

情報探索者の探索行動分析に基づく情報探索効率化のためのメタデータ体系化技法の開発

Development of metadata systematization technique to improve search efficiency based on searching behavior

深尾浩¹ 岡田伊策^{1,3} 阿部雅昭² 笈田佳彰^{1,3} 齋藤稔¹ 稗方和夫³

Hiroshi FUKAO¹, Isaac OKADA^{1,3}, Masaaki ABE², Yoshiaki OIDA^{1,3}, Minoru SAITO¹, and Kazuo HIEKATA³

¹ 富士通株式会社

¹ FUJITSU LIMITED.

² 株式会社富士通システムズ・ウエスト

² Fujitsu Systems West Limited

³ 東京大学

³ THE UNIVERSITY OF TOKYO

アブストラクト: 社内実績資産の再利用によるスピード・コスト・品質の向上は重要であるが、情報に適切なメタデータが付与されていないため、情報探索者が有用な情報を発見するために多大な時間と手間がかかっている。未経験者から熟練者に至る様々な探索者毎に探索行動が異なり、探索時に必要なメタデータも探索者毎に異なることに着目し、情報に付与すべきメタデータを体系化する技法を開発した。本稿では、提案書作成時におけるケーススタディを行い、本技法の有効性について考察を行う。

1. はじめに

プロジェクト遂行を円滑に進め成功に導くためのポイントは、過去のプロジェクトの豊富な現場実績による実践知の再利用によるスピード・コスト・品質のさらなる向上である。そのためには、社内の技術情報の中から有用な情報を素早く探索・再利用できることが必要である。

しかしながら、社内の実績資産には適切なメタデータが付与されていないため、情報の探索者（SE）が有用な情報を発見するのに多大な時間・手間がかかるというムダが生じている。あるいは、再利用できる情報があるにも関わらず、それを発見できずに重複開発するというムダも生じている。結果として、それらのムダにより、スピード・コスト・品質の向上が達成できない。社内の実績ある資産の再利用による QCD 向上のサイクルを図 1 に示す。

このような社内における情報共有に際して増大する人間への負担を軽減するための、情報共有システムの研究が盛んに行われている[1][2]。これらの研究は、情報提供者側から見た、情報をどのように検索させるかという観点に基づいている。

これに対して、筆者は、社内技術情報の探索作業における情報探索者の行動に着目し、情報探索の目的と、その探索行動を分析・分類し、その行動モデルに応じた、最短・最速・効率的に情報を発見・再利用できるための、最適なメタデータの整備・体系化技法を開発した。本稿では合わせてその成果について論述する。

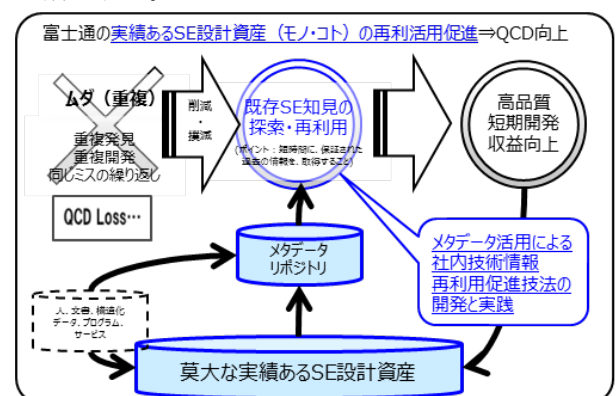


図 1 資産再利用による QCD 向上サイクル

2. 課題と背景

社内に存在する莫大な実績資産が再利用されてい

ない理由として、以下の課題・背景が存在する。

(1) 技術資産が未整理

大半の技術資産は、それを再利用されることを前提としておらず、再利用しやすいように整理されていない。資産にメタデータが付与されていない状態なので、どのような情報かは、その内容を直接見なければ判らない。そのため再利用しようとしても、有用な情報を探しだすためには多くの手間がかかる。

(2) 役に立たない中途半端なメタデータ

一方、実績資産の再利用を促進しようと情報にメタデータを付与する際、情報提供者視点からメタデータを付与すると、利用者を増やしたいあまり、「万人（初心者～熟練者）向け」といったメタデータを付与してしまう例が多く見受けられる。資産にメタデータを付与しても、その値で情報探索者が内容を区別できないメタデータでは意味をなさない。結果として、利用者を増やしたい意図とは裏腹に、本当の利用対象者が不明確になり、かえって再利用されないなど、適切でないメタデータが害になっていることも多い。

