

要求獲得のためのシソーラスの統合

加藤潤三¹ 佐伯元司² 大西淳³ 林晋平⁴ 海谷治彦⁵ 山本修一郎⁷

概要: 同一ドメインに属する複数の情報源から個別に構築した複数のシソーラスを1つのシソーラスに統合する方法を述べる。この統合したシソーラスは、要求獲得において多くの要求を獲得できるという点で有用であることを事例研究によって示す。

キーワード: 要求獲得, ドメイン知識, シソーラス

Integration of thesauruses for Requirements elicitation

KATO JUNZO^{†1} SAEKI MOTOSHI^{†2} OHNISHI ATSUSHI^{†3} HAYASHI SHINPEI^{†4}
KAIYA HARUHIKO^{†5} YAMAMOTO SHUICHIRO^{†6}

1. はじめに

対象ドメインのドメイン知識がない、または不足している要求分析者が要求を獲得した結果、ドメイン知識不足のため要求の欠落が発生する。THEOREE[1]は、要求分析者のドメイン知識をシソーラスによって補い要求を獲得することによって要求の欠落を減らすことを目的に開発した要求獲得手法である。

THEOREE で活用するシソーラスが持つドメイン知識の質と量が獲得した要求の欠落の量に影響する。我々は、

高品質かつ多量のドメイン知識を THEOREE で活用するシソーラスが持つように、我々は技術文書を情報源とするシソーラス構築支援(手法と支援ツール)[2]を開発した。このシソーラスの構築支援は、1つの情報源から1つのシソーラスを構築するための手法とツールを提供している。

1つの情報源で要求獲得に必要なすべてのドメイン知識が満たされることは一般には起こりえず、複数の情報源が互いに不足を補うことによって必要なドメイン知識が得られる。つまり、複数の情報源から構築された複数のシソーラスを統合したシソーラスを THEOREE で利用するための手法が必要である[2]。本論文では、[2]の手法を用いて構築した複数のシソーラスを統合する手法と、統合したシソーラスの効果の評価について述べる。

本論文の構成を述べる。本論文の目的を2章、本論文の理解を助けるために3章で THEOREE と THEOREE で使

用するシソーラスについて述べる。4章では、統合したシソーラスへの要求とその使い方を、5章で統合する手順を5章で述べ、6章ではシソーラスの統合事例および統合したシソーラスのドメイン知識は多くなったことを述べる。7章では、6章で述べた統合したシソーラスを用いて THEOREE で獲得を行った結果とその分析を述べ、獲得した要求は統合前のシソーラスで獲得した要求より多くなっていること確認した。当研究で得た知見を8章、解決できたことと残った課題を9章で述べる。

2. 研究の目的

シソーラスを統合することによって1つの情報源から構築したシソーラスよりもドメイン知識の量を増やすことができ、シソーラスを統合する手法はシソーラスのドメイン知識の量を増やす手段と位置付けることができる。シソーラスが持つドメイン知識の量が THEOREE で獲得した要求の数に影響する、ドメイン知識の量が多いほど獲得要求は多くなり、要求の欠落を軽減できる。

異なる情報源のドメイン知識を持つシソーラスは1つの情報源から構築したシソーラスより要求獲得に必要なドメイン知識の量が多いことと、ドメイン知識の量が多いシソーラスを THEOREE に活用すればより多くの要求を獲得できると仮定する。

情報検索の分野では、分野横断的な検索を行うためのシソーラス統合技術が開発されてきたが[3]、これらの手法は我々の要求獲得への活用といった目的とは異なっている。

本論文では、[2]の手法で1つの情報源から構築した1つのシソーラスを複数集め、これらを統合することによってドメイン知識を多く含むシソーラスを得る。このような手法で統合したシソーラスは、統合するシソーラスの数が多いほどドメイン知識が増加する特性を持つ。この特性を持つ統合したシソーラスを明らかにし、統合方法を確立する必要がある。

1 独立コンサルタント
Independent Consultant

2 南山大学
Nanzan University

3 立命館大学
Ritsumeikan University

4 東京工業大学
Tokyo Institute of Technology

5 神奈川大学
Kanagawa University

6 名古屋国際工科専門職大学
International Professional University of Technology

本研究の目的を以下とする

- 統合したシソーラスの満たすべき要求を示す
 - 同じドメインの異なる情報源から生成されたシソーラス同士の統合手法を確立する
 - 統合したシソーラスを用いた要求獲得が、統合しないシソーラスを用いた要求獲得に比べて有用であることを確かめる
- 以降、統合したシソーラスを統合シソーラスということにする。

