi-language · Knowledge sharing · Content category · Geographic region

---

A. Pariyar (✉)

Institute of Social Informatics and Technological Innovations,  
University Malaysia Sarawak, Sarawak, Malaysia  
e-mail: amitpariyar@gmail.com; pamit@animas.my

Y. Murakami

Unit of Design, Kyoto University,  
91 Chudoji Awata-cho, Kyoto 600-8815, Japan  
e-mail: yohei@i.kyoto-u.ac.jp

D. Lin · T. Ishida

Department of Social Informatics, Kyoto University,  
Kyoto 606-8501, Japan  
e-mail: lindh@i.kyoto-u.ac.jp

T. Ishida

e-mail: ishida@i.kyoto-u.ac.jp

© Springer Nature Singapore Pte Ltd. 2018

Y. Murakami et al. (eds.), *Services Computing for Language Resources*,  
Cognitive Technologies, [https://doi.org/10.1007/978-981-10-7793-7\\_9](https://doi.org/10.1007/978-981-10-7793-7_9)

## 1 Introduction

Recent advances in online community participation has accelerated knowledge sharing beyond human imagination. Wikipedia aptly represents such a prodigy in action serving as a storehouse for millions of articles contributed by 200+ language communities. Another evolving community are global brands publishing country-specific websites to represent 100+ countries exceeding 40 languages. The implication of such emergent multi-language knowledge sharing communities is the publication and sharing of massive content in several languages. The daunting and challenging task is to manage inconsistency in the shared content, both *common* as well as *distinct* languages. Logical and Factual inconsistency widely studied in database and knowledge-based system are also equally anticipated in multi-language knowledge sharing system (MLKSS). Though approaches to explicitly state consistency rules as *foreign key constraints* or *using logical formulae* are popular with past systems, it is not practically viable solution for MLKSS due to complexity with growing number of languages. The alternative approach adopted in this chapter is to highlight cases that are expected to cause inconsistency in logic and fact as community share content, analyze their consequences and design consistency rules.

Cases such as *content omitted*, *content updates not shared* and *content conflicts* are expected to occur during collaboration. The cases may seem trivial at first, nonetheless the complexity to handle inconsistency is raised when they are dispersed across several languages. Some consequences are (a) *community bias with one community preferred over another* (b) *inconsistency at scales leading to globally and locally shared inconsistent content* and (c) *regional discrepancies*. MLKSS should address such issues to enable consistent knowledge sharing. Another issue is the constraint in content consistency meaning '*opposing views to support consistency*'. Knowledge sharing goals of communities vary along the continuum, where one end puts emphasis on leveraging knowledge equally while other end supports customized knowledge sharing to cater to community preferences. For example, documents such as product manual, technical specification are usually produced with the intention to share same information in several language editions. A '*rigid consistency policy*' is a better match to enforce one-to-one correspondences in multiple languages. The growing cultural homogeneity among communities with 'one size fits all' notion or legislative rules could have demanded such a stern policy. On the contrary, the persistence of cultural difference among communities widely valued in past studies stresses on relevant knowledge shared to suit specific community preferences. Knowledge sharing is viewed as not uniform among communities and exact correspondences in the shared content is not always preferred. For example, the need to restrict publication and description of content in specific languages in country-specific websites makes '*non-rigid consistency policy*' a better choice. Such opposing views has to be supported in the design of MLKSS. Consistency analysis will give a context to design language services adhering to knowledge sharing goals.

To avoid *undesirable consequences of inconsistency* in knowledge sharing and to allow *consistency constraints* resulting from divergent knowledge sharing goals

among communities, this chapter contributes in the design of MLKSS to support content consistency. Section 2 illustrates an example of inconsistency revealing its causes and consequences. Section 3 details approach to leverage knowledge equally. Section 4 illustrates underlying community preferences for customized knowledge sharing. Section 5 discusses and Sect. 6 concludes this chapter.

## 2 Illustrating Cause and Consequences

We refer to corporate related content ‘3M at a glance’ of a global brand 3M, published in its country-specific websites for Switzerland, Canada, United States, France, India and Australia. Note that Switzerland and Canada have multiple official languages; ‘French’ is a common language among France, Canada and Switzerland. Websites also represent geographic regions: North America, Europe and Asia Pacific. Following observations are compiled.