(3) 技術資産の蓄積・保有・継承が属人化

技術資産の再利用が進むように適切に整理するには、資産の内容が判るだけでは不十分である。その資産をいつ、どのタイミングで使うと効果的かを判断するには、熟練者の「経験」「知見」「勘」が必要である。社内における情報普及は熟練者の属人的な熱意に依存している。しかし熟練者自身がメタデータ付与を行うことは高コストに繋がる。また、その熟練者が現場に出ている間はメタデータの質の低下・再利用の停滞が起こる。そのため、熟練者の「経験」「知見」「勘」の分析・体系化・伝承が必要となる。

3. 解決施策

これらの課題を解決するためには、以下の施策を実施する必要がある。

(1) 社内技術資産を整理

莫大な技術資産を、そのファイルに内包されている整理に使える情報（タイトル、作者、日付、ファイル形式、存在場所、等）を用いたメタデータを付与し、分類・整理する。

(2) 探索者像に合わせた適切なメタデータ付与

技術資産の探索者を分析・分類し、それぞれの探索者像の行動様式に合わせた、探索者視点で探索・発見に必要なかつ適切なメタデータを整備、体系化して、付与する。

(3) 熟練者の「経験」「知見」を手順化

熟練者の持つ経験・知見に頼っていたメタデータ付与（識別・分類・適用）を細分・手順化する。セ

マンティック技術（機械学習による自動分類）を用い、(2)のメタデータを使って、自動付与する基準とロジックを作る。それにより、属人性を排除した、永続的な仕組みにすることが可能となる。

3.1. 具体的な解決技法

(1) 技術資産の「在り処」の特定

現場にとって真に再利用価値のある技術資産とその在り処を特定し、そのファイルに内包されている明示的な属性を自動抽出してメタデータとして付与する。そのメタデータにより資産を分類・整理でき、再利用資産を発見しやすくなる。

また合わせて、再利用価値の低い資産も選別し、削除することで、社内の資産全体の再利用価値の向上を図ることができる。

(2) 情報探索者像のセグメンテーションと行動のモデル化

情報の探索者が求める情報の6W2H*を分析し、求める情報を基点に探索者を科学的にセグメンテーションする。さらに、セグメント毎に探索者の探索行動をモデル化する。そのことにより、探索者の行動モデルの詳細段階に合わせた、探索者視点での適切なメタデータを網羅的に設定し、関係性を可視化してナビゲーションに役立てる。つまり、探索者の次の行動を予測し、先手を打って誘導することが可能となる。

探索者の「さがす」行動ログの分析から、さがす行動様式のパターンがあることが判る。さがす対象に対する熟練度が高い探索者はさがしたい物が「あるはずだ」という具体的イメージを持って「探索的」にさがす。一方、熟練度が低い探索者は何をさがせば良いか判らず「あるかどうかわからない」物を「探索的」にさがす。熟練度が増すにつれて、さがし方が「探索的」から「検索的」へと変化していくと考えられる。

* What, Who, When, Where, Why, for Whom, How, How much

(3) 熟練者の「経験」「知見」をロジック化・自動化

細分化・手順化した熟練者のメタデータ付与ロジックを収集し、かつ暗黙知的作業実績をリバーエンジニアリングする。セマンティック技術による自動分類・自動付与をシステム実装することで、付与コストを低減し、維持メンテナンスも容易になる。

