

SIプロセス効率化に向けた 知識の構造化と再活用の取組み

須間裕一*
Hirokazu Suma
相川勇之†
Takeyuki Aikawa
内出隼人†
Hayato Uchide

相馬仁志*
Hitoshi Soma
藤村圭介*
Keisuke Fujimura

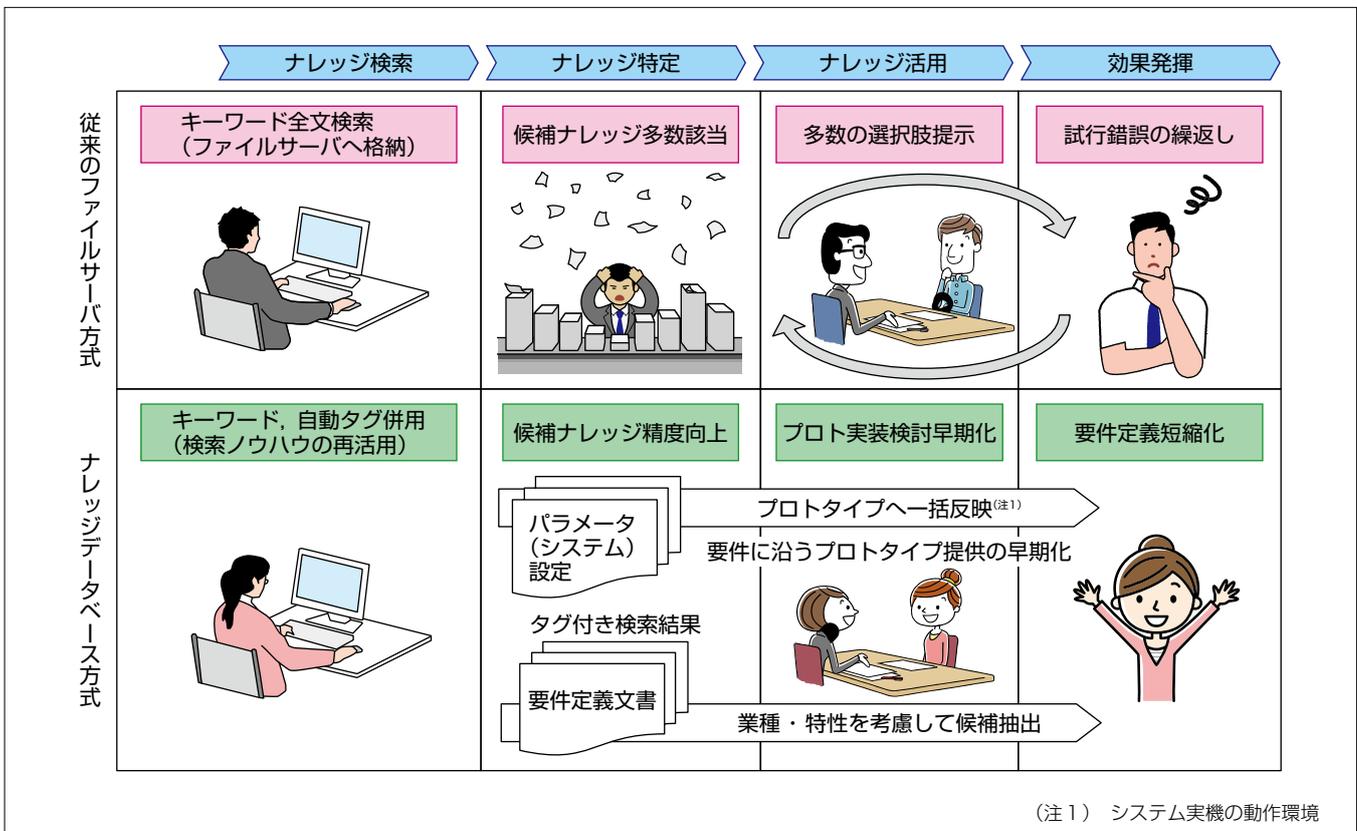
Approach to Structuring and Reusing Knowledge to Make System Integration Process Efficient