**Cause.** As shown in Fig. 1, the information about ‘product donations’ for the year ‘2013’ is omitted in ‘French’ version of websites for France, Switzerland and Canada. The latest information related to ‘2014’ is not propagated to most of the websites except for India and United States. This also means that content updates are not available in languages ‘French’, ‘Deutsch’ and even ‘English’. The information about

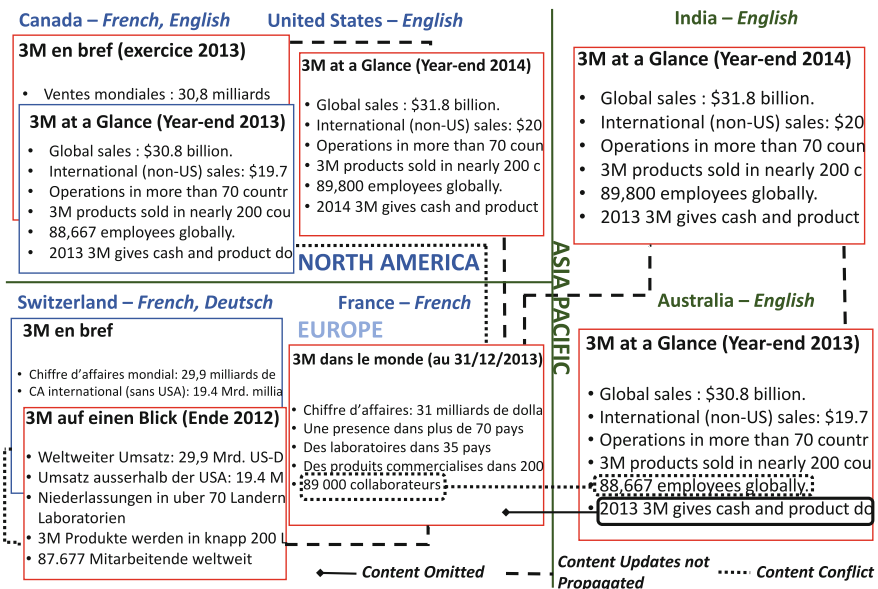


Fig. 1 Causes of inconsistency among country specific websites

‘number of employees’, ‘global sales’ also appears to conflict among France, Canada, Australia and Switzerland. This means conflict among languages. We compiled cases as *content omitted*, *content updates not propagated* and *content conflicts* as potential cause for inconsistency. Next we show consequences.

**Consequences.** The presence of cases *content omitted*, *content updates not propagated* and *content conflicts* in knowledge sharing have following consequences.

- (a) **Community Bias.** The delay in the simultaneous release of content updates among communities creates bias, as one community or language is assumed to be prioritized over another community. In this example, ‘English’ seems to be a language choice and India and United States seems to be country choice in sharing content. Within Canada, it seems more information is in ‘English’ compared to ‘French’ language, room for bias inside a country.
- (b) **Global and Local Inconsistency.** Content updates for the year ‘2014’ published by websites for India and United States is not shared with remaining countries in multiple languages ‘French’ and ‘Deutsch’. Also not shared in common language ‘English’ with Canada and Australia. Failing to propagate content updates either in common and distinct languages among countries gives rise to globally shared inconsistent content. Content conflict between languages ‘Deutsch’ and ‘French’ for information on ‘statistics for the number of employees’ within Switzerland gives rise to local inconsistency.
- (c) **Intra and Inter Regional Discrepancies.** Intra-regional discrepancies occur among countries inside same geographic region such as Asia Pacific (India and Australia) and North America (United States and Canada) as updates for ‘2014’ is not propagated. The statistics in ‘global sales’ and ‘number of employees’ offered are also conflicting in France, Canada and Australia leading to inter-regional discrepancies.

We are motivated to design MLKSS to address community bias, global and local inconsistencies and regional discrepancies. Next we will detail approaches to support consistency for specific knowledge sharing goals.

### 3 Leverage Knowledge Equally

Multilingual correspondences is achieved when content updates are allowed to propagate consistently across languages. Language neutral representation is used in [3] to automate generation of consistent multilingual instruction. Since technical skill is required to modify underlying knowledge representation to amend changes, its use is limited to domain experts. Cosyne [9] on the other hand uses language processing with state of art machine translation, concept network, cross-lingual entailment to pinpoint differences and overlapping among languages. Ziggurat [1] another automated system uses self-supervised learning to align info boxes in multilingual Wikipedia articles. Both systems support resource rich languages mostly to

European languages and replicating them to resource poor language is not practical due to limited linguistic resources. Restructuring multilingual correspondences with MLHTML [14] and alignment tools [2] suffers management overhead to manage central representation in order to ensure consistent updates propagation. The collaborative wiki style translation in [5] is also inadequate in highlighting specific inconsistent cases and keeping track of language from which the information originates. Referring to limits of past studies, particularly inadequate support for resource poor languages, the goal of this section is to leverage knowledge equally in variety of languages and communities. The approach to detect inconsistency in multilingual content is presented next.

