

クラウドコンピューティングによるオフショア開発支援環境の提案

M2010MM047 張 錦峻

指導教員 青山 幹雄

1. はじめに

オフショア開発における文書交換の支援に着目する。中国人向けの文書作成と自動翻訳を支援するサービスを中心に、文書をクラウドに一元管理するクラウドコンピューティング開発環境を提案する。オフショア開発において、ブリッジSEの能力に依存せず、効率よく正確な情報伝達の実現を目指す。

2. 背景と問題点

近年、グローバル化するソフトウェア開発において、オフショア開発市場が拡大している。しかし、オフショア開発における情報伝達では次の問題が起きている。

(1) 発注側と受注側の直接情報伝達

遠隔地でコミュニケーション効率が悪く、情報伝達は効率的にできない。さらに、言葉や文化が異なるため、情報を誤解することが多く、情報を正確に伝達するのが難しい。従って、異文化コミュニケーションにおける情報伝達が困難になっている [2][6]。

(2) ブリッジSEによる情報伝達

ブリッジSEの能力と権限により、情報を正確に効率的に伝えられないことがある。そのため、システム開発の成否がブリッジSEに左右されることになる。ブリッジSEはオフショア開発のボトルネックとなる可能性がある。従って、ブリッジSEによらない情報伝達の仕組みが必要となる [2][3]。

3. 関連研究

3.1. 中国人向けの文書作成方法

オフショア開発における文書を作成するには、相手が日本語に習熟していない外国人を意識し、平易な文章を工夫して作成する必要がある。中国人向けの文書作成方法がまとめられている (表 1) [2][3]。

しかし、中国人向けの文書における不適切な表現を細分化していないため、人の能力によって中国人向けの文書作成方法を十分に活用できない可能性がある。

表 1. 中国人向けの文書作成方法の一覧

番号	項目	概要
1	カタカナ禁止	カタカナ英語は辞書にないことが多いため、英語で書く
2	否定文禁止	語順が違うため、否定文を肯定文に置き換える
3	冗長文章禁止	短い簡潔な文章で記述、1つの文章に1つの意味
4	曖昧言葉禁止	不明確な言葉を避けて言葉を厳密に使う
5	項目番号付与	箇所書きを多用して項目番号を使って会話
6	記述粒度統一	記述内容の詳細を明確する上で粒度を揃えて記述

3.2. 言語グリッドによる翻訳

言語グリッドでは、機械翻訳などの言語資源をクラウドコンピューティングとして提供されている。ユーザ辞書を利用する翻訳サービスの作成が可能である [1][4]。

しかし、ソフトウェア開発向けの翻訳精度の高い翻訳サービスが提供されていない。

4. アプローチ

4.1. 日本語表現アンチパターンの細分化

不適切な日本語表現を日本語表現アンチパターンとして定義し、分類する。さらに、問題を起こした原因を分析することにより、日本語表現アンチパターンの分類を階層的に細分化する。日本語表現アンチパターンに基づき、日本語表現をチェックし、適切な表現に改善する方法を定義する。

4.2. 経験知の有効利用

コミュニケーションによる問題を解決するには、相手国の言語や文化を学ぶ以外には、積み上げた経験やノウハウの活用が重要である [5]。

オフショア開発において、文章に不適切な日本語表現、日中IT専門用語、ブリッジSEによる改善の訳文を経験知として定義する。さらに、経験知を活用できる仕組みを提案する。

4.3. クラウドによるオフショア開発の支援

オフショア開発に関わるリソースを、Web経由でクラウドコンピューティングとして提供することにより、リアルタイムなデータの共有も、開発環境の統一も可能になる。さらに、業務に役立つクラウドコンピューティングを利用し、オフショア開発を支援できる (図 1)。

