

(続紙 1)

京都大学	博士 (情報学)	氏名	AMIT PARIYAR
論文題目	Supporting Consistencies in Multi-Language Knowledge Sharing (多言語知識共有における一貫性支援)		
<p>(論文内容の要旨)</p> <p>The goal of this research is to support multi-language knowledge sharing with a focus on content consistency among local, regional and international communities. Due to the rapid growth of online collaboration, the probability of sharing inconsistent content in various languages and communities is increased. This is problematic for multi-language knowledge sharing systems because (i) consistency rules for the shared content cannot be explicitly stated in advance and (ii) there exist constraints from diverging views in supporting content consistency. To this end, this research proposes techniques and guidelines on content consistency to support knowledge sharing among multi-language communities. This thesis consists of seven chapters.</p> <p>Chapter 1 outlines the thesis including the research overview, objectives, issues and approaches.</p> <p>Chapter 2 describes the background of this thesis. This chapter examines previous research to understand knowledge sharing goals among communities. Previous research explored the processes of leveraging knowledge equally and customizing knowledge sharing. These researches highlighted the constraints in handling content consistency for knowledge sharing. The reliance of previous techniques that use content processing and their limited support to resource poor languages also motivated this research to offer solution for content consistency to larger communities, including the resource deprived ones.</p> <p>Chapter 3 proposed a process-based approach to detect inconsistent content shared among communities in order to leverage knowledge equally. The proposed approach is not language specific as it is based on the concept of synchronizing user editing actions. To realize this approach, a state transition model is introduced to define the states in multilingual content, the set of actions and transition functions. Using the combination of states, rules are designed to represent inconsistent multilingual content. Experimental results from applying the proposed approach to revision history of Wikipedia articles in English and Nepali languages showed a satisfactory result with an average precision of 88% and a recall of 86% in detecting inconsistency.</p> <p>Chapter 4 shifted the focus to customization in knowledge sharing and proposed guidelines for web managers to promote consistency while sharing specific content categories among communities. The approach based on propagation is proposed to qualitatively compare the content in webpages of country-specific websites and examine their propagation in (a) website graph and (b) website pair. For this study, 480 webpage samples from 80 websites representing ten global brands (Nivea, 3M, Starbucks, Acer, Samsung, KPMG, HP, Nestle, Avon, John Deere) are analyzed. A total of 480 comparisons of webpages in website graph and 1680 comparisons in website pairs are performed to examine propagation of content. The results revealed scales in sharing such as, corporate related information that is shared globally, customer support related information that is shared locally and product related information that is shared</p>			

regionally and locally. The results also revealed coupling phenomena in sharing, such as high coupling between websites when sharing corporate related information. These findings are essential in the development of the guidelines for content consistency, including global consistency needed for corporate related information and the local consistency needed for customer support information. Such guidelines are also useful in avoiding cross-site content inconsistencies.

Chapter 5 further explored customization in knowledge sharing with respect to geographic regions and proposed guidelines on content consistency for sharing in specific regions. The approach based on propagation is proposed to qualitatively compare webpages of country-specific websites from various geographic regions and examine their propagation (a) within the regions and (b) among the regions. Eighty websites from the geographic regions of North America, Asia Pacific, Europe and Middle East / Africa are analyzed. A total of 240 comparisons of webpages within regions and 1440 comparisons among regions are performed to examine propagation of content in specific regions. The results revealed high coupling in websites among countries in Europe and low coupling in websites inside North America. This suggested high probability for intra-regional discrepancies in content shared inside the European region. The results also revealed high coupling in websites among Asia Pacific, Europe and Middle East / Africa, and suggested a high probability for inter-regional discrepancies. Implications of such findings are the guidelines, example strict consistency policy needed for sharing content inside European region compared to North America. Such guidelines are useful in avoiding regional discrepancies caused by sharing inconsistent content.

Chapter 6 proposed a simple technique to execute the guidelines determined in former chapters. Based on propagation, pattern of sharing (a) internationalization, (b) regionalization and (c) localization are proposed to restrict content updates confined to specific communities and their languages. Besides simplicity, the advantage is its suitability for both monolingual and multilingual cases, and its automated or manual support for execution.

Chapter 7 summarizes the original contributions and examines future directions.

From the techniques that are simple and applicable in variety of languages along with the guidelines for web managers to deal with inconsistencies in (a) multilingual content and (b) cross-site content, this research contributed to the design of a multi-language knowledge sharing system that caters to the goals of communities which require both for leveraging knowledge equally and customization in knowledge sharing.

注)論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせ

て、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 words で作成し

審査結果の要旨は日本語500～2,000字程度で作成すること。

(続紙 2)

(論文審査の結果の要旨)

本論文は、言語の異なるコミュニティ間で共有されるコンテンツの一貫性を維持するための方法を検討したもので、得られた主要な成果は以下の通りである。

1. 状態遷移モデルを用いた多言語コンテンツの非一貫性の検出

言語の異なるコミュニティ間で知識共有を実現するために、コンテンツの削除や更新によって生じるコンテンツの非一貫性を検出する手法を考案している。具体的には、修正中、確定、翻訳完了の3つを編集状態とし、各状態における実行可能な編集行為と状態遷移によって定義される状態遷移モデルを提案している。この状態遷移モデルを用いて、言語の異なるコミュニティの編集行為の履歴から、コンテンツの非一貫性を推定する手法を提案している。この手法は、ユーザの編集行為にのみ基づき、コンテンツの内容や言語に依存しないため、東南アジアなどコンピュータによって処理可能なデータの少ない低資源言語に対しても適用可能である。提案手法をWikipedia記事の編集履歴に適用した結果、高い精度で非一貫性を検出できることを示している。

2. コンテンツカテゴリに着目したWebサイトの一貫性分析

グローバル企業等のWebサイトでは、単純にコンテンツを共有するのではなく、国や地域に応じた記事の改編が行われることがある。そこで、コンテンツのカテゴリに応じて共有方針にどのような差があるかを分析している。その結果、会社概要はグローバルに共有され、製品情報は地域内で共有され、カスタマーサポートの情報は国内で共有される傾向を見いだしている。さらに、こうした知見に基づき、Webサイト間のコンテンツの一貫性を維持するためのガイドラインを提案している。

3. コンテンツの地域性に着目したWebサイトの一貫性分析

欧州、北米、アジア太平洋、中東アフリカなどの地域に応じて、Webサイト間でコンテンツの共有方針にどのような差があるかを分析している。この分析では、コンテンツを地域内と地域間で伝播する場合に分け比較を行っている。その結果、地域内の伝播では、欧州各国間で頻繁に共有が図られること、北米では各国の独立性が高いことを見いだしている。また、地域間の伝播では、欧州、アジア太平洋、中東アフリカ間で頻繁に共有が図られることを見いだしている。これらの知見に基づき、地域内や地域間でのコンテンツの一貫性を維持するためのガイドラインを提案している。

以上のように、本論文は、言語の異なるコミュニティ間のコンテンツの非一貫性を検出する手法を提案すると共に、コンテンツのカテゴリや地域性に着目して、その一貫性を維持するためのガイドラインを提案しており、学術上、実用上寄与するところが少なくない。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成27年7月30日に実施した論文内容とそれに関連した試問の結果合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。
要旨公開可能日： 年 月 日以降