### 3.1 Process-Based Approach

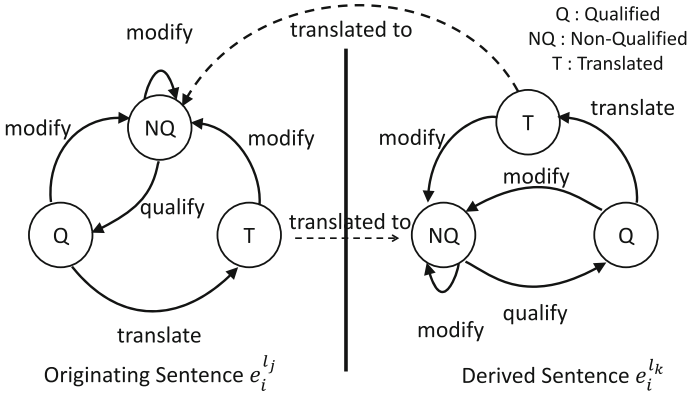
Our work is based on the concept of *synchronizing user editing activities* and detect inconsistency as it occurs in the process of creating a multilingual document. For this we extend multilingual correspondences structure in [2] by augmenting information about states of parallel aligned content, keep track of their modification and employ rules to detect inconsistency.

**Notation.** A Monolingual Document  $d^l$  is the document with the content available in language  $l$ . A sentence  $e_i^l$  in the document  $d^l$  is the  $i^{th}$  sentence in language  $l$ . Content in monolingual document are organized into a collection of sentences  $d^l = \{e_i^l \mid 1 \leq i \leq n\}$ . If  $L$  is the set of languages used in the multilingual document then parallel multilingual document is the collection of several monolingual documents  $D^L = \{d^l \mid l \in L\}$ . With this granularity we will focus on consistency of multilingual content at a sentence level. We refer to [4] for basic concepts in automata theory. Next we will present state transition model to define states, actions and state transitions. Then we will define inconsistency detection rules to check inconsistent states.

#### 3.1.1 State Transition Model

The state transition model is described as a tuple:  $M = (S, \Sigma, \delta, S_0)$  where, (1)  $S = \{Q, NQ, T\}$  is the set of states of sentences corresponding to Qualified, Non-Qualified and Translated states respectively.  $S_0 = \{Q, NQ\}$  is the set of initial states, (2)  $\Sigma = \{\text{modify, qualify, translate}\}$  is the set of actions performed on sentences and (3)  $\delta$  is the state transition function given by  $\delta : S \times \Sigma \rightarrow S$ .

**States.** To define states  $S$  in multilingual content we need to consider (i) relation of content originating in one language with content derived from translation in another language and (ii) change in relation as the content is modified with either contextual changes (addition or deletion of facts or information) or surficial changes (e.g. paraphrasing the text).



**Fig. 2** State transition diagram of a parallel aligned sentence  $(e_i^{l_j}, e_i^{l_k})$

- **Qualified:** A sentence  $e_i^l$  in the multilingual document is said to be in Qualified state Q if the sentence holds updated facts or new information. Such sentence is eligible for translation to other languages.
- **Non-Qualified:** A sentence  $e_i^l$  in the multilingual document is said to be in Non-Qualified state NQ if the sentence holds paraphrased text, grammatical corrections or derived information from another language. Such sentence do not require translation.
- **Translated:** A sentence  $e_i^l$  in the multilingual document is said to be in Translated state T if the sentence is translated into another language.

**Transition Function.** The state transition diagram of parallel aligned sentence  $e_i^{l_j}$  (originating) and  $e_i^{l_k}$  (derived) in Fig. 2 shows the change in states for actions corresponding to *modify*, *qualify* and *translate*. Multilingual Document  $d^{l_j}$  and  $d^{l_k}$  have several such parallel aligned sentences and information about their states.