言語グリッドを連携する文書翻訳サービス、日本語表現チェックサービスを含め、オフショア開発を支援するクラウドコンピューティングアーキテクチャを提案する。

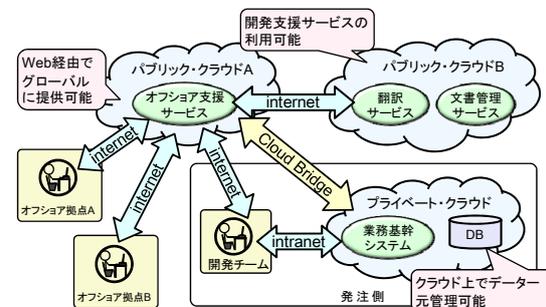


図 1. クラウドによるオフショア開発の支援

5. 日本語表現アンチパターン

文書による情報伝達において、誤解を招く、不適切な表現を日本語表現アンチパターンとして定義する。

5.1. 日本語表現アンチパターンの分類

習慣、言葉難易度、文化に着目し、日本語表現アンチパターンを曖昧表現と難解表現と文化背景表現に分類する。さらに、表現の不適切さの違いを抽象化し、日本語表現アンチパターンの分類を細分化する。これにより、不適切さの原因を明確でき、共通の対応方法で文章を改善できる。クラス図を用いて階層構造モデルを表現し、日本語表現アンチパターンの所属関係を示す(図2)。

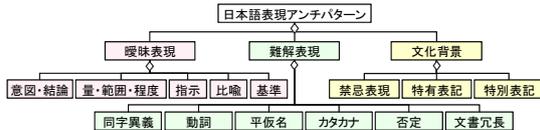


図2. 日本語表現アンチパターン階層モデル

- (1) 曖昧表現: 日本人が習慣的によく使用する、意味がはっきりしない、場合による意味が変わる言葉を含む(表2)。
- (2) 難解表現: 語の意味の複雑さと、中国語との違いによる、中国人にとって理解し難い言葉を含む。
- (3) 文化背景表現: 日本で特有の表現と、使用してはいけない差別用語などの言葉を含む。

表2. 日本語表現アンチパターン一覧(曖昧表現)

分類	番号	項目	例
曖昧表現	1	意図・結論表現	しなくてもよい、かもしれない、ほうがいい
	2	量・範囲・程度表現	おおよそ、大体、大抵、ほど、ぐらい、多分、できるだけ、ほとんど、あたり、非常に
	3	指示表現	下記の、上記の、前述の、後述の、これ
	4	比喩表現	~ようである、~らしい、~みたいな感じで
	5	基準表現	必要な場合、(状況・場合・必要)に応じて

5.2. 日本語表現の改善方法

中国人向けの文書作成方法に基づき、改善方法として次の3パターンを定義する。さらに、日本語表現アンチパターンを改善方法と対応付けることにより、共通な方法で文章を改善できる(表3)。

- (1) 付記: 語の意味を間違えないように、中国語、英語による意味の解釈を付け加える。
- (2) 書換え: 不適切な表現を適切な表現で置き換える。
- (3) 削除: 不適切な表現を削除する。文章の意味が変わらないことを前提とする。

表3. アンチパターンと改善方法の対応付け(曖昧表現)

分類	項目	付記	書換え	削除	例、説明
曖昧表現	意図・結論表現	×	○	×	「しなくてもいい」⇒「しないでください」
	量・範囲・程度表現	×	×	○	「大体15画面」⇒「15画面」
	指示表現	×	△	×	
	比喩表現	×	△	×	前後の文脈を分析する必要がある。改善した文章を登録した場合、完全一致で検索し書換え可能
	基準表現	×	△	×	

5.3. 日本語表現の改善メッセージ

文章改善のプロセスをユーザに明示する必要がある。日本語表現アンチパターンの分類と改善方法に基づき、次の3種のメッセージを定義する。さらに、3種のメッ

セージを組み合わせる改善メッセージを作成する(表4)。

- (1) 促進メッセージ: 問題点を指摘し、改善を促進する。
- (2) 説明メッセージ: 表現の不適切さを説明する。
- (3) 対策メッセージ: 改善の対策を提案する。

表4. 日本語表現の改善メッセージ(曖昧表現)

分類	促進メッセージ	項目	説明対策メッセージ
曖昧表現	曖昧表現なので文書を誤解する可能性がある。	意図・結論表現	説明 結論または意図が不明確である。 対策 選択肢を2つを分け、1つのみ指定し、伝えたいことを明確してください。
		量・範囲・程度表現	説明 量・範囲・程度を表す副詞がついている。 対策 不要な副詞を削除してください。
		指示表現	説明 指示のものが不明確である。 対策 項目番号を付与し、利用して会話してください。
		比喩表現	説明 比喩表現である。 対策 比喩表現を避けてください。
		基準表現	説明 物事の基準が不明確である。 対策 基準の判別がつかない言葉を削除してください。

