3C2-OS-13b-5

標準化活動者へのメタデータ付与による活動の見える化と合理的な後任候補者選定

Appropriate selection of a successor by visualization of the relationship by applying meta-data to the standardization activities and members.

岡田 伊策[*1] 齋藤 稔[*1] 大和 裕幸[*2] 稗方 和夫[*2] 笈田 佳彰[*2] 三浦 慎也[*3]
Isaac OKADA, Minoru SAITO, Hiroyuki YAMATO, Kazuo HIEKATA, Yoshiaki OIDA, and Shinya MIURA

[*1] 富士通株式会社共通技術本部ナレッジ推進統括部
[*1] System Engineering Knowledge Improvement div., SYSTEM ENGINEERING TECHNOLOGY UNIT, FUJITSU LIMITED.
[*2] 東京大学大学院新領域創成科学研究科
[*2] Graduate School of Frontier Sciences, THE UNIVERSITY OF TOKYO
[*3] 東京大学工学部システム創成学科
[*3] Faculty of Engineering, THE UNIVERSITY OF TOKYO

A model system which can visualize the relationship between knowledge and actual experiences in the Fujitsu's standardization activities has been developed. By applying meta-data to profiles and annual activity-reports of members who are engaged in the standardization activities, the relationship between the profiles and the annual activity-reports has been visualized as RDF graph. It facilitates nominating the successor who has similar knowledge and experience to current members, reasonably and appropriately.

1. 緒言

企業において、蓄積されるリソースは増加の一途をたどり、莫大な量のリソースが社内 DB に存在する。各リソースは、リソース種別（人、モノ、文書など）毎に個別に管理されることが一般的で、リソース間の関係は管理できていないことがある。

例えば、富士通の対外標準化活動では、標準化活動従事者の『名簿 DB』と、標準化活動従事者による『年次活動レポート DB』は、各々独立した DB として蓄積・管理されてきた。

異なるデータソース間のリンクを利用した研究事例としては、鹿島[鹿島 2007 年]や松村ら[松村 2011 年]の研究がある。鹿島の研究はネットワーク構造解析（リンクマイニング）によるリンクの予測であり、松村らの研究は博物館情報と地域情報の連携活用である。本研究は、それらと比して、企業内の実データにおいて、限定されたビジネスニーズに特化して、有効な結果を得ようという点が異なる。

一般的に標準化活動従事者の交代時は、前任者と知識・実績・人脈面で類似性・共通性の高い後任者が選任される。この選任を DB データの有機的関係性を活用して合理的・効率的に支援できれば、後任者の適切性・正当性・妥当性を客観的に高められる。

そこで、メタデータ技術を用いて、リソース群間の紐付けと関係性の『見える化』を、課題解決の糸口とした。（図 1）

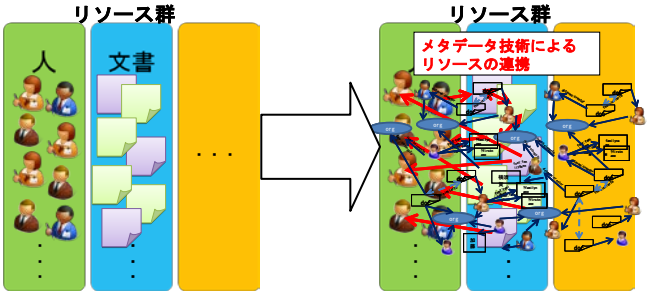


図1 メタデータ技術によるリソース間のリンク付け

連絡先：岡田伊策、富士通株式会社 共通技術本部 ナレッジ推進統括部、東京都大田区新蒲田 1-17-25 富士通ソリューションスクエア

構成要素間のリンクの『見える化』で、類似の関係性を検索・発見して、同類リソースの発見効率が向上すると想定した。

具体的には、前述の対外標準化活動の『名簿』（人）と『活動レポート』（文書）リソースにメタデータを付与して、同一メタデータが DB をまたがってリンクしている関係性を『見える化』した。