### 3.1.2 Inconsistency Detection Rules

To design rules we will use the combination of states between parallel aligned sentences in documents,  $d^{l_j} = \{e_i^{l_j} \mid 1 \leq i \leq n\}$  and  $d^{l_k} = \{e_i^{l_k} \mid 1 \leq i \leq n\}$ . Table 1 illustrates case of interest *content omitted*, *content updates not propagated* and *content conflict* represented as rules using state combination. For example the presence of Qualified Q states in both parallel aligned sentences  $e_i^{l_j}$  and  $e_i^{l_k}$  corresponds to content conflict as both sentence holds updated information. The rules presented here are naturally extended in the case of parallel multilingual documents ( $|L| \geq 2$ ) meaning more than two languages. The tabular representation of aligned sentences and their states in [10] highlights the ease in tracking inconsistencies.

**Table 1** Inconsistency detection rules

Rules description for inconsistent cases	State combination of parallel aligned sentence ( $e_i^{l_j}, e_i^{l_k}$ )
(i) <b>Content Omitted</b> $\forall e_i^{l_j}, e_i^{l_k} : isCreated(e_i^{l_j}) \Rightarrow isInconsistent(e_i^{l_j}, e_i^{l_k})$	(Q, $\_$ )( $\_$ , Q)
(ii) <b>Content Updated not Propagated</b> $\forall e_i^{l_j}, e_i^{l_k} : isStateOf(e_i^{l_j}, Q) \wedge (isStateOf(e_i^{l_k}, NQ) \vee isStateOf(e_i^{l_k}, T)) \Rightarrow isInconsistent(e_i^{l_k}, e_i^{l_j})$	(Q, NQ)(NQ, Q) (Q, T)(T, Q)
(iii) <b>Content Conflicts</b> $\forall e_i^{l_j}, e_i^{l_k} : isStateOf(e_i^{l_j}, Q) \wedge isStateOf(e_i^{l_k}, Q) \Rightarrow isInconsistent(e_i^{l_j}, e_i^{l_k})$	(Q, Q)

### 3.2 Experimental Evaluation

**Setup.** We referred to edit histories of multilingual articles titled “2013 ICC World Cricket League Division Three” (referred as Article 1) and “2014 ICC World Twenty20” (referred as Article 2) available in English and Nepali languages. We extracted 71 parallel contents in Article 1 from May 4 to May 9 and 72 parallel contents in Article 2 from March 16 to April 6, the duration of tournament. Content directly copied and appearing as English text in Nepali articles are ignored. We then labeled 30 modification actions from Wikipedia Edit Summary Legend as qualifying and non-qualifying modification.

**Evaluation.** We compared inconsistencies detected applying the proposed technique with inconsistencies identified from manual inspection to compute precision and recall. For Article 1 overall precision of 94% and recall of 85% is achieved. Precision is higher due to detection of majority of missing contents in Nepali article. The missing content (*matches between Nepal-America, Nepal-Uganda, Nepal-Oman* in Nepali article, Revision Id: 337549) is detected as missing content in English article. For Article 2 the overall precision achieved is 82% and the recall is 87%. Inconsistency between content “Round 1 Group B” in English Article (Revision Id: 600298773) and Nepali article (Revision Id: 384275) is detected as updated content (*match entries to Netherland vs. Zimbabwe*) is not propagated to Nepali article. However, the decrease in precision for Article 2 accounts from the absence of content processing involved in checking semantic relatedness in parallel content. Content conflict due to updating same content in both languages is also detected. The content (*entries for score points for Nepal*) in English article (Revision Id: 600444676) and Nepali article (Revision Id: 384304) is detected as conflicting content and hence inconsistent. With the proposed technique for detecting inconsistency in the selected articles, we find an average precision of 88% and recall of 86% which is satisfactory in detecting inconsistency given that only user editing actions are used.

The result can be improved if integrated with NLP to apply semantics to confirm content consistencies. Towards leveraging knowledge equally the proposed approach requires minimal language processing and hence support resource poor languages. When supplemented with advanced NLP techniques the accuracy can be improved. Next section will highlight knowledge sharing involving community preferences.

## 4 Customize Knowledge Sharing

Several managerial strategies are adopted by organization in their business processes, content management and so on. Centralization is exercised to control and dictate business activities from headquarter office. Knowledge sharing is unidirectional where what content is to be published for a certain market, what is to be translated, what not to be translated are all decided by the central authority. There are ample chances to lose relevance yet the main focus is to create unified brand presence worldwide. Decentralization on the other hand encourages country offices to independently execute their business activities. Inconsistent branding, fragmented localization, inappropriate content published in the absence of well-defined guidelines are some problems associated with knowledge sharing. Hybrid strategy ensures brand preservation while country offices develop local programs that complies with corporate goal and standards. From a technical standpoint, collaborative tools that support global consistency and local flexibility is required. This is where the need to customize knowledge sharing is crucial so that content consistency is imposed only where relevant. Customization is also attributed to cultural [13] and non-cultural [6] differences when sharing content features such as corporate information, communication/customer support, financial information and so on in websites. Past studies have also depicted cultural differences among geographic regions in the use of instant messaging [8] and stimuli to website effectiveness. This raise an important concern whether there exists preferences when sharing content categories with specific regions. This section will explore the relation of content categories and their scope in publication to specific geographic regions.

