

業務制約を利用した 共有データオブジェクト方式のWeb サービス連携

山口大貴[†] 坂下善彦[†]

重要な Web 技術の一つであるサービス連携だが