3. THEOREE とシソーラス

THEOREE で利用するシソーラスのドメイン知識は、機能と機能に付随する制約からなる。機能は、動作と動作対象の次いで表現される。これらの要素の間には関係があり、機能間の関係は Has-a 関係、動作対象間の関係は HasA, IsA 関係がある。

THEOREE は、顧客が初期要求リストを提示することから始まる。分析者は要求リストに出現する語句とシソーラス中の動作や動作対象を対応付け、シソーラスの要素間の関係を使って、顧客に問い合わせる質問や提案を作成し、顧客からの回答をもとに既存の要求の詳細化や新たな要求を獲得し、獲得した要求を改訂要求リストに追加する。これを質問や提案する項目がなくなるまで繰り返す。

たとえば、「資料を貸出す」機能は「貸出す」という動作と「資料」という動作対象から構成される。この機能に「貸出期間」と「貸出冊数の上限」の制約が付随する。THEOREE はシソーラスを参照し、質問、提案を作成し要求を獲得する手法である。

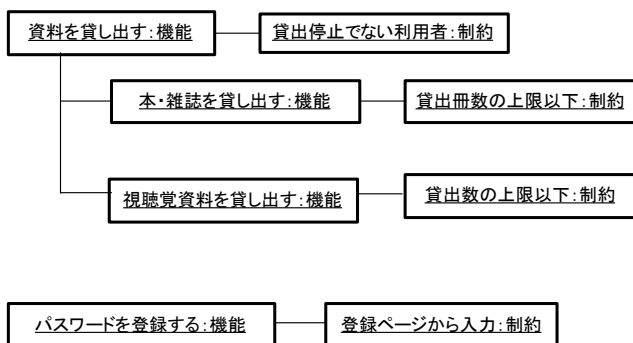


図1 B 市立図書館のシソーラス (抜粋)

構築支援(手法と支援ツール)[2]で構築する THEOREE 用シソーラスは、1つの情報源から構築されたシソーラスである。情報源は計算機で読める技術文書(テキストファイル)である。詳細は参考文献[2]を参照。情報源である技術文書のドメイン知識に、両立しない機能、同時に存在しない動作対象、HasA 関係を辿ると元の HasA 関係に戻る階層関係、HasA 関係が逆転していることがあれば、これを矛

盾ということにする以降、矛盾をこの意味で使う。情報源のドメイン知識に矛盾があれば、これを情報源として採用しない。

4. 統合シソーラス

4.1 統合シソーラスの統合の方針

同じ文字列で同じ意味を持っている動作、動作対象、制約などの構成要素は同じものとして統合する。シソーラスの構築手法[2]ではドメイン知識に矛盾がない情報源のみ扱うため、この構築手法で構築したシソーラス内では矛盾はない。一方、情報源の背景にある方針、環境などが異なるため、異なる情報源から得たシソーラス内に矛盾するドメイン知識を発見することがある。矛盾の例としては、ある情報源に含まれる機能と別の情報源に含まれる機能が両立しない場合がある。両立しない機能の例を示す。

パスワードを登録する機能が、市立図書館で異なっていた。A 市立図書館は初期パスワードを発行し、初期パスワードでログインし、初期パスワードをパスワード(自分で考えた本パスワード)に変更することを「パスワードを登録する」とし、B 市立図書館はパスワードを登録するページに利用者がパスワード入力することを「パスワードを登録する」としている。これは、利用者カードの発行時にパスワードを図書館がログインするためのパスワードを初期パスワードとして発行する方式とログインするためにはパスワードを登録しなければならない方式があり、これらは両立しない。この「パスワードを登録する」をA 市立図書館の「パスワードを登録」とB 市立図書館の「パスワード登録」として統合する。