本論文では、『見える化された関係性そのもの』を検索できるプロトタイプシステムを開発して、『関係性』をクエリにして検索すれば、限定されたビジネスニーズにおいては、同類リソースの発見が容易になり、かつ選択結果の妥当性も『見える化』で共有できることを示すことを目的としている。

2. 課題解決のためのアプローチ手法

2.1 検証対象としたサンプルデータ

今回の検証では、以下の 2 つの DB を対象にした。

① 対外標準化活動従事者『名簿 DB』（人）：
DB 内容：氏名／所属／標準化活動分野／団体、等

② 対外標準化活動年次レポート（文書）：
DB 内容：文書（Word／PDF；執筆者名／団体／等）

従来は、これらの DB は、物理的にも管理上も別々の DB として管理されており、相互には連携していなかった。

（注）本研究では、元データの氏名を仮名に置換して実験した。

2.2 メタデータの設計

名簿（人）に対しては「FOAF」（The Friend of a Friend project:http://www.foaf-project.org/）を、標準化活動レポート（文書）については「Dublin Core」を参考にして、メタデータを設計した。さらに、本システムにユニークなメタデータを「fkw」として追加して設計した。本システムにユニークなメタデータは「fkw:belongsTo」（所属している標準化団体）、「fkw:similarTo」（キーワードを元にした類似文書）などである。（表 1）

表1　設計したメタデータの一覧

人に対するメタデータ（標準化活動者名簿より）

メタデータフィールド（プロパティ）	説明	定義域（rdfs:domain）	値域（rdfs:range）	メタデータバリュー例
foaf:familyName	姓名の「姓」。	foaf:Person	rdfs:Literal	富士
foaf:firstName	姓名の「名」。	foaf:Person	rdfs:Literal	太郎
fkw:belongsTo	所属している標準化団体。	foaf:Person	fkw:StdOrganization	<http://intap.org>
fkw:stdOrgOfficialPosition	所属する標準化団体における役職。	foaf:Person	rdfs:Literal	メンバ
fkw:company	所属する会社名	foaf:Person	rdfs:Literal	富士通株式会社
fkw:country	国籍	foaf:Person	rdfs:Literal	日本

文書に対するメタデータ（標準化活動レポートより）

メタデータフィールド（プロパティ）	説明	定義域（rdfs:domain）	値域（rdfs:range）	メタデータバリュー例
fkw:similarTo	キーワードを元にした類似文書	fkw:Document	fkw:Document	<http://id103.doc>
dcterms:creator	執筆者	fkw:Document	foaf:Person	<http://id96.per>
dcterms:subject	テーマとする標準化活動	fkw:Document	fkw:StdOrganization	<http://intap.org>
fkw:fileName	ファイル名	fkw:Document	rdfs:Literal	ActRepo.pdf
fkw:publishedcomany	所属する会社名	fkw:Document	rdfs:Literal	富士通株式会社

標準化団体に対するメタデータ（標準化活動者名簿より）

メタデータフィールド（プロパティ）	説明	定義域（rdfs:domain）	値域（rdfs:range）	メタデータバリュー例
fkw:orgName	団体名	fkw:StdOrganization	rdfs:Literal	INTAP

2.3　メタデータの付与

設計したメタデータは、標準化活動従事者の氏名など、DB が明示的に保有している情報については、機械的に付与した。

一方、特に標準化活動レポート（文書）については、効果的なメタデータ付与が、名簿との関係性を強固にすると想定して、特に以下の 2 点を工夫した。

(1) 活動レポート間の類似性

類似性の高い文書の執筆者は同様の知識・経験を保有している可能性が高い。具体的には、松尾ら[3]の「語の共起の統計情報に基づくキーワード抽出手法」と逆文書頻度を組み合わせた。