要旨

経済産業省による“デジタルトランスフォーメーション(DX)レポート”⁽¹⁾では、企業は競争力維持・強化のために、DXをスピーディーに進める必要がある一方で、“既存システムの老朽化”“IT技術者の人材不足”等が課題に挙げられている。三菱電機インフォメーションシステムズ株式会社(MDIS)では基幹システムパッケージ(PKG)を用い、プロトタイプを活用したシステム構築(SI)でユーザー企業のDX推進に貢献しているが、過去事例のノウハウは暗黙知にとどまり、その効率的な再利用が課題であった。そこでMDISでは、過去事例の蓄積文書を構造化して形式知として再利用するために、機械学習による自動化技術を活用したナレッジデータベースの試作と評価を行った。ナレッジデータベースは、登録機能、検索機能とともに、自動タグ

付与機能を持つ。自動タグ付与では、文書テキストを形態素解析して数値ベクトル化することで、入力文書の内容と最も関連性の高い文書の推定タグを付与する機械学習モデルを適用した。このナレッジデータベースを用いて基幹システムPKGを用いたSIの要件定義プロセスを模擬した結果、従来のファイルサーバ方式(過去事例の文書階層を再利用するため整理することなく全文検索して参照)に比較してシステム生成(プロトタイプ)の効率が大幅に向上した(従来比で85%の改善)。

DX推進での課題に対し、ナレッジデータベースの活用によってノウハウ再利用とSE(System Engineer)育成の底上げを行い、プロジェクト品質・コスト・納期の改善を実現し、SIプロセス効率化を図っていく。



SIの要件定義プロセスでの知識の再利用方式の比較

ナレッジデータベース方式は、過去事例の暗黙知を構造化して形式知にすることで再利用を図るが、そこにSEの労力を充てることを削減するため、機械学習による自動化技術を活用する。ルール型自動タグ付与機能と学習型自動タグ付与機能によって、原則全ての文書にタグ推定を行い、推定タグを付与する。SIプロセスで経験の浅い技術者でも過去事例の蓄積文書を再利用するため、タグ検索と従来の全文検索とを併用することにした。

1. ま え が き

経済産業省による“デジタルトランスフォーメーション(DX)レポート”⁽¹⁾では、企業は競争力維持・強化のためDXをスピーディーに進める必要性があるとする一方で、経営戦略としてDXを実践する上で“既存システムの老朽化”、“IT技術者の人材不足”等が課題に挙げられている。とりわけ、国内企業では、2025年に既存基幹システムの6割が21年超の運用を迎える中で各企業の基幹システム刷新が進まない状況から“既存ITシステムの崖(2025年の崖)”として警鐘が鳴らされている。

MDISでも、ユーザー企業の基幹システム構築(SI)や刷新を推進する上で、人的リソースの育成と生産性向上を体系的に行っていく必要がある、SIプロセス効率化の技術的な底上げを行ってきた。

本稿では、その手段として、機械学習を活用したナレッジとノウハウの共有を焦点にした知識構造化基盤を構築し、生産性の向上に向けて実施した取組みについて述べる。

2. MDISのSIプロセスでの課題と技術的解決の観点

2.1 SIプロセスでの人的ノウハウの課題

ユーザー企業がDXを実行してシステムを刷新する上では全体最適化の視点が不可欠であるが、個別最適化された事業部門の既存システムがある中で、刷新システムの要件決定が必要になる場合が多い。これらのSIプロジェクトでは、要件定義工程での検討が不十分である場合、後続の設計工程や開発工程、また成果物の品質・コスト・納期に影響を及ぼすため、上流工程の品質確保が重要になる。

MDISがSI事業を推進する中でこの点は事業分野に関わらず共通的な要素であるが、MDISの製造業向けSI事業では基幹システムPKGを用い、要件定義でプロトタイプを実施することで要件工程での齟齬(そご)を防止している。現状でも、プロトタイプの土台に過去のSI事例のノウハウを集約した標準モデル(テンプレート)を用いているが、ユーザー企業別の要件に対するPKG機能適用では、まだベテランSEのノウハウに依存する割合が高い。これから加速するユーザー企業のDX推進に貢献するためにはIT人材の確保と並行して過去事例でのノウハウの効率的な再利用が必要であるが、ノウハウの多くが暗黙知であることが課題であった。

2.2 SIプロセスの課題に対する技術解決の観点

MDISがSIプロセス効率化に向けてノウハウの再利用を

加速させるため、要件定義プロセスの構成要素を四つに分けて(図1)、各要素の課題を検討した結果、過去事例で生成した文書が十分に再活用されていない問題が明らかになり、この問題の解決を行った。