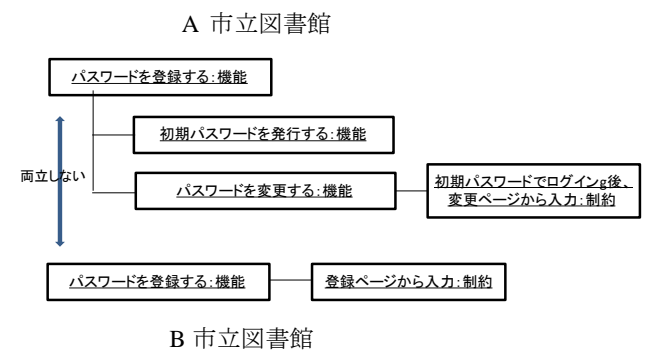


図2 シソーラス間で発生する矛盾の例

このシソーラス間に矛盾がある場合、矛盾を解決しないで統合する。矛盾を含めて統合するため、元のシソーラスのドメイン知識はそのまま統合される。

複数の情報源から構築したシソーラスを統合シソーラス内に矛盾があっても、1つの情報源から構築したシソーラスを復元できれば、復元したシソーラスには矛盾はない

ので、複数の情報源の間で矛盾があっても、この復元したシソーラスで THEOREE を使えば問題は生じないと考える。したがって、統合シソーラスには、機能、動作、動作対象などに加えて、どの情報源から得たかの情報を付加する。この情報をシソーラスの識別子(TID)と呼ぶ。同じ識別子を持つシソーラス要素をたどることにより、1つの情報源から得たシソーラスを統合シソーラスから復元することができる。

統合するシソーラスの前提は以下である。

- 1つの情報源のドメイン知識の構成要素と構成要素間の関係には矛盾がないこと。
- これを満たす情報源から構築したシソーラスは無矛盾である。

1つの情報源のドメイン知識から構築したシソーラスの構成要素には属性として同じシソーラスの識別子(TID)が振られ、シソーラスのすべての構成要素はこの属性 TID を持っている。以降、属性 TID を単に TID という。TID は情報源の識別子でもある。統合シソーラスの構成要素は TID の集合を持つ。

4.2 統合シソーラスの使い方

シソーラスを使った要求獲得手法である THEOREE を適用する。統合前の個別のシソーラスの利用であり、1つのシソーラスの識別子を決め、決めたシソーラスの識別子を持つ統合シソーラスの要素を優先して検索し、検索結果がないとき、別のシソーラスの識別子を持つ統合シソーラスの要素を検索する。1つの情報源では初期要求と顧客の回答に対応ができないうとき、提案可能な話題に移るときに他の情報源のドメイン知識を参考にする。

統合シソーラスを THEOREE に利用するとき、ドメイン知識を探す優先順が決まっている。分析者は要求獲得を始める前に、TID を一つ定め、この TID のシソーラスを復元し、ドメイン知識を探すとき復元したシソーラスを優先し、復元したシソーラスにないドメイン知識が初期要求または顧客の回答にあれば統合シソーラスを探す。

同じ TID を持つ統合シソーラスの構成要素、要素間の関係を使い統合シソーラスを検索し、検索結果があれば検索結果から質問、提案の作成し、見つからないときは統合シソーラスのすべての構成要素、構成要素間の関係を使い検索し見つければ、矛盾しないことを確認して質問、提案の作成で使う。

例えば、「予約された資料」を取り置いたら取り置いていることを利用者に通知したいという回答が顧客からあったとする。このとき分析者は A 市立図書館から得たシソーラスを使って要求獲得を進めていたとする。A 市立図書館から得たシソーラスには「通知手段」はあるが、「何時通知するか」がなかったが、B 市立図書館には、「予約された資料を取り置いたら」(制約)、「予約された資料を取り置いていることを通知する」(機能)があった。-分析者は、識別

子に A 市立図書館を持つ構成要素ではなく、B 市立図書館を持つ要素へと移り、以降は B 市立図書館の要素をたどることにより要求獲得を進めていく。

統合シソーラスの定義を以下に示す。

統合シソーラスの定義

シソーラス: $TH = (N, E, Atr)$

$N = F \cup C$

A: 動作の集合 (表現は文字列)

O: 動作対象の集合(表現は文字列)

F: 機能: 機能, $F \subset A \times O$

C: 制約: 制約の集合

E: N 間の関係 $E \subset N \times N$

Label: シソーラスの識別子の集合:

$atr = N \cup A \cup O \cup E \rightarrow 2^{Label}$

$f = \langle a, o \rangle \in F$ のとき, $atr(f) = atr(a) = atr(o)$ である

1つの情報から論文[2]の手法で構築したシソーラスを個別シソーラスと呼び、1つの情報源 X から構築された個別シソーラスに対し、X 情報を付加したシソーラスを TH_x とすると、

$TH_x = (N_x, E_x, atr_x)$ では

$atr(f_x) = atr(r_x) = \{X\}$

ここで、 $f_x \in F_x, r_x \in E_x$

5. シソーラスの統合方法

5.1 統合の方針

特定のドメインの情報源から構築した個別シソーラス群を解析して、統合シソーラスを作成するアプローチを採用した。

この方法の特徴は、以下である。個別シソーラス間では、同じドメインの入力情報源から構成されるため、多くの構成要素と構造は同じものとなるが、入力情報源が完全に同じではなく、幾分の違いから個別シソーラス間で差異や矛盾が生じる。この差異に対応する個別シソーラスの要素に識別子 TID を付して、統合シソーラスを作成する。統合シソーラスを正確かつ効率良く更新する必要が生じた場合も、TID を活用できる。

統合シソーラスでは、個別シソーラス群の差異部分にあたる構成要素と構成要素間の構造を元のまま参照できるようにしている。換言すれば、個別シソーラス間の構成要素間の構造の不一致は解決(構造を崩す)しないで、統合シソーラスを構築している。

5.2 統合手順

統合シソーラスでは、個別シソーラス群の差異部分にあたる構成要素と構成要素間の構造を元のまま参照できるようにしている。換言すれば、個別シソーラス間の構成要素間の構造の不一致は解決(構造を崩す)しないで、統合シソーラスを構築している。n(n>1)個の情報源から構築した n

個の個別シソーラスがあるとする。統合は、 n 個の個別シソーラスから、最初に2つを選択して、統合シソーラスを構築する。次に未選択の個別シソーラスを1つ選択して、構築済の統合シソーラスと合わせて、新たな統合シソーラスを構築する。この手順を繰り返して n 個の個別シソーラスから最終的に1つの統合シソーラスを構築する。

手順1 n 個の個別シソーラスから2個を選択し、統合シソーラスを構築する。

手順2 未選択の個別シソーラス群から1個を選択し、統合シソーラスと合わせて新たな統合シソーラスを構築する。

手順3 以降、未選択の個別シソーラスがなくなるまで、手順2を繰り返す。

5.3 統合する規則

統合シソーラスと個別シソーラスから新たな統合シソーラスを構築する規則は以下とおりである。

2つの統合シソーラス TH_1 と TH_2 を統合して得られたシソーラス $TH'=(N',E',atr')$ とする。

5.3.1 規則1:

機能(動作と動作対象の対) f (統合シソーラスの機能を f_1 , 選択したシソーラスの機能を f_2 とする)については、動作と動作対象の文字列が一致するペアが存在するとき、

$$atr(f_1)' = atr(f_1) \cup atr(f_2)$$

このとき、下位の機能について矛盾がないことを確認し、矛盾があれば、一致するペアが存在しないとする。一致するペアが存在しないとき

統合する機能 f_2 を統合シソーラスに追加する

動作対象 o (統合シソーラスの動作対象を o_1 , 選択したシソーラスの動作対象を o_2 とする)については、動作対象の文字列が一致するペアが存在するとき、

$$atr(o_1)' = atr(o_1) \cup atr(o_2)$$

このとき、下位の機能について矛盾がないことを確認し、矛盾があれば、一致するペアが存在しないとする。

統合する動作対象 o_2 を統合シソーラスに追加する

一致するペアが存在しないとき

統合する動作対象 o_2 を統合シソーラスに追加する

5.3.2 規則2:

機能間の関係については

関係 $RF_i \subset F_i F$ 例;