これにより抽出できた 10 個のキーワードについて、集合の類似度を評価するシンプソン係数を用いて評価し、閾値を超えるものを類似文書とした。（図 2）

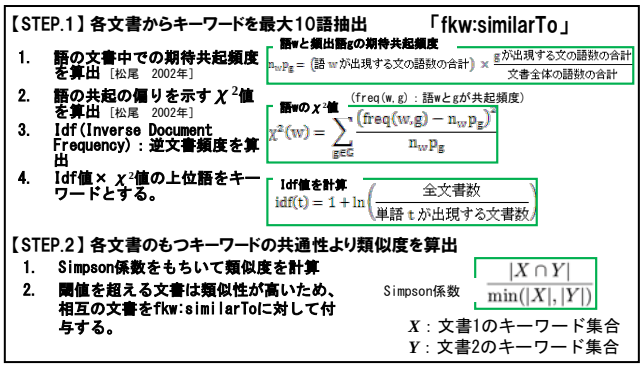


図 2　文書間の類似性の判定

(2) 活動レポートがテーマとする標準化団体名の付与。

名簿から標準化団体名を辞書化、活動レポートに頻出する上位 3 つの団体名をメタデータとして付与することにより、活動レポートが主題とする標準化活動と、その類似・近似の活動の発見を容易にした。

2.4　メタデータの RDF 記述

付与したメタデータを RDF で記述することにより（図 3）、標準化活動従事者と年次標準化活動レポートを RDF グラフ表現でき、DB をまたがった関係性（データリンク）を『見える化』できるようになった。（図 4）

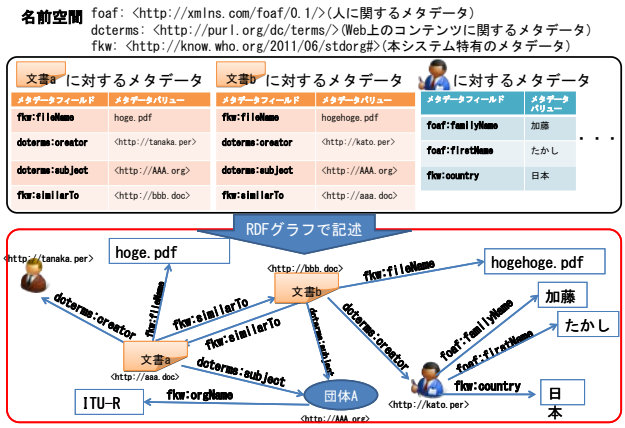


図3　メタデータ付与と RDF 記述

この RDF グラフ群（すなわちオントロジー化された両 DB の構成要素群）に対して、SPARQL クエリを与えることにより、同一のグラフパターンを持つ部分が容易に抽出できるようになった。

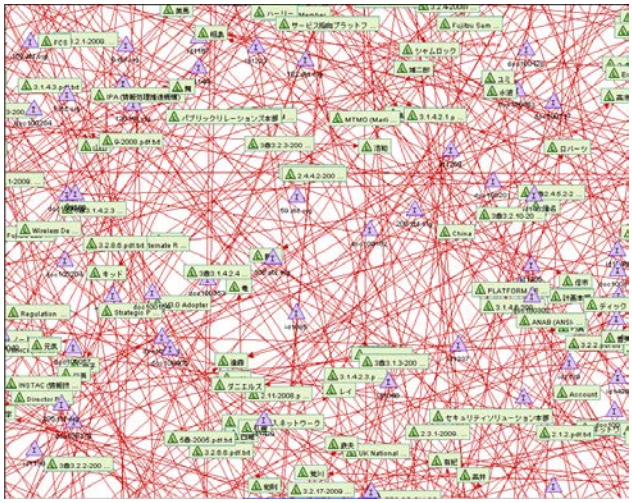


図4　RDF Gravity (RDF Graph Visualization Tool) による可視化

2.5　標準化活動従事者交代時の後任者選定論理の仮定

従来、ある標準化活動従事者が転任などで交代する場合、その後任者は、当該従事者の同僚が慣習的に選任される傾向が強い。結果的に、前任者と類似のナレッジおよび人脈を保有しているので、妥当な判断となる。

この慣習的な後任選任論理を、当事者ヒアリングに基づいて、以下の 5 つのクエリとして仮定し（表 2）、SPARQL で記述した。

表2　同様の技術分野を持つ人を検索するためのクエリ

クエリ	内容
Q1	前任者と同一の標準化団体に属している
Q2	前任者が執筆した活動レポートのテーマとなる標準化団体に属している
Q2'	前任者が加わっている団体をテーマとして活動レポートを書いている
Q3-1	前任者が執筆した活動レポートと類似した*レポートを執筆している
Q3-2	前任者が執筆した活動レポートと類似した**レポートを執筆している