既存の文書管理システムで、過去文書の再利用を阻害する要因の一つは“入力情報”整理の労力にあった。SIプロジェクトでは各プロセスで固有ノウハウ、知識を含む多くの文書を作成するが、ノウハウ、知識共有のためプロジェクト完了後に改めて有用な文書に絞って文書属性を含め情報整理をするためには多大な労力を要する。その労力が一つのハードルになり、ノウハウ、知識を含む文書の格納及び共有がなされないまま暗黙知にとどまるケースがあった。このため、入力情報としてのSIプロジェクトノウハウを含む文書の作成、登録にできる限り手間をかけないようにすることを第一の観点にした。

次の観点は、全文検索機能だけでは知識、ノウハウが十分に再利用できないことである。MDISでは既に文書管理システムは持っており、文書区分に応じた階層と共通の属性管理を行っている。しかし、事業特性に応じた固有知識の再利用では、全文検索機能によるキーワード指定が手掛かりになるため、目的の情報をたどるためのキーワード指定自体も均一的ではなく、ユーザー企業別の用語の揺らぎや、PKGシステム機能固有の用語・略語の考慮が必要になる。これらに対応するため、現状は利用する技術者(SE)が過去事例を踏まえたノウハウを活用してキーワード指定をしており、IT人材の確保を行う中で事業経験が浅いSEには、正しくキーワード指定を行うこと自体が難しいケースがあった。そこで、手間をかけず蓄積された文書を知識として再利用するため、全文検索以外の属性を自動付与することを第二の観点として設定した。

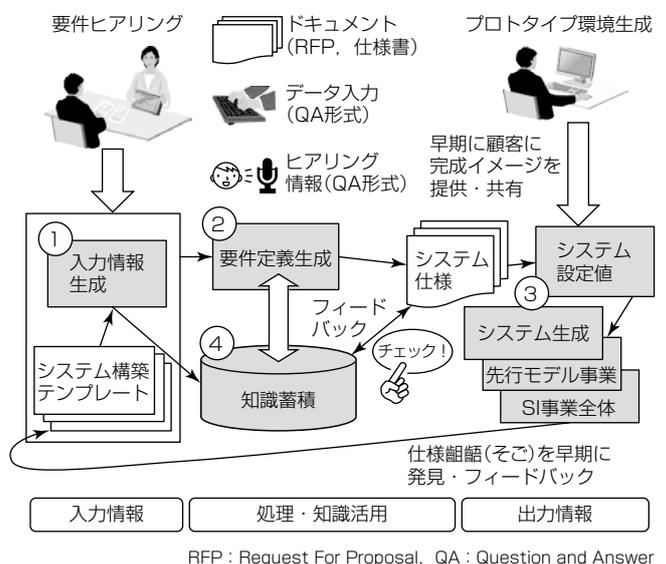


図1. SIプロセスの構成要素(要件定義)

最後に、“知識蓄積”に対する属性の網羅性と精度の問題である。既存の文書管理システムでもSEが一部の属性管理を行っているが、様々な文書に各SEが人手で属性を付与するため文書内容と属性を対応付ける基準が一定ではなく、蓄積された文書を属性情報から検索しても合致しない、又は属性自体が正しく付与されていないケースがあった。このため、文書に対する属性の自動付与では、文書内容から一定基準で判定を行うこと、全ての文書に属性を付与すること、及び誤った属性にはフィードバックを反映できることを第三の観点として設定した。

3. 機械学習を活用したナレッジデータベース

SIプロセスでの課題をITで解決するため、過去事例の暗黙知を構造化して形式知にすることで再利用を図るが、そこにSEの労力を充てることを避けるために自動化技術を適用する。MDISでは、SIプロセスの自動化に対し、過去事例の知識を蓄積・再利用する文書管理向けに三菱電機類似文書検索技術を土台にして機械学習による自動化技術を適用したシステムを開発した。これを、“ナレッジデータベース”と称する。この章では、ナレッジデータベースのシステム機能について述べる。

3.1 ナレッジデータベースの機能構成

従来は開発案件ごとに文書を共有サーバ上に蓄積し、全文検索機能を活用してきた。ナレッジデータベースでは、開発案件ごとに蓄積された提案資料や仕様書などの文書をサーバ上に一括格納する際に機械学習で自動的に構造化を行う。すなわち、文書検索で利用頻度の高い“開発工程”や“業務区分”などの属性を機械学習によってタグ情報として文書に付与し、タグ検索と従来の全文検索とを併用する構成にして、経験の浅いSEがプロジェクトの文書を検索・再利用することを可能にした。