3.2. 情報探索行動のモデル化

細分化したセグメント毎に、情報を探索する際の行動様式をモデル化した。

SEが情報を探索する行動は、以下のような流れと

表 1 セグメント毎の探索行動における障壁の分析

| 探索対象に対する 熟練度 情報 探索行動の流れ | <初級者> 何を知るべきか、よくわ かっている状態 | <初級者・中級者> 知るべきことに関して考え はあるが、なんと表現する かわからない状態 | <中級者・熟練者> あると想定される情報を 探している状態 | <熟練者> 以前見た情報を再度検 索している状態 |
|----------------------------------|--|---|-------------------------------------|---|
| 対象に関心を持つ／注意を 向ける | どこにどんな情報がある かも判らない。 | どこを探せば良いかが 判らない。 | どこを探せば良いかが 判らない。 | どこを探せば良いかが 判らない。 |
| 対象に関する情報を探す | 検索エンジンに入力す るキーワードが判らない。 検索結果を絞り込め ない。 | 検索結果を絞り込むの に適切なキーワードが 思いつかない。 | 認識や言葉の揺らぎで 欲しい結果が得られな いことがある。 | 正解の存在を知ってい るが、検索結果として 得られないことがある。 |
| 探索結果を比較・検討する ／選択する | 判断基準がないため、 検索結果を比較／検 討することができない。 | 検索結果を十分に絞 り込めないため、取捨 選択に時間がかかる。 | コンテンツを開かず判 断する情報が無い。 | ファイルを開いて判断す る時間がかかる。 |
| 選択した情報に従い行動に 移す／取り込む | 具体的にどう行動した ら良いか判断できない。 | 行動の結果の妥当性 を判断できない。 | 行動の結果の妥当性 を判断できない。 | 行動の結果をフィード バックする手段がない。 |

なる。

- ① 探索対象に関心を持ち、注意を向ける。
- ② 探索対象に関する情報を探す。
- ③ 探索結果を比較・検討し、選択する。
- ④ 選択した情報に従い行動する／情報を取り込む。

さらに、そのモデルの詳細段階でも探索行動を分析し、探索の鍵となるメタデータを抽出・特定した。セグメント毎に、探索を進めていく上で、どのような障壁が存在するかを分析したものを表 1 に示す。ここに示した障壁を回避するために有効なメタデータを整備することで効率的に探索が行える。

このように、探索者の行動に着目・予測することで、情報の探索をより柔軟・効率的にナビゲートすることが可能となる。

4. 「提案書作成支援システム」上での検証

ここまで述べた、情報探索者のセグメンテーションとその行動モデルによる分析を、プロジェクト提案のためのプレゼンテーション作成を支援するシステム[3][4]、「提案書作成支援システム」上で検証した。

4.1. 「提案書作成支援システム」の概要

SE の一般的な提案書作成の流れを図 2 に示す。

新規商談において提案書を作成する際、過去の提案書を再利用することで効率化を図る。しかし提案内容は商談毎に毎回異なるため、提案シナリオ立案から既存資料の探索・収集には多くの時間がかかり、提案書の品質を向上させる時間が十分にとれないという問題が起きていた。「提案書作成支援システム」は過去の実績ある提案素材を効率的に探索・再利用するシステムである。

従来は提案書 (PowerPoint ファイル) を開いて必

要なスライドを取捨選択するため手間がかかっていたが、「提案書作成支援システム」ではスライド単位で探索できることで、素材収集の時間を大幅に短縮できる。

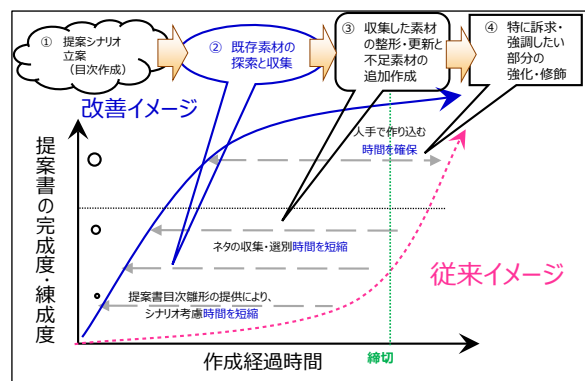


図 2 提案書作成の流れ[4]

「提案書作成支援システム」の概念図を図 3 に示す。

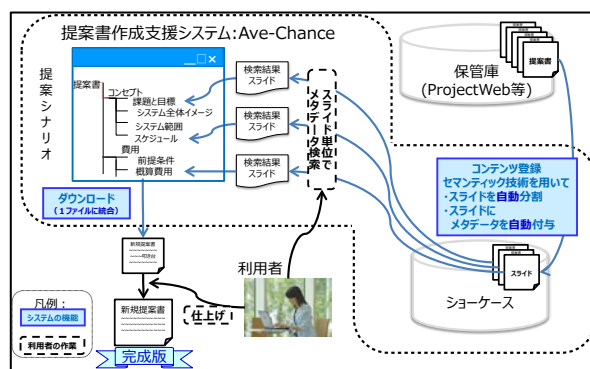


図 3 「提案書作成支援システム」概念図[4]