5.4. 日本語表現の改善の実例

5.1節で分類した日本語表現アンチパターンと、5.2節で示した改善方法に基づき、日本語表現をチェックし、改善した例を示す(図3)。

- (1) 「ファイル」はカタカナ表記のため、英文で書き換えるか、または英文を付記する。例では、カタカナ表記を英文で書き換える。
- (2) 「消さない」は「消す」の否定文になる。「消す」は五段動詞なので、この場合ではサ変動詞である「削除」で置き換える。「消さない」は「削除しない」に書き換える。
- (3) 「しない」は「を禁止する」の肯定文に変更する。さらに、「削除」前の格助詞である「を」を削除する。

このように文書を改善することにより、例え日本語が堪能でなくても、文章を誤解することを防ぎ、情報を正確に伝達できる。

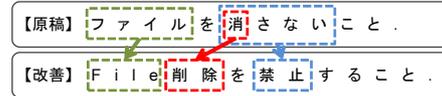


図3. 日本語文章改善の実例

5.5. 日本語表現アンチパターンデータモデル

日本語表現アンチパターンを表現チェック項目としてデータを記述するには、次の属性を必要とする(表5)。

- (1) 分類: 日本語表現アンチパターンのカテゴリを示す。曖昧表現、難解表現、文化背景表現を記述する。
- (2) 項目: 不適切さの違いにより抽象化した日本語表現アンチパターンの細分項目を記述する。
- (3) 元表現: 不適切な日本語表現の内容を記述する。
- (4) 改善表現: 改善した表現を記述する。
- (5) 改善方法: 付記、書換え、削除の改善方法を記述する。
- (6) 改善メッセージ: 改善促進メッセージ、説明メッセージ、対策メッセージ、3種のメッセージを記述する。
- (7) 推奨度: 改善方法の推奨度を記述する。推奨度を付けることにより、場合によって適切な改善方法を選択でき、改善効果を高める。

表5. 日本語表現チェックデータの例

分類	項目	元表現	改善表現	改善方法	改善メッセージ	推奨度
難解表現	同字異義	年度末	年度末("3月末")	付記	-	1
		3月末	書換え		2	

6. 専門用語の定義

文書の自動翻訳を行う場合、専門用語辞書による翻訳精度を大きく向上できる。ユーザ辞書共通フォーマットである UTX(Universal Terminology eXchange) 仕様に従い、IT 専門用語辞書を定義する(表 6)。

表 6.IT 専門用語辞書のデータ(抜粋)

#UTX 1.11; ja-JP/zh-CN; 2011-04-15T10:00:00Z+09:00; Copyright:CHOU (2012); bidirectional				
src	tgt	src:pos	term status	concept ID
ネットワーク	网虫	noun	approved	100
ネチズン	网民	noun	approved	101
アクセス	存取	verb	approved	105
アクセス	访问	verb	forbidden	105

7. オフショア開発支援環境の提案

7.1. オフショア開発支援クラウドアーキテクチャ

オフショア開発を支援するクラウドコンピューティングアーキテクチャを図 4 に示す。文書管理、文書翻訳、経験知蓄積、日本語表現チェックを含め、オフショア開発を支援する機能を、プライベート型サービスとして提供する。発注側が必要なデータと機能のみを公開し、オフショア開発の支援アプリケーションをパブリッククラウド上で構築する。

- (1) 文書管理サービス: 文書の作成、参照、更新、削除、通知送付などの機能を提供し、文書をクラウドに一元管理するサービスである。文書交換、文書管理を支援する。
- (2) 文書翻訳サービス: 翻訳精度の向上の仕組みを組み込み、文書を翻訳するサービスである。ブリッジSEの翻訳作業を支援する。
- (3) 経験知蓄積サービス: 日本語表現アンチパターン、開発技術者が定義した専門用語、ブリッジSEによる改善の訳文といった、実際の業務から得られた経験を再利用するために、知識として蓄積するサービスである。蓄積された経験知が文章改善、文書翻訳などに利用される。
- (4) 日本語表現チェックサービス: 文書による情報伝達が正確に行うために、日本語表現をチェックし文章を改善するサービスである。中国人向けの文書作成を支援する。