***同じ主題（dcterms:subject）を持つもつ文書を「類似している」とした。**

****類似（fkw:similarTo）で結ばれた文書を「類似している」とした。**

3. リンクされたデータ活用による有用性検証のケーススタディ

本システムを活用して、過去の後任選定実例と同様に、後任執筆者を発見できるかを検証することにした。

具体的には、

①前述の 5 つのクエリ

②対外標準化活動者『名簿 DB』(人)のうち約 550 名分

③『年次標準化活動レポート DB』(文書)のうち 5 カ年分

(抽出された対象社外標準化団体は、約 200 団体)

のメタデータを使用して、同一テーマで標準化活動従事者が交代したケースを選定。同一の結果を本システムが再現できるか実証した。

3.1 正解ケースの選定

5 年間で実際に標準化活動レポートの執筆者が交代した実績を表 3 に示す。

表3 標準化活動レポートのテーマを引き継いだ5名

No	前任者	後任者	レポート内容	年次
1	石川高志	木下京子	OMA	2005 → 2006
2	石川高志	村山ゆず季、羽田紀代	ITU-R	2005 → 2006
3	木下京子	尾崎澄江	OMA	2006 → 2007
4	菊地ミノル	横須賀成寛	INTAP	2007 → 2008
5	羽田紀代	遠藤美月	ITU-R	2008 → 2009

3.2 仮定した後任選任論理クエリによる探索

前述の 5 つのクエリによる検索を、全体候補者 DB(550 人)に対して実行することにより、

① 全体候補者の中から該当者を何人に絞り込めたか

② 絞り込まれた候補者の中に、実際の後任者は存在するかを判定した。

3.3 結果

対象 550 人のデータに対して、5 つのクエリを 15 パターンの組合せで実行した。後任候補者を選別して 10 人以下まで絞り込むことができた結果が、表 4 である。

表4 前任者 5 人に対する、クエリの組合わせと約 550 人の候補の中から絞り込まれた結果

No	前任者名	Q1	Q2	Q2'	Q3-1	Q3-2	Q1∩Q2	Q1∩Q2'	Q1∩Q3-1	Q1∩Q3-2	Q2∩Q2'	Q2∩Q3-1	Q2∩Q3-2	Q2'∩Q3-1	Q2'∩Q3-2	Q3-1∩Q3-2
1	石川高志	46	44	19	37	12	27	4	3	5	3	6	5	17	6	9
2	石川高志	46	44	19	37	12	27	4	3	5	3	6	5	17	6	9
3	木下京子	98	81	24	38	11	48	6	3	6	3	8	6	18	6	9
4	菊地ミノル	166	138	18	55	21	71	7	1	13	1	5	3	8	6	13
5	羽田紀代	72	32	19	30	7	10	5	1	5	1	6	3	6	5	7

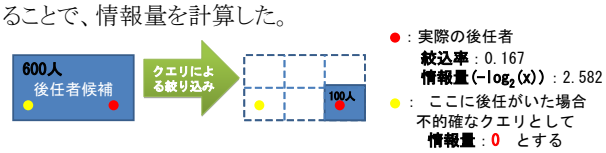
□ : 絞込結果の中に後任者が存在する。
■ : 絞込結果の中に後任者が存在しない。

全前任者に対していずれかのクエリを投げれば、10人以下にまで絞込が出来ている。

(1) SPARQL クエリの有用性の評価

図 5 の左上の四角は後任者候補の全体を示すイメージ図である。右上の四角はクエリにより後任者候補が絞り込まれた様子を示すイメージ図である。点線部分がクエリにより除外された後任候補者であり、青い部分がクエリにより絞り込まれた後任候補者である。赤い丸で示すように、青い部分に実際の後任候補者がいれば、クエリにより選択した情報量であるクエリの選択情報量がゼロ以上であり、黄色い丸で示すように点線部分にいればクエリの選択情報量はゼロとする。