### 4.1 Propagation Based Approach

Our work is based on the concept of *analyzing propagation of content* occurring among communities during knowledge sharing and use it to determine their preferences when sharing specific content categories and to specific geographic regions. We then infer information about ‘scales’ and ‘coupling’ in sharing to generalize required content consistency policy. Country-specific websites managed by global brands is an ideal example of cross country propagation in knowledge sharing. The managerial challenge in such knowledge sharing remains difficulty in propagating content updates where required, which leads to inconsistent cross-site content. By



comparing content in webpages belonging to country-specific websites and analyzing their propagation allows us to understand preferences when sharing specific content categories and for specific geographic regions.

**Propagation.** Propagation is said to occur between websites  $w_1$  and  $w_2$  managed in a global brand if webpages  $p_a \in w_1$  and  $p_b \in w_2$  have exactly same or comparable content. This literally means comparing webpages to check whether exactly same text or comparable (paraphrased text with same information) or completely different text appears between websites. Since comparable content has to be checked between webpages, manual effort is needed to examine their propagation among websites which means existing text-based method cannot be applied. We base this notion of propagation in network of websites and their affiliated regions.

#### 4.1.1 Website Graph and Website Pair

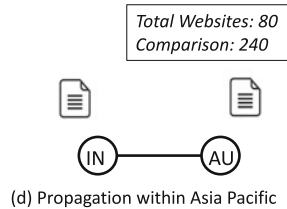
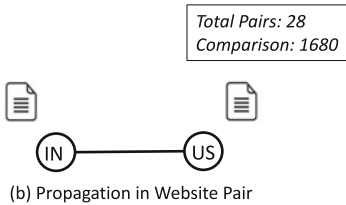
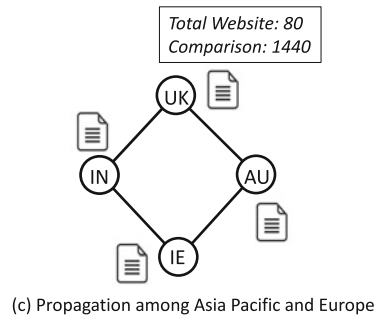
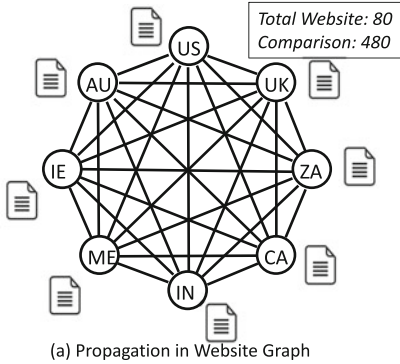
Website graph is a structure that interconnects all country-specific websites. The concept here is to examine propagation occurring to and from all interconnected websites. If timestamp information is available it can be used to assign the source website producing content and follow its propagation to remaining websites. Else we consider each country-specific website as a potential source for publishing content in a webpage shared with the remaining websites, as in this chapter. Propagation in all interconnected potential sources websites is represented as a website graph in Fig. 3a. The purpose is to examine ‘scales’ in sharing for specific content categories. The scale is represented with three options of propagation to (i) *all country-specific websites* meaning tendency for being global (ii) *some country-specific websites* meaning tendency for being regional and (iii) *no propagation* meaning tendency for being local. The purpose of applying concept of propagation in website pair as in Fig. 3b is to examine ‘coupling’ between websites when sharing specific content categories. The higher occurrences of (i) *propagation* in website pair meaning high coupling between websites and (ii) *no propagation* meaning low coupling between websites.

#### 4.1.2 Within and Among Geographic Region

Since country-specific websites also represent specific geographic regions, abstracting the concept of propagation at this level, expands our understanding of preferences when sharing within and among specific regions. Figure 3d illustrates intra-regional propagation occurring in websites of India and Australia, both countries within Asia Pacific. Coupling among countries within a region is measured by the occurrence of propagation and no propagation among websites. Figure 3c represents inter-regional propagation occurring among countries in Asia Pacific (India, Australia) and Europe (UK and Ireland), in a structure which is a subset of website graph. Coupling among regions is again determined from occurrences and no occurrence of propagation. In the subsequent sections, we will determine community preferences when sharing

**Content Category:** Corporate, Product and Customer Support Information

**Geographic Regions:** Asia Pacific, Europe, North America and Middle East-Africa



**Fig. 3** Propagation in content categories and geographic regions

specific content categories and specific geographic regions by analyzing propagation in website graph, website pair, within and among geographic region.