$rf_1 \in RF_1$ と $rf_2 \in RF_2$ で

一致する ($rf_1=rf_2$) が存在するとき

$$atr(rf_1)' = atr(rf_1) \cup atr(rf_2)$$

一致するペアが存在しないとき

統合するシソーラスに rf_2 を追加する

動作対象間の関係については

関係 $RO_i \subset O_i \times O_i$ において

$ro_1 \in RO_1$ と $ro_2 \in RO_2$ で

一致する ($ro_1=ro_2$) が存在するとき

$$atr(ro_1)' = atr(ro_1) \cup atr(ro_2)$$

一致するペアが存在しないとき

統合するシソーラスに rh_2 を追加する

5.3.3 統合例

動作対象をとりあげ、統合規則を例示する。統合前の統合シソーラスを s_1 , 統合させるシソーラスを s_2 , とし統合後の統合シソーラスを s_1' とする, 図3に例示の凡例を示し, 一致するペアが存在するときの統合結果を図4で, 存在しないときの統合結果を図5で, 矛盾が生じる場合の統合結果を図6で示す。

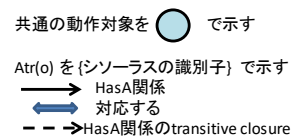


図3 例示の凡例

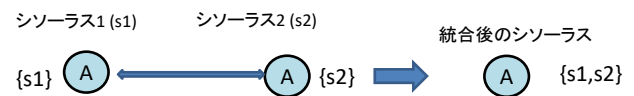
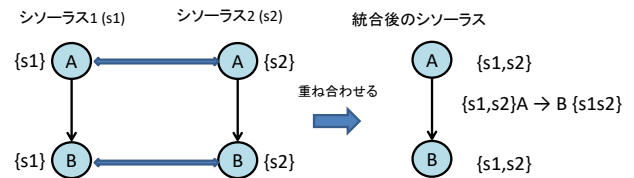


図4 一致するペアが存在するときの統合

不一致のとき

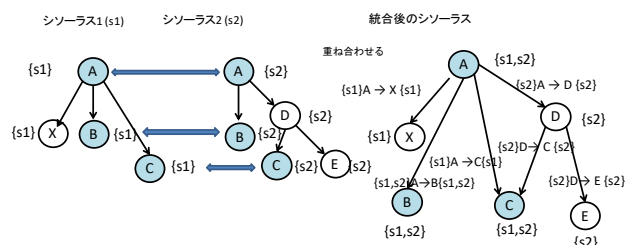
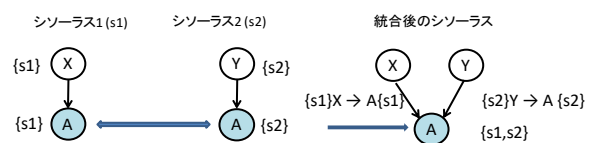


図5 一致するペアが存在しないときの統合

矛盾 (関係がループ) の例

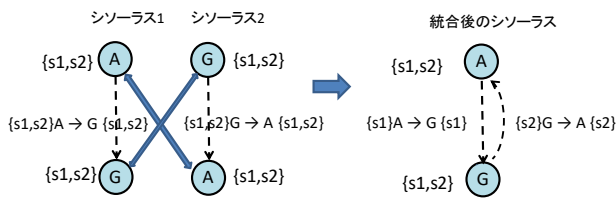


図 6 矛盾（関係がループ）したときの統合

6. シソーラスの統合方法の事例研究

6.1 事例研究

ドメインを図書館のカウンタ業務とする。市立図書館のホームページから収集した技術文書（テキスト部分）を情報源として、個別シソーラスを構築しており、2つの個別シソーラスから統合シソーラスを構築する。

6.2 統合前のシソーラス

個別シソーラスの細かい内容は紙面の都合上省略する。情報源である技術文書の大きさをテキストの行数で示し、シソーラスのノードとなる、機能数、動作対象数とエッジとなる関係を表1と2に示す。表1は統合する前のA市立図書館、表2はB市立図書館である。