次に、クエリの選択情報量の計算方法について説明する。クエリによって、絞り込まれた後任候補者の人数を分母として何分の 1 まで絞り込めたかということで絞込率を計算し、その log をとることで、情報量を計算した。



例)No.3（前任者：木下京子）の場合　　全549人

クエリ	Q1	Q2	Q2'	・・・	Q2∩Q2'	Q2∩Q3-1	Q2∩Q3-2	Q2'∩Q3-1
絞込数	98	81	24	・・・	3	8	6	18
絞込率	0.179	0.148	0.044	・・・	0.005	0.015	0.011	0.033
絞込精度	○	○	○	・・・	尾崎×	○	引継×	○
情報量	2.49	2.76	4.51	・・・	7.5157	6.10	6.52	4.93
不的確なクエリの除外	2.49	2.76	4.51	・・・	0	6.10	0	4.93

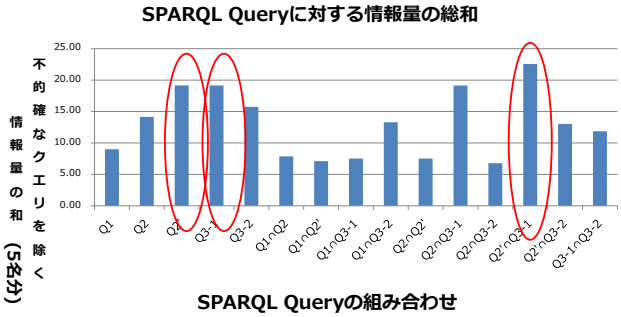
図5 クエリの持つ選択情報量の算出

(2) クエリの組合せと情報量の和

実際に引き継がれた 5 名の後任者について、情報量を合算すると、図 6 の通りとなる。赤い丸で囲んだクエリの選択情報量が大きいことから、Q2',Q3-1 のクエリの情報量が大きいことがわかり、そのアンド検索はさらに情報量が大きくなった。

つまり、本研究で設計したメタデータを用いたクエリのうち、この 2 つのクエリが後任者を捜すうえで最も適していたということになる。このようにクエリの選択情報量を評価することで、今回のシナリオで有効なクエリが Q2',Q3-1 であることを示した。