## 4.2 Content Preferences

Websites from 10 global brands ranked highly in the web globalization report card (Yunker 2014) are selected for this study. Sample of 8 country-specific websites from each brand representing countries: India (IN), Australia (AU), United Kingdom (UK), Ireland (IE), United States (US), Canada (CA), Middle East (ME) and South Africa (ZA) and representing geographical regions: Asia Pacific, North America, Europe and Middle East-Africa are chosen. A total of 80 country-specific websites are collected as the source for webpages to be used for comparison. From 8 country-specific websites, we also have 28 possible websites pairs representing content sharing in country pairs such as (IN, AU), (IN, UK) and so on.

Webpages offering content in ‘English’ language are selected. Content categories used for sampling webpages are: (a) *Corporate Information*: webpages that provide background information of a company such as mission statements, history and its people (b) *Product Information*: webpages on description, usage, and specification of product and (c) *Customer Support Information*: webpages on ways to contact

company or find answer to queries. We then manually label webpages to their specific content categories. From each global brand we collected 48 webpage samples making a total of 480 webpage samples. We applied propagation-based approach to check for content propagation in website graph and website pair. A total of 480 webpages are qualitatively compared for their propagation in website graph. For each website pair there are 60 comparisons of webpages making a total of 1680 comparison for all 28 website pairs.

### 4.2.1 Propagation in Website Graph

Table 2 illustrates that out of 160 comparisons of webpages in “Corporate Information”, 50% of cases are identified in which propagation occurs among all country-specific websites while 32% of cases in which propagation occurs in some websites and 18% of cases in which no propagation occurs among the websites. As for more than 80% cases propagation occurs from at least a single country-specific website, it can be agreed that the suitability of content related to “Corporate Information” at a global scale. This means the dissemination of up-to-date knowledge is required globally. Only 15% cases in which propagations occur in all websites are identified for “Product Information” which strongly suggest that such content are not globally suitable. However, 36% cases of propagation to some websites and 49% cases of no propagation are comparable which infers “Product Information” may be suitable both regionally and locally among countries. This means the dissemination of up-to-date knowledge is either restricted among several countries within and across regions or limited to specific country. Contrary to this 66% cases of no propagation among countries strongly suggested that in “Customer Support Information” is locally suitable within a country. Local scale also suggests for synchronization of content updates to occur in local languages within a country. For example, content synchronized in official languages (English and French) within Canada.

### 4.2.2 Propagation in Website Pair

Table 2 illustrates that out of 560 comparisons of webpages in “Corporate Information” 71% of cases with propagation occurs in website pairs. This suggests high

**Table 2** Results of comparing webpages with content categories

Content category	Propagation in website graph			Propagation in website pair	
	All (%)	Some (%)	None (%)	Yes (%)	No (%)
(i) Corporate information	80 (50)	52 (32)	28 (18)	397 (71)	163 (29)
(ii) Product information	24 (15)	57 (36)	79 (49)	200 (36)	360 (64)
(iii) Customer support information	32 (20)	23 (14)	105 (66)	142 (25)	418 (75)

coupling when sharing content related to corporate information. Consistency has to be strictly enforced as such content are more likely to be updated frequently. 75% of cases with no propagations are identified for “Customer Support Information”. This suggest low coupling in websites and consistency is not strictly enforced except in local languages. The coupling in a website pair while sharing content for “Product Information” tends to be neutral (no significant difference in occurrence of propagation and no propagation). This suggest that policy for consistency to be moderately enforced while sharing product related information. This section expanded our understanding of community preferences in sharing that differs for specific content categories. Next we will detail customization involving geographic regions.

### **4.3 Geographic Preferences**

As with previous setup we categorized country-specific websites from the global brands into four geographic regions: Asia Pacific, North America, Europe and Middle East-Africa. We again sampled webpages related to “Corporate Information”, “Product Information” and “Customer Support Information”. A total of 480 webpages samples were manually labeled to their specific content categories and geographic regions. We applied propagation-based approach to check for content propagation within geographic region and among geographic regions. A total of 240 comparisons of webpages are performed to check for propagation occurring within all four geographic regions. A total of 1440 comparisons of webpages are performed to check for propagation among geographic regions.