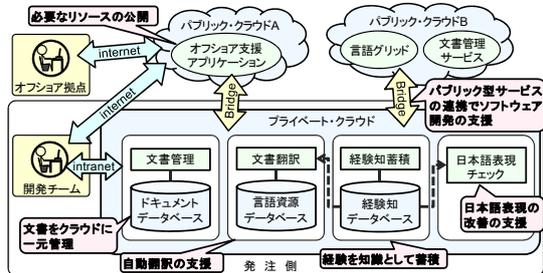


図 4. オフショア開発支援クラウドアーキテクチャ

7.2. 日本語表現チェックサービス

日本語表現チェックと日本語文章改善の 2 つのモジュールで日本語表現チェックサービスを構成する(図 5)。

- (1) 日本語表現チェック: 言語グリッドの形態素解析原子

サービスを利用し、形態素を解析する。形態素ごとに表現チェック項目を検索し、改善の方法、改善のメッセージなどの情報を含む、日本語表現チェック項目一連のデータをチェック結果ファイルにまとめる。チェック配置子にはチェック対象となる品詞の種類を記述する。日本語表現チェックモジュールの振る舞いを図 6 に示す。

- (2) 日本語文章改善: 日本語表現チェックの結果に基づき、文章を改善する。チェック結果ファイルを取得できない場合、日本語表現チェックモジュールを先に呼び出す。

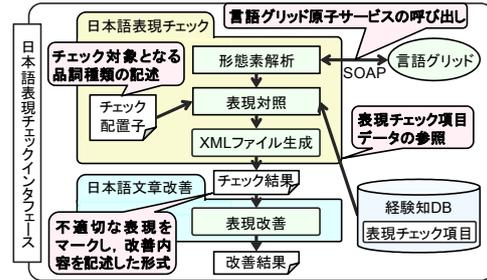


図 5. 日本語表現チェックサービスアーキテクチャ

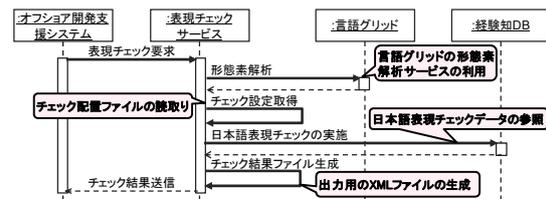


図 6. 日本語表現チェック振る舞いモデル

7.3. 文書翻訳サービス

定型文、専門用語、翻訳リペアを用いて、翻訳精度を向上する(図 7)。言語資源データベースには、Web 日本語 n グラムと類語辞書を設置する。Web 日本語 n グラムは、Google が Web ページから抽出した、大規模な日本語データ (約 2550 億単語) である。Web 日本語 n グラムにより、言葉の出現頻度を測定する。

- (1) 定型文翻訳: ブリッジSEによる改善の訳文を定型文として定義する。定型文データを参照し、翻訳できた場合、専門用語翻訳、翻訳リペアを実行しない。
- (2) 専門用語翻訳: 形態素ごとに、IT 専門用語のデータを取得する。原文と専門用語データを引数として、言語グリッドの機械翻訳サービスに渡し、翻訳を行う(図 8)。
- (3) 翻訳リペア: 翻訳精度を低下させる不適切な箇所を折り返し翻訳で抽出し、Web 日本語 n グラムによる言換え候補をフィルタリングし、出現頻度の高い類語で置き換える(図 9)。

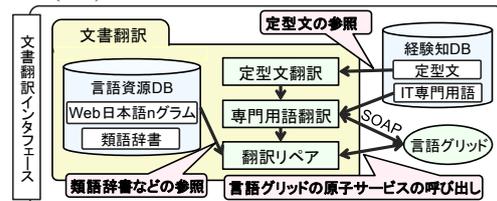


図 7. 文書翻訳サービスアーキテクチャ

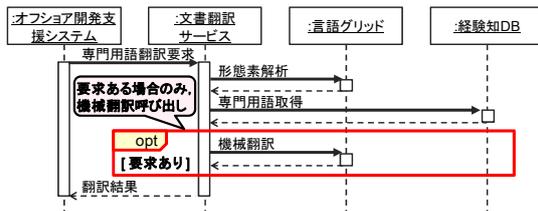


図8. 専門用語翻訳振る舞いモデル

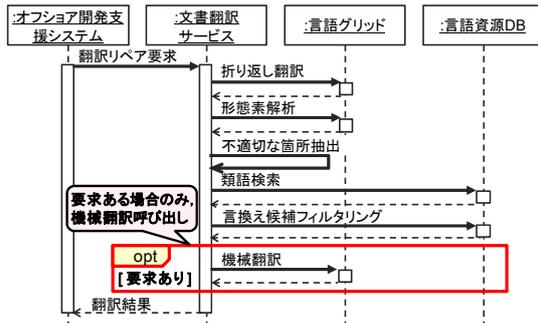


図9. 翻訳リペア振る舞いモデル

8. 評価と考察

8.1. 日本語表現の改善

日本語表現アンチパターンを階層的に細分化し、不適切さの原因を明確し、共通の対応方法で文章の改善を可能にする。これにより、中国人向けの文書に不適切な表現を避け、文章の誤解を防ぐことができる。読み手の能力に依存せず、文書による正確な情報伝達が可能となる。