表 1 統合前 A 市立図書館のシソーラスの情報

A 市立図書館

入力行数	645
機能数	71
機能HasA関係数	38
動作対象数	44
動作対象HasA関係数	24

表 2 統合前 B 市立図書館のシソーラスの情報

B 市立図書館

入力行数	499
機能数	44
機能HasA関係数	19
動作対象数	23
動作対象HasA関係数	9

6.3 統合シソーラス

統合規則に従って統合した結果を機能については、表3で動作対象については表4で示す。

表3は機能数と機能間のHasA関係数を示している。

表 3 統合シソーラスの機能情報

機能

機能間の HasA 関係

	統合後の件数		統合後の件数
A,Bにある	32	A,Bにある	10
A	39	A	28
B	12	B	9
計	83	計	47

表4は動作対象数と動作対象間のHasA関係数を示している。

表 4 統合シソーラスの動作対象情報

動作対象

動作対象間の HasA 関係

	統合後の件数		統合後の件数
A,Bにある	18	A,Bにある	5
A	26	A	19
B	5	B	4
計	49	計	28

6.4 統合するときの課題と対応

6.4.1 同じものを異なる文字列で表現している

A市立図書館とB市立図書館では、同じ動作対象をそれぞれ別の名前で表現していた。これらを合わせる必要があった。

例 A市立図書館の「貸出状況一覧」とB市立図書館の「貸出状況」は同じ動作対象であったので統合時に名前を訂正して対応した。統合を担当した分析者が、違う文字列については、同じ概念かどうかを判断し、「同じ概念の場合は同じ文字列にした。

6.4.2 異なる意味のものを同文字列で表現している

「パスワードを登録する」は、A市立図書館では初期パスワードを変更するとき（本人のパスワードを登録する）の機能であり、B市立図書館では、パスワードを登録するときの機能である。両立しない機能なので、矛盾したときの統合例のように、「パスワードを登録する」はA,Bの市立図書館の間では対応関係がないとみなしA市立図書館のTIDを持つ「パスワードを登録する」とB市立図書館のTIDを持つ「パスワードを登録する」をそのまま登録した。

統合した結果、統合シソーラスは次のように元の個別シソーラスよりもドメイン知識が増加している

A市立図書館と比較しての増加数

機能：12，動作対象：5，機能間の関係：9，動作対象間の関係：4

B市立図書館と比較しての増加数

機能：39，動作対象：26，機能間の関係：28，動作対象間の関係：19

統合シソーラスは統合前の個別シソーラスよりもドメイン知識量が多い（機能、動作対象、機能間の関係、動作対象間の関係の数が多い。これは、同じドメインの複数の情報源から構築したシソーラスの持つドメイン情報に差異があっても、そのまま活用できるようにし、元のドメイン知識を残したたことによる。

7. 要求獲得の事例と分析

7.1 事例

7.1.1 目的

統合シソーラスをTHEOREEで利用するとより多くの要

求を獲得できる。特定の TID の個別シソーラスにない初期要求，利用者の回答への対応も統合シソーラスを利用することにより，多くの要求が獲得できることを確認する。

7.1.2 体制

3名で実施した。この3名は図書館を知り分析経験も豊富であり，要求工学にも見識がある社会人である。

3名の役割は，以下である。

分析者 1名

利用者(THEOREE の顧客に相当する)

一般利用者 1名，

図書館職員 1名

分析者が事例の結果を分析した

7.1.3 概要

(1) 期間

2か月

(2) 提示した初期要求

(3) 第9回 ICSE のワークショップの課題にある図書館問題[6]の要求を初期要求とした。日本語訳を図7に示す。

1. 以下のトランザクションが発生する小さな図書館データベースシステムを考えよ。

(1)本の貸し出しと返却

(2)図書館への本の登録と削除

(3)特定の著者の本，あるいは特定の分野の本に関するリストの作成

(4)特定の利用者が借りている本のリストの作成

2. システムのユーザは，図書館職員と一般利用者の2種類で，

(1)3番目のトランザクションのみ一般利用者が扱える。

(2)4番目のトランザクションで，利用者自身が借りている本のリストの作成は一般利用者でも扱える

3. データベースの満たす要件は，

(1)すべての本は貸し出し可能か貸し出し中

(2)貸し出し可能かつ貸し出し中の本はない

(3)利用者が借りられる本の冊数には上限がある

図7 初期要求リスト

(4) 事例の要求獲得手順

分析者と利用者との間のやりとりは E-Mail を使い，利用者である図書館職員と一般利用者間のやりとりを禁止した。

① 利用者から初期要求を分析者に E-Mail により配布

② 分析者が初期要求から，統合シソーラスを参照し，話題と取り上げる優先順位を決める。最初の話題について，統合シソーラスを参照し，初期要求の確認と質問文を作成して E-Mail により利用者へ配布

③ 以降，利用者が分析者へ確認・質問・提案について

の承認，質問および新しい要求を回答として E-Mail で分析者に送る。この回答をもとに分析者は承認された要求は改訂要求リストに追加し，回答にある質問への回答と新しい要求については統合シソーラスを参照して質問または確認文を作成して E-Mail により利用者へ配布する。これを話題の終了まで続け，話題が終了したら残した話題の優先順位中に取り上げ，取り上げた話題についての確認と質問を E-Mail により利用者へ配布することから始める。