#### **4.3.1 Propagation Within Geographic Region**

From Table 3 we found that the number of occurrences of propagation and no propagation among websites are comparable within Asia Pacific and Middle East-Africa. But for websites within North America, the majority of cases 67% show no propagation. This suggest low coupling in sharing content among countries in North America. In contrast the majority cases of propagation almost 60% occur among country-specific websites in Europe. This means high coupling in websites among countries in Europe.

#### **4.3.2 Propagation Among Geographic Region**

As in Table 3 there tends to be noticeable differences in the number of occurrences of propagation and no propagation while sharing content with countries in North America. More than 60% cases of no propagation in websites for sharing content from Asia Pacific, Europe and Middle East-Africa with customer in North America. This suggest low coupling in websites when sharing content with North America meaning

**Table 3** Results of comparing webpages with geographic regions

Propagation among geographic region		
	Yes (%)	No (%)
(i) Asia Pacific, Europe	107 (45)	133 (55)
(ii) Asia Pacific, North America	78 (33)	162 (68)
(iii) Asia Pacific, Middle East-Africa	105 (44)	135 (56)
(iv) Europe, North America	83 (35)	157 (65)
(v) Europe, Middle East-Africa	110 (46)	130 (54)
(vi) North America, Middle East-Africa	87 (36)	153 (64)
Propagation within geographic region		
	Yes (%)	No (%)
(i) Asia Pacific	29 (48)	31 (52)
(ii) Europe	35 (58)	25 (42)
(iii) North America	20 (33)	40 (67)
(iv) Middle East-Africa	32 (53)	28 (47)

websites within North America tend to have less interaction with websites from other region. Content categories wise we find that the occurrences of propagation tends to be higher among regions Asia Pacific, Europe and Middle East-Africa for sharing corporate related information in comparison to North America. This suggests high coupling among regions except with North America. Also corporate related information tend to be globally suitably. Higher occurrences of no propagation among region while sharing “Product Information” suggest that such content tend to be region specific and either locally or regionally suitable. In fact more than 70% no propagation cases occur for product related with North America meaning region specific product information is mostly preferred. The differences in the occurrences of propagation and no propagation for “Customer Support Information” seem to be consistent among all regions suggesting websites in all regions are more likely to prefer locale content. Next we will summarize the findings of this chapter.

## 5 Discussion

In designing a multi-language knowledge sharing system, we first shifted our focus from stating consistency rules explicitly to highlight source of inconsistency. Cases such as *content omitted*, *content updates not propagated* and *content conflicts* were considered as they are obvious in collaborative setting. Their occurrences in real world example of country-specific websites managed in a global brand also shed light on some of the consequences such as (a) community bias, (b) global and local inconsistency and (c) regional discrepancies. By no means are the mentioned cases

trivial; the complexity to detect their presence meaning inconsistency increases with the number of languages and communities participating in knowledge sharing. Then we touched on an important issues about consistency constraints due to opposing knowledge sharing goals of communities. We made following contributions.

1. ***Approaches for consistency adhering to knowledge sharing goals.*** We proposed a process-based approach to detect inconsistency in multilingual content when leveraging knowledge equally. The concept is to synchronize user editing activities realized with a state transition model and design of rules to detect inconsistent states. We evaluated approach to be satisfactory with an average precision of 88% and a recall of 86% in detecting cases of inconsistency. We also proposed a propagation-based approach to determine scales and coupling a key indicator for community preference in customized knowledge sharing. We used propagation in website graph, website pair, within and among geographic region to show preferences for sharing content categories/geographic regions.
2. ***Guidelines to enforce content consistency.*** We suggested guidelines for content consistency. We showed that community prefer to share corporate related information globally, customer support related information locally and product related information regionally. Implication is ‘global consistency policy’ for corporate related, ‘local consistency policy’ for customer support related information is required. Patterns of sharing Internationalization, Regionalization and Localization in [12] are useful to propagate updates consistently based on scales. We showed that community prefer high coupling when sharing corporate related and low coupling when sharing customer support related information. High Coupling means frequent interaction for sharing content, more vulnerable for inconsistency and higher priority for consistency policy required. Websites in Europe tend to be more dependent due to high coupling and share most content compared to websites in North America. Websites inside North America tend to be autonomous and participate less in sharing. This means customer inside European region is more vulnerable to ‘intra-regional discrepancies’, a higher priority for ‘intra-regional consistency policy’ for European countries. High coupling among Asia, Europe and Middle East-African countries suggest higher priority for ‘inter-regional consistency policy’. We also revealed that websites in North America have higher preferences for specialized product related information not shared with other region; while customer support related information are specialized inside all regions and not shared; for both cases ‘intra-regional consistency policy’ is suited. Details also examined in [11].
3. ***Support for resource poor language communities.*** We proposed approaches that require minimal language processing and supported knowledge sharing for resource poor communities. The problem surfacing limited support of existing approaches to content consistency in resource poor languages is scarce linguistic corpuses to perform advance NLP operations. Better results can be anticipated when proposed approaches is integrated with NLP at a preliminary stage of inconsistency management and using framework such as Language Grid [7].