したがって、クエリの選択情報量は、SPARQL クエリを用いた検索インタフェースを設計する際に、「どのクエリに重みをつければ良いか」の指針となり得る。



クエリ	内容
Q2'	前任者が加わっている団体をテーマとして活動レポートを書いている
Q3-1	前任者が執筆した活動レポートと類似したレポートを執筆している

図6 情報量の和とクエリの組み合わせ

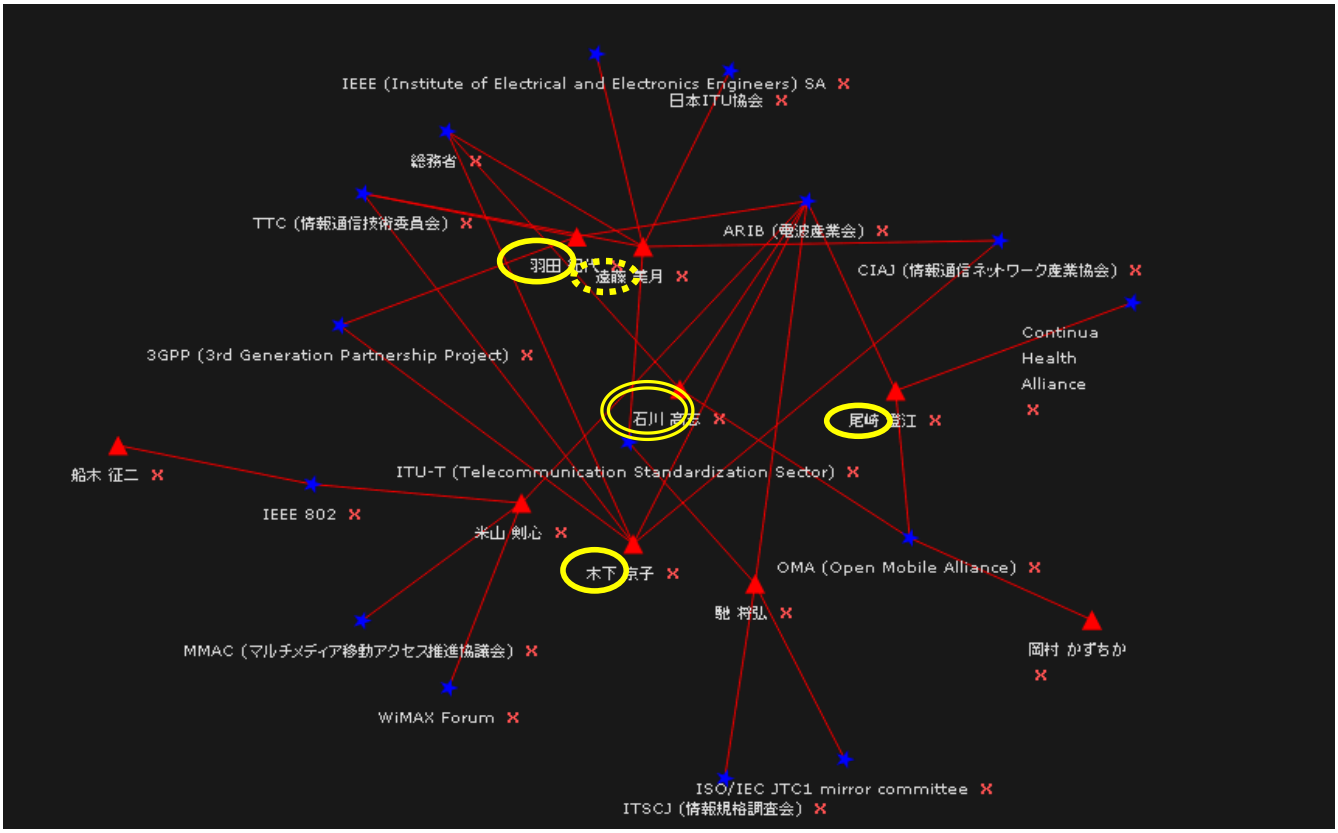


図7　他の後任候補の可能性

3.4 他の後任候補者の発見

さらに、データの関係性を『見える化』したことで、実際に選任された後任者の他にも、後任候補者を発見できた。

図 7 の例では、『石川』の後任には、実際に選任された『木下』『羽田』の 2 名以外にも、類似の複数の関係性を持つ『遠藤』『尾崎』も候補となりうることがわかる。

表 3 にあるように、後日、『木下』の後任に『尾崎』、『羽田』の後任に『遠藤』が選任されたので、図 7 の示唆は妥当である。

このように名簿（人）と、活動レポート（文書）の関係性を『見える化』したことにより、他の選択肢の提示も可能である。

4. 結論

異なる DB に対して、共通かつ横断的なメタデータを付与することにより、データのリンク関係を『見える化』し、その『見える化』されたデータリンクモデルを鍵に、類似性・共通性を、効率的に検索できるようになった。

このことにより、慣習と直感で選任されていた標準化活動従事者交代時の後任選出が、前任者との知識・人脈・実績の類似性・共通性の『見える化』に基づいて判断され、合理性・合目的性の合意形成が容易になった。

4.1 今後の課題

データのリンクの『見える化』による情報照会の効率化を一層促進するためには、同様のアプローチを公開データ全般に広める取り組みが必要である。

参考文献

[鹿島　2007 年]　鹿島久嗣: ネットワーク構造予測, 人工知能学会誌, Vol. 22, No. 3, pp. 344-351 (2007)

[松村　2011 年]　松村冬子, 小林巌生, 嘉村哲郎, 加藤文彦, 高橋徹, 上田洋, 大向一輝, 武田英明: Linked Open Data による博物館情報および地域情報の連携活用, 情報処理学会, 人文科学とコンピュータシンポジウム, じんもんこん 2011 論文集,2011(8),403-408 (2011-12-03)

[松尾　2002 年]　松尾豊, 石塚満: 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム, 人工知能学会論文誌, Vol. 17, No. 3, pp. 213-227 (2002)