④ 事例では，期間を優先したので話題を残したまま途中で終了した。

(5) 分析の観点

統合シソーラスの活用に関する分析

以下の観点で分析した。

観点1：統合シソーラスの有用性

獲得した要求において，寄与したドメイン知識の分類（初期要求，統合シソーラス，利用者）し件数を測る。

観点2：統合シソーラスのドメイン知識の利用場面

統合シソーラスのドメイン知識が利用される場面（初期要求の補足等，分析者からの質問・提案文作成，分析者と利用者との対話による要求獲得）の比率を測る。

観点3：要求獲得で利用されたドメイン知識の情報源

要求獲得で利用された統合シソーラスのドメイン知識とドメイン知識の TID の集合別の件数を測る。

7.2 事例の結果と分析

機能と動作対象を分析の対象とする。機能間の関係と動作対象間の関係は機能と動作対象を分析するときに参考にした。

統合シソーラスの有用性

獲得した要求の内，統合シソーラスのドメイン知識の寄与は表5から機能で63%，表8から動作対象で50%であった。

統合シソーラスのドメイン知識の利用場面

機能は，表6から分析者の質問・提案，利用者との対話，初期要求の補足の順であった。表9から動作対象も同様であった。

要求獲得で利用されたドメイン知識の情報源

利用されたドメイン知識の情報源の割合は，機能については表7から，双方にあるドメイン知識が81%で残りはA市立図書館，B市立図書館であった。動作対象については，表10から双方にあるドメイン知識が64%，A市立図書館，は36%，B市立図書館は0%であった。

改訂要求リストには，統合シソーラスの複数のドメイン知識を利用して獲得された機能と動作対象があり，対話により統合シソーラスのドメイン知識から複数の動作対象を獲得できた。このため，機能では，表5のシソーラス利用の機能件数と表6，表7の引用された統合シソーラスの要

素数とは一致していない。動作対象では表 8 のシソーラス利用の動作対象件数と表 9, 表 10 の引用された統合シソーラスの要素数とは一致しない

表 5 改訂要求リストの機能の出所

改訂要求リストの機能	件数	比率
シソーラス利用	19	0.63
内 単独(提案)	8	0.27
内 初期要求と併用	4	0.13
内 利用者への影響(質問、提案)	7	0.23
初期要求のみ	2	0.07
初期要求と利用者	2	0.07
利用者へのみの回答	7	0.23
計	30	1.00

表 6 機能における統合シソーラスの利用形態

引用シソーラス	件数	比率
内 単独(提案)	9	0.43
内 初期要求と併用	4	0.19
内 利用者への影響(質問、提案)	8	0.38
計	21	1.00

表 7 機能における利用した統合シソーラスの TID

使用したシソーラスのID	件数	比率
A	20	0.95
内 共通	17	0.81
内 Aのみ	3	0.14
B	1	0.05
計	21	1.00

動作対象

表 8 改訂要求リストの動作対象の出所

改訂要求リストの動作対象	件数	比率
シソーラス利用	16	0.50
内 単独(提案)	8	0.25
内 初期要求と併用	2	0.06
内 利用者への影響(質問、提案)	6	0.19
初期要求のみ	4	0.13
初期要求と利用者	1	0.03
利用者へのみの回答	11	0.34
計	32	1.00

表 9 動作対象における統合シソーラスの利用形態

引用シソーラス	件数	比率
内 単独(提案)	7	0.50
内 初期要求と併用	2	0.14
内 利用者への影響(質問、提案)	5	0.36
計	14	1.00

表 10 動作対象における利用した統合シソーラスの TID

使用したシソーラスのID	件数	比率
A	14	1.00
内 共通	9	0.64
内 Aのみ	5	0.36
B	0	0.00
計	14	1.00

表 7 は機能の要求獲得で利用した個別シソーラスの機能に関するドメイン知識の件数を示している。A 市立図書館と B 市立図書館の双方にあったドメイン知識は 17 件, A 市立図書館のみにあったドメイン知識は 3 件, B 市立図書館個のみにあったドメイン知識は 1 件であった。1 件でも多くの要求を獲得するという意味では, 統合シソーラスは有用である。