## 6 Conclusion

This chapter addressed two important issues in multi-language knowledge sharing: (i) impracticality in stating consistency rules explicitly in advance and (ii) constraint in content consistency. We showed that even minor cases such as *content omitted*, *content updates not propagated* and *content conflicts*, when considered as potential cause for inconsistency, it leads to undesirable consequences: *community bias*, *global and local inconsistency* and *regional discrepancy*. We also showed that opposing knowledge sharing goals impose consistency requirements from rigid to non-rigid. To avoid inconsistencies while adhering to knowledge sharing goals of communities, we contributed in the design of multi-language knowledge sharing system with (a) process-based approach for multilingual content synchronization to leverage knowledge equally and (b) propagation-based approach to analyze community preferences when sharing specific content categories/geographic regions, to customize knowledge sharing. We also extended support to resource poor language communities by basing our approach on minimal language processing requirements.

**Acknowledgements** This research was supported by the Grant-in-Aid for Scientific Research (S) (24220002, 2012–2016) from Japan Society for the Promotion of Science (JSPS).

## References

1. Adar, E., Skinner, M., Weld, D.S.: Information arbitrage across multi-lingual Wikipedia. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pp. 94–103. ACM (2009)
2. Al Assimi, A.B., Boitet, C.: Management of non-centralized evolution of parallel multilingual documents. In: Proceedings of Internationalization Track, 10th International World Wide Web Conference, Hong Kong, pp. 1–8 (2001)
3. Hartley, A., Paris, C.: Multilingual document production from support for translating to support for authoring. *Mach. Transl.* **12**(1), 109–129 (1997)
4. Hopcroft, J.E., Motwani, R., Ullman, J.D.: Automata theory, languages, and computation. *Int. Ed.* **24** (2006)
5. Huberdeau, L.P., Paquet, S., Désilets, A.: The cross-lingual wiki engine: enabling collaboration across language barriers. In: Proceedings of the 4th International Symposium on Wikis, WikiSym '08, pp. 13:1–13:14. ACM, New York (2008)
6. Huizingh, E.K.: The content and design of web sites: an empirical study. *Inform. Manage.* **37**(3), 123–134 (2000)
7. Ishida, T.: The Language Grid: Service-oriented Collective Intelligence for Language Resource Interoperability. Springer Science & Business Media, Berlin (2011)
8. Kayan, S., Fussell, S.R., Setlock, L.D.: Cultural differences in the use of instant messaging in Asia and North America. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 525–528. ACM (2006)
9. Monz, C., Nastase, V., Negri, M., Fahrni, A., Mehdad, Y., Strube, M.: Cosyne: a framework for multilingual content synchronization of wikis. In: Proceedings of the 7th International Symposium on Wikis and Open Collaboration, pp. 217–218. ACM (2011)
10. Pariyar, A., Lin, D., Ishida, T.: Tracking inconsistencies in parallel multilingual documents. In: 2013 International Conference on Culture and Computing (Culture Computing), pp. 15–20. IEEE (2013)

11. Pariyar, A., Murakami, Y., Lin, D., Ishida, T.: Content sharing in global brand from geographic perspective. In: 2015 International Conference on Culture and Computing (Culture Computing), pp. 151–158. IEEE (2015)
12. Pariyar, A., Murakami, Y., Lin, D., Ishida, T.: Content sharing in global organization: a cross-country perspective, pp. 1–10 (2015)
13. Robbins, S.S., Stylianou, A.C.: Global corporate web sites: an empirical investigation of content and design. *Inform. Manage.* **40**(3), 205–212 (2003)
14. Tonella, P., Ricca, F., Pianta, E., Girardi, C.: Restructuring multilingual web sites. In: International Conference on Software Maintenance, 2002. Proceedings, pp. 290–299. IEEE (2002)