4.2. 提案書素材の整理

検証を行った環境には、約 4,000 スライドの提案書素材を格納した。その際、各スライドの属性 (タイトル、作者、日付、提案書提出先企業名、業種、

売上高、等)を、セマンティック技術を用いて抽出し、スライドのメタデータとして自動で付与した。自動付与したメタデータの一覧を表2に示す。これは実際に提案書を作成する現場SEの探索行動を分析し、探索に有効なメタデータを選択したものである。

表2 自動付与した主要メタデータ一覧

| | |
|-------|----------|
| タイトル | 顧客の売上高 |
| 頻出単語 | 作成日 |
| 作者 | 最終更新日 |
| サイズ | 類似色スライド |
| 顧客名 | 類似形スライド |
| 顧客の業種 | オリジナル提案書 |

4.3. 提案書作成者のセグメンテーションと行動モデル

「提案書作成支援システム」での検証は、5名の現場SEに実際に提案書を作成してもらい、メタデータの効果を測定した。そのうち、熟練者SEと初心者SEの行動の相違を表3に示す。熟練者SEと初心者SEのプロフィールは以下の通り。

- 熟練者：提案書作成に関して熟練者のSE。提案の技術的内容についても熟知しており、過去の提案書素材についてもある程度当たりがつけられる。自律的に提案シナリオを作成できる。
- 初心者：提案書作成に関しては初心者レベルのSE。提案の技術的内容に関してはある程度理解しているが、過去の素材については判らない。

与えられた提案シナリオに従って素材を探索できる。

2名の被験者それぞれが、実際に「提案書作成支援システム」を用いて素材収集を行った際の探索行動を分析したところ、以下のような違いがあることが確認できた。

熟練者の探索行動を分析すると、提案シナリオに沿ってどのようなスライドが必要か、どのような素材があるかある程度判っているため、「検索的」な探し方を行っており、キーワード選択や、複数キーワードでの絞り込みが的確である。そのため、検索回数や検索ヒット数は少ないが、短時間で多くのスライドを収集できている。

これは、熟練者の情報探索行動として、素材の在り処は覚えていないが、過去に作成した提案書の特徴的なキーワードや、スライドのイメージを覚えているため、それを基に探すことができる。実際の検証において、過去に提案した「顧客名」やスライドの「頻出単語」、またスライドのサムネイル表示と「類似形スライド」などのメタデータを有効に利用して、容易に欲しいスライドを見つけることができた。

一方初心者の場合、提案シナリオに沿った適切なキーワード選択ができず「探索的」な探し方を行っている。そのため、ヒットしない検索回数が多くなっている。また、多くヒットした検索結果から適切に絞り込み検索ができず、拡大表示回数も多くなっている。

これは、初心者の情報探索行動として、どこにどのような情報があるかが判らず、過去の素材として何があるのかも判らないため、多くのキーワードを

表3 検索行動と探索行動の相違

| | 熟練者 | 初心者 |
|--------------------|--------------------|-------------------------|
| システム利用時間 | 14分 | 20分 |
| 検索回数 | 15回 | 25回 |
| 探し方 | 検索的 | 探索的 |
| 検索キーワードの思いつき方 | 提案シナリオに沿って自発的に思いつく | 目次項目名やメタデータ一覧から思いつく |
| 有効なメタデータ | 顧客名、頻出単語、類似形スライド | 頻出単語、顧客名、顧客の業種、オリジナル提案書 |
| 検索ヒット数 | 857スライド | 2,278スライド |
| ヒット数0の検索回数 | 1回 | 8回 |
| スライドの拡大表示数 | 90スライド | 114スライド |
| スライドのダウンロード数 | 8提案書から 16スライド | 9提案書から 10スライド |
| 拡大表示数に対するダウンロード数比率 | 17.8% | 8.8% |
| 従来手法での収集時間(予測) | 2時間 | 5時間 |
| 時間短縮効果 | 8.57倍 | 15倍 |