8.2. 文書翻訳精度の向上

言語グリッド翻訳サービスをベースにし、定型文と専門用語辞書と翻訳リペアによる翻訳精度の向上を可能にする。これにより、定型文と専門用語の蓄積にとともに、翻訳精度を徐々に向上できる、オフショア開発向けの翻訳サービスの作成が可能となる。

8.3. 経験知の有効利用

日本語表現アンチパターン、専門用語、定型文といったデータを定義し、経験を知識としての蓄積を可能にする。日本語表現アンチパターンが日本語文章の改善に利用され、定型文と専門用語が翻訳精度の向上に役立つ。経験知の有効利用により、オフショア開発がブリッジSEを含む開発者の能力への依存度を低下させることができる。さらに、第三者の技術者がその知識を学習したり適用したりすることで生産効率の向上も期待できる。

8.4. クラウドによる開発支援環境の利用

(1) リアルタイムなデータ共有

日本語表現チェックサービス、文書翻訳サービスを中心に、文書をクラウドに一元管理する、提案クラウドコンピューティングにより、リアルタイムなデータ共有を可能にする。これにより、文書による効率的な情報伝達が可能となる。さらに、作業環境が統一され、オフショア開発全体の生産効率の向上も期待できる。

(2) 文章改善の自動化

定義した日本語表現アンチパターンと文章改善方法に基づき、文章改善のプロセスをクラウドコンピューティングとして提供することにより、文章改善の自動化を可能にする。これにより、中国人向けの文書の作成を支援することができ、書き手の能力に依存せず、中国人向けの文書の作成が可能となる。

(3) 文書翻訳の自動化

経験知を利用した翻訳精度向上の仕組みに基づき、文書翻訳のプロセスをクラウドコンピューティングとして提供することにより、文書翻訳の自動化を可能にする。これにより、ブリッジSEの翻訳作業を支援でき、ブリッジSEによる影響の低減が可能となる。

9. 今後の課題

(1) サービス利用規約とインタフェースの定義

提案したオフショア開発を支援するクラウドアーキテクチャでは、サービスの利用規約とインタフェースが未定義である。今後、サービスのインタフェースを定義し、サービスの利用規約を明確する必要がある。

(2) 適用による経験知の蓄積

提案したオフショア開発を支援するクラウドアーキテクチャでは、経験知を利用するため、今後提案環境を実際に適用し、経験知を蓄積する必要がある。

(3) 文書管理と経験知サービスアーキテクチャの明確化

文書管理と経験知サービスのアーキテクチャが未定義であるため、文書管理と経験知サービスのアーキテクチャを明確にする必要がある。

10. まとめ

日本語文章における不適切な表現を日本語表現アンチパターンとして定義し、細分化した。日本語表現をチェックし、適切な表現に置き換える方法を提案した。また、既存翻訳サービスと連携し、翻訳精度を向上する仕組みを提案した。さらに、中国人向けの文書作成と自動翻訳を支援するサービスを中心に、文書をクラウドに一元管理するクラウドコンピューティング開発環境を提案した。提案開発環境の妥当性を評価した。

参考文献

- [1] T. Ishida (ed.), The Language Grid, Springer, 2011.
- [2] 幸地 司, オフショア開発に失敗する方法, ソフト・リサーチ・センター, 2008.
- [3] 幸地 司ほか, 標準テキストオフショアプロジェクトマネジメントSE編, 技術評論社, 2009.
- [4] Language Grid, <http://langrid.org/en/index.html>.
- [5] 辻 洋ほか, オフショア・ソフトウェア開発の進化と技術者の経験知, 情報処理, Vol. 49, No. 5, May 2008, pp. 551-557.
- [6] L. Zhang, et al., Solution Proposals for Japan-Oriented Offshore Software Development in China, SEAFOOD, 2009, pp. 14-24.