ナレッジデータベースは、“登録機能”“検索機能”“自動タグ付け機能”から構成される(図2)。

3.1.1 ナレッジデータベース登録機能

登録機能では、まず各プロジェクト文書をプロジェクト完了時に文書一覧(ファイルリスト)に一括登録を行う。次にファイルリスト内の文書からテキストを抽出(Excel^(注2)ファイルはシート別に抽出)し、テキストを全文検索システムの索引に登録する。また、ファイルリストにはルール型自動タグ付与ツールによってタグ検索用の索引データを作成し、さらに学習型自動タグ付与機能を用いてタグ検索用の索引データを補完する。

(注2) Excelは、Microsoft Corp.の登録商標である。

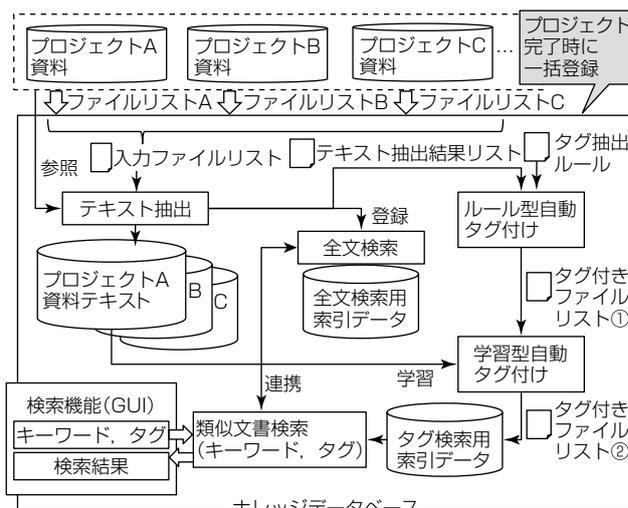


図2. ナレッジデータベースの機能構成

3.1.2 ナレッジデータベース検索機能

検索機能は、複数端末での利用に向けてナレッジデータベースサーバからクライアントへ分離した。検索機能はGUI(Graphical User Interface)を備え、GUIに対しタグ検索用のタグ情報と全文検索用のキーワードを入力する。この際、タグ情報はプルダウンメニューから選択(カスタマイズ機能を提供)する。入力に基づき、ナレッジデータベース内で全文検索結果とタグ検索結果を統合し、スコア上位の検索結果を表示する。なお、4章で述べる評価の検索GUIには、検索結果の優劣性評価記録のため評価ダイアログを追加した。

3.1.3 ナレッジデータベース自動タグ付与機能

自動タグ付け機能は、“ルール型自動タグ付与機能”と“学習型自動タグ付与機能”の二つの処理を行う。

“ルール型自動タグ付与機能”では、ファイルリストの文書に対してファイル名やファイルパスから、その属性を特定可能な文書にだけ自動タグを付与する。

一方、“学習型自動タグ付与機能”では、事前処理でサンプルの文書とタグのセットを学習データにして、統計的に機械学習モデルを生成する。機械学習モデルは、文書テキストを形態素解析して文書を数値ベクトル化することで、入力文書の内容と最も相関性の高い文書の推定タグを出力する。ファイルリストの文書登録時に、対象ファイルを機械学習モデルへの入力にすることで“ルール型自動タグ”を補完し、機械学習によって推定した“学習型自動タグ”を付与する。

4. SIプロセスへのナレッジデータベース適用と試行評価

ナレッジデータベースがSIプロセスでノウハウ再利用

と生産性向上に寄与するか、基幹システムPKG(SAP^(注3))を用いたSIの要件定義プロセスを模擬して評価試行を実施した。

4.1 ナレッジデータベース試行評価の概要

MDISの経験の浅い若手SE 3名でSAP業務アプリケーションの要件定義を模擬する作業に対して、ファイルサーバ方式(過去事例の文書階層を再活用のため整理することなく全文検索し参照)と、ナレッジデータベース方式の両環境を準備し(表1)、また、要件定義プロセスの模擬にはMDISが基幹システム構築で用いる“SAP ERP^(注3)(Enterprise Resource Planning)テンプレートMELEBUS”のモデルシナリオのうち販売管理(8シナリオ)と在庫購買管理(8シナリオ)について各31種類のパラメタ設定を準備した。各被験者はモデルシナリオと利用方式を両方被験する組合せで試行し、一定時間内で作成できたプロトタイプシステム生成(パラメタ設定)の件数を比較した。