試して、多くの検索結果を端から吟味して選択することを示している。実際の検証において、最初に目次項目名そのもの、あるいはそこに含まれる単語をキーワードとして検索する例が多く見られた。そして、得られた検索結果スライドの「メタデータ一覧」から、「頻出単語」に含まれるキーワードや「顧客名」「顧客の業種」、さらに「オリジナル提案書」からスライド分割前の提案書そのものを見る行動が見られた。これはメタデータを絞り込むためではなく、一旦探索範囲を広げるために利用していることが判る。こうして与えられたメタデータから網羅的に手がかりを掴み、欲しいスライドを発見することができた。

このように、被験者によって、利用するメタデータ項目が異なったり、同じメタデータ項目でも利用の仕方が異なったりすることが確認できた。

被験者にヒアリングしたところ、従来の提案素材探索と比較して、熟練者で8.6倍、初心者では15倍もの時間短縮効果があったことが判った。探索行動は異なるのに、どちらも短時間で探索できたのは、それぞれの探索行動にマッチした探索手段、適切なメタデータが用意されていたからである。

5. 解決策の効果と今後の展開

5.1. 効果

技術情報の探索者をセグメンテーションし、その探索行動を予測することで、探索者が資産を探索する際の障壁は何かを想定することができる。そして、その障壁を回避するために必要なメタデータをあらかじめ想定し、提供することで、資産の探索・再利用の効率化が図れることを「提案書作成支援システム」における検証で実際に示すことができた。

メタデータを付与するにあたっては、セマンティック技術による自動抽出、自動付与により、人手でメタデータを付与することに比べ、大幅にコストを削減できた。

そして、メタデータを用いた探索により、効率的に対象を絞り込むことができ、素早く欲しい情報にたどり着くことができた。また、適切な探索キーワードを思いつかない初心者でも、資産に付与されたメタデータを見ることによって、自分の知識にない他の関連・類似キーワードに気づくことができ、欲しい情報に効率的にたどりつくことができた。

このように、探索が効率化されることで、網羅的かつ短時間で素材の収集が可能になり、提案書の仕上げに十分な時間をかけることで提案書品質の向上を図ることができた。

以上のことから、資産に適切なメタデータを体系的に付与することで、提案活動のスピード・コスト・

品質に貢献できることが判った。

5.2. 今後の展開

今回、「提案書作成支援システム」を利用した提案書作成業務において、情報探索者のセグメントと探索行動の違いにより、探索に有用なメタデータも異なることが判った。今後は、情報探索者をより緻密にセグメンテーションし、セグメントに応じたメタデータ項目のバリエーションを増加させることで、さらなる探索の効率化実現を目指す。そのためには、情報探索者をより具体的にイメージする必要がある、ペルソナを用いて探索者像の具体化を図ることが有効であると考えられる。ペルソナを用いての検証が今後の課題である。

6. まとめ

今回筆者は、技術情報の活用最適化技法として、情報探索者のセグメンテーションおよびその行動モデルを分析することで、それが既存資産の再利用に効果的であることを、「提案書作成支援システム」を用いて検証した。

この技法は、提案書素材だけではなく、技術資産全般の再利用に広く適用可能である。その対象資産や対象業務毎に情報探索者像も異なり、それを分析・具体化することで、様々な分野において、スピード・コスト・品質の向上に貢献できると考える。

今後筆者は、提案書素材に留まらず、設計ドキュメントをはじめとした多くの技術資産に関し、その再利用を最適化するために、さらに開発を進め、貢献していきたい。

参考文献

- [1] 森本由起子, 間瀬久雄, 平井千秋, 阿部琢哉, 大野治. システムエンジニア向け情報共有システムの開発. *Journal of the Society of Project Management* Vol.7, No.2, pp.40-45, 2005
- [2] 中山康子, 真鍋俊彦, 竹林洋一. 知識情報共有システム(Advice/Help on Demand)の開発と実践. 「インタラクティブ・97」, pp.103-110, 平成9年2月
- [3] 稗方和夫, 大和裕幸, 笈田佳彰, 岡田伊策, 齋藤稔. プレゼンテーション作成支援システムの開発. SIG-KST-2011-03-01 (2012-03-01)
- [4] 岡田伊策, 齋藤稔, 松岡伸治, 笈田佳彰, 大和裕幸, 稗方和夫. プロジェクト提案のための文書情報管理システムの開発と実用化. *The 27th Annual Conference of the Japanese Society for Artificial Intelligence*, 2013