「パスワードを登録する」という矛盾した(両立しない)機能が統合シソーラスにあったが, 先に A の個別シソーラス情報を参照した質問で図書館職員が承認したため, 矛盾は生じなかった。

8. 考察

8.1 統合シソーラスの利用による効果

表 5, 表 6 からドメイン知識が増えれば, 獲得できる要求が多くなることを確認できた。統合シソーラスのドメイン知識を利用して利用者とは対話して獲得した要求が機能で 23%, 動作対象で 19% である。利用者との対話によって要求を獲得する THEOREE の特徴も確認できた。

8.2 ドメイン知識が少ないシソーラスを優先した場合

優先するシソーラス A, B を入れ替え, B (市立図書館) のシソーラスを優先させたら, B からのドメイン知識の利用は多くなるが, A からのドメイン知識の利用も多くなると予想できる。そして, 「パスワードを登録する」という両立しない(両立しない)機能については, B のシソーラスを参照した質問では, 図書館職員から別の提案を求められることが想定でき, ここで A を参照し, 代替案を提案することが可能である。

8.3 両立しない機能と動作対象の扱いについて

ドメイン知識には, 例えば A 市立図書館のパスワード登録と B 市立図書館のパスワード登録は両立しない。A 市立図書館の xxx のように A 市立図書館という修飾付きの機能と動作対象がある。TID はこの修飾語と同じ働きをする。修飾語を付けて識別する機能と動作対象があれば, 矛盾した機能, 動作対象の存在を示している。修飾付きの機能と動作対象をそのまま統合することにより, 機能同士は同じ文字列で異なる意味があっても, 修飾部分が異なっていれば統合できる。ドメイン知識には必須の機能にいくつかの種類があるし, 動作対象でも同じである。この問題を解決できる。この種類毎にシソーラスをグループ化することも可能である。

9. おわりに

統合シソーラスとは何かを示し, 異なるいくつかの情報源からの個別シソーラスを統合し, 統合シソーラスを作成する方法を開発した。2 つの市立図書館の個別シソーラス

に本論文で述べた統合方法を利用して、図書館の統合シソーラスを作成し、作成した統合シソーラスのドメイン知識の量が増えていることを確認した。

この統合シソーラスを THEOREE に利用し、単独の個別シソーラスを利用した場合よりも獲得した要求が増えていることも示した。本論文で述べた統合シソーラスは、要求獲得に有効であり、効果があった。

今後の課題として、①取り残した話題がある場合に、次の話題から再開する仕組みを提供することによる、統合シソーラスの効果の確認、②既存の公開されている仕様、例えば OPAC の仕様、と特定のサービスを提供する製品の仕様から構築したシソーラスを統合した統合シソーラスの作成と、この統合シソーラスを利用した要求獲得実験の実施、③システム改修後のシソーラスを統合することによる変更容易性の確認、例えば B 市立図書館のホームページ改訂後のホームページから構築したシソーラスの適用による確認、④システムの利用者向けのドメイン知識はあるが、システム提供者のドメイン知識がない分析者がシソーラスを利用する場合の問題点の洗い出し、⑤個別シソーラス間で矛盾する箇所が多い場合の対応と矛盾の識別規則の補強、がある。

参考文献

- [1] 加藤潤三, 佐伯元司, 大西淳, 海谷治彦, 山本修一郎: シソーラスを利用した要求獲得方法 (THEOREE), 情報処理学会論文誌 Vol. 50, No. 12 pp. 3001-3017, 2009...
- [2] 加藤潤三 佐伯元司 大西淳 海谷治彦 林晋平 山本修一郎: 要求獲得のためのシソーラス構築支援, 情報処理学会論文誌 Vol.57, No.7 1576-1589, 2016.
- [3] 大村, 情報検索支援のための語彙拡張: シソーラス統合の事例, http://www.mslls.jp/am2014yoko/01_omura.pdf
- [4] Arbib et al A Bases for theoretical computer science 1981 Springer Verlag
- [5] Sommerville, I. and Sawyer, P. (著), 富野壽 (監訳): 要求定義工学プラクティスガイド, p.58, 共立出版(2000)
- [6] Proc. the 9th International Workshop on Software Specification and Design, 1987