

日本語設計書作成支援

企業のグローバル展開が加速される中、外国語による文書作成の機会が増加している。システム開発分野では中国でのオフショア開発が活発で、中国人技術者が日本語文書を作成することへの期待は大きい。しかしながら、作成時間の短縮や文書品質の向上の余地はあり課題は多い。NTT データでは、日本語の用法理解を促進するための用例文検索機能と、誤りを指摘する日本語チェック機能を開発した。これらにより、グローバルな開発体制の日本語設計書作成を支援する。

日本語設計書作成支援

NTT データにおけるオフショア開発への発注割合は 2011 年度の実績が 7.5%であり、2012 年度の目標は 10%と上昇傾向にある。オフショア開発で作成される日本語文書は、設計書、試験項目表といったプログラム製造に係わるものが主であるが、今後は、要件定義書といったシステムの全体構成に係わる文書の作成にも期待が寄せられている。このように、日本語が外国語であるシステム開発のエキスパートが、日本語文書を作成する機会は増加傾向にあると言える。

しかしながら、日本語教育を受けるのは就業してからという技術者も多く、開発現場にて様々な工夫がされているものの、作成に時間を要したり、誤った表現が多く含まれたりするのが現状である。

例えば、技術者の母国語にはない日本語特有の助詞や、動詞の活用といった文法で疑問点が生じると、日本語の文法書を確認したり、社内にいる日本語のエキスパートへ問合せたりしている。この場合、該当箇所の探索や、回答を受けるまでの時間を要することが、効率的な文書作成のさまたげとなっている。

さらに、表現のチェックについては、日本への文書による問い合わせや文書の納品の前に、リーダーの確認やメンバでのレビューを行っている。ここでは、目視により表記チェックや内容確認が行われるが、日本語として誤った表記は事前に洗い出し、その後、内容確認を行う方が作業効率の

観点から望ましい。

NTT データでは、オフショア開発で生じるこれら課題解消の取組として、特に中国人技術者向けの日本語設計書作成を支援する技術開発を行っている。具体的には、中国語にはない日本語特有の文法に対して、用法の理解を支援する用例文検索機能と、表記を自動でチェックする日本語チェック機能である。

開発プロジェクトの現状

開発プロジェクトで作成された中国人技術者の日本語文書の誤りは、中国語には存在しない助詞、動詞の活用、カタカナ語の誤りが 60%を占めている。

具体的には、表 1 の「動詞の活用」の例のように、時制を考慮せず活用を誤ったり、「助詞の省略」の例のように、助詞を過剰に省略したりする誤りが多い。これは、日本語と中国語の文法の違いである、助詞の有無、動詞活用の有無、時制の有無などの影響と言える。

表 1 日本語誤りの例

動詞の活用	ボタンを押下する後の処理
助詞の省略	ログに出力し処理開始する

また、カタカナ語の誤りの例として、英語の「message」に対して「メッセージ」と記載すべきところを、「メセッジ」と表記する誤りが多く見られる。これは、英語での発話の原音と日本語表記の差異に対し、英語の原音に影響された結果

と言える。

これらの誤りやすい文法や表記に対して、作成時の支援を行うのが用例文検索機能であり、作成後の確認時に支援を行うのが日本語チェック機能である。

用例文検索機能

技術者の記載する文は、プログラムで実行する処理ロジックが多い。この場合、処理対象となる名詞や処理方法となる動詞を要件定義書の記載から抜粋して日本語文を作成している。そのため、文を構成する名詞や動詞は明確で、文の作成作業は名詞と動詞をつなぐ助詞や、動詞の活用を選定する作業と言える。

これに対して、用例文検索機能では、上記の名詞や動詞をキーワードとして過去に作成した文法的に正しい文書集合から用例文を検索し、その結果を参考にすることで作成の支援を行う。検索結果の提示にあたっては、入力した名詞や動詞が助詞を介して係受け関係にある句を含む文を選び出すことで、助詞や動詞の活用の選定に適した用例文を優先的に表示することを実現している。

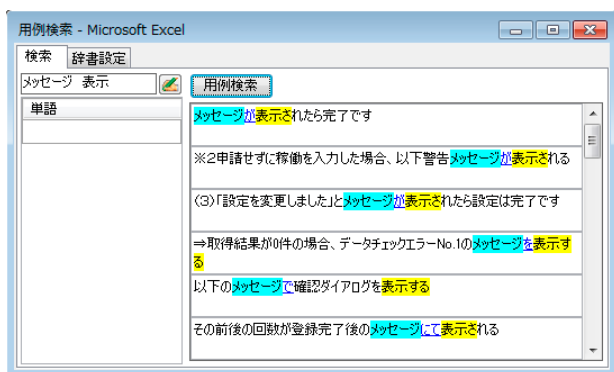


図 1. 用例文検索機能の出力例

図 1 は「メッセージ」と「表示」をキーワードに検索した出力例である。これは能動表現と受動表現の使い方が提示されており、それぞれに適した助詞と動詞の活用の使い方が把握できる。

このように、用例文検索機能を活用することで、

用法の理解に対して文法書の確認が不要となり、文書作成作業が効率的に行えるようになる。

日本語チェック機能

技術者の作成した文書は、日本に提出する前にレビューの機会が取られる。レビューでは、内容確認を行うことが本来の目的である。しかしながら、日本語表記の誤りがあれば、あわせて修正する必要がある。

これに対して、日本語チェック機能では、誤りの多い助詞とカタカナ語の表記をチェックし、日本語の誤りを自動で検出する。誤りの自動検出は、設計書の表記が定型的な表現が多いことから、設計書データの統計情報をもとに助詞やカタカナ語が標準的な用法から外れているものを検出している。

助詞は記載されたものが適切であるかだけでなく、名詞と動詞をつなぐ助詞が過剰に省略されていないかもチェックしている。また、カタカナ語のチェックは音素的な類似性を考慮する方式を採用し、誤り検出だけでなく正解候補の提示を行い修正の効率化につなげている。

このように、日本語チェック機能を活用することで、日本語として誤った表記を事前に洗い出せるようになり、文書品質の向上が実現できる。

今後の課題

グローバル化の進展した社会ではコミュニケーション・ギャップがこれまでに大きくなくなっていくと考えられる。NTT データでは、グローバルな開発体制で行われる文書によるコミュニケーションに対して、文書作成支援を中心に研究開発を行ってきた。今後は、理解支援のための翻訳技術、多言語文書活用のための検索技術など、さらなるギャップの解消に向けたメディア分析技術のソリューションを活用した、アプリケーション開発を進めていく予定である。

2012 年 10 月版