両利用方式とも過去10プロジェクトの文書計18,652文書(Excelシート別計66,181件)を蓄積し、被験者は各12時間ずつ試行した。

(注3) SAP, SAP ERPは、SAP SEの登録商標である。

4.2 ナレッジデータベース試行評価結果

今回、各被験者のパラメタ設定数を基に達成率(実績件数/全対象31件)を算出した。その結果、被験者3名共にナレッジデータベース方式(提案方式)の達成率がファイルサーバ方式(従来方式)の達成率を上回る結果を得た(図3)。また、3名での平均値は従来方式での41%に対し提案方式では76%になり、作業効率を85%改善できた。

この要因として、ルール型自動タグ付与機能で全体の30~70%(タグ区分による)、さらに学習型自動タグ付与機能で96%(テキスト化できない文書以外の全件)にタグ

表1. 被験者ごとの利用方式及び対象業務

被験者(SE)	利用方式	
	ファイルサーバ方式	ナレッジデータベース方式
A	販売管理	在庫購買管理
B	在庫購買管理	販売管理
C	販売管理	在庫購買管理

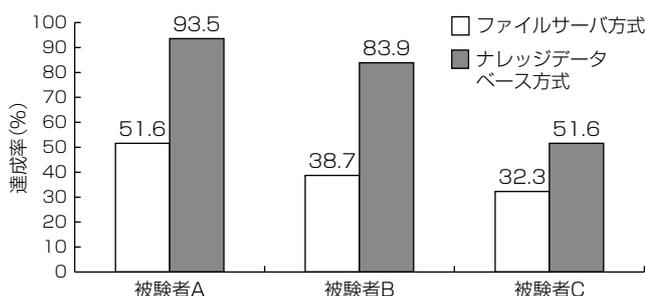


図3. パラメタ設定の達成率の比較

付与ができたこと、また、評価時の検索ログから検索結果の有用率を算出した結果、20位以内に目的文書が含まれる割合は98%、5位以内に含まれる割合は66%であったことが挙げられる。

4.3 ナレッジデータベース活用で期待される効果

ナレッジデータベースの活用によって、経験の浅いSEでも従来のファイルサーバ方式以上にプロトタイプ生成のパラメタ生成の生産性を大きく向上させることができる効果が得られた。SIプロセスでは、上流工程の要件定義でのプロトタイプが“標準化への誘導(PKG標準の適用度向上)”“課題解決日数短縮”に有効であることから、要件定義の品質向上に大きく寄与する。また、プロトタイプ活用がPKG活用のSIプロセスで重要になる。さらに、過去事例から構造化された知識の再利用は経験の浅いSEの技術の底上げに寄与し、ベテランSEへの依存度を下げることになる。結果として、DX推進上の課題である“既存システムの老朽化”、“IT技術者の人材不足”に対して、ノウハウ再利用とSE技術者育成の底上げで貢献し、プロジェクト品質・コスト・納期の改善とともにSI事業でDX案件を同時並行する多重度(事業推進力)を高めることが可能になる。

5. むすび

SIプロセス効率化に向けた知識の構造化と再活用の取組みとしてナレッジデータベースを焦点に試作と評価を述べた。今後は、ナレッジデータベース自体への機能拡張として“関連文書推論機能”によって文書間の相関性を紐(ひも)づけて文書活用の加速を図るとともに、“検索結果フィードバック機能”で文書の自動タグの精度向上を計画している。また、ナレッジデータベース活用が進み、タグ情報の精度が高まるとともに、MDISが並行して取り組んでいるSIプロセスでのシステム生成(プロトタイプ)の自動化と連動させ、将来的には要件定義より上流の受注前工程から省力化されたプロトタイプ提供を狙う。

また、今回の試作と評価は基幹システムPKG(SAP)を用いたSIを対象にしたが、タグ自動付与を行う機械学習モデルを生成する開発環境をクラウド上に整備し、他SI事業でも活用の目的に応じた“学習型自動タグ付与機能”が活用できるように汎用化も目指す。これによって、ユーザー企業のDX推進に対する具体的なイメージを早期に示し、顧客への一層の貢献を行っていく。

参考文献

- (1) 経済産業省：DXレポート
https://www.meti.go.jp/shingikai/mono_info_service/digital_transformation/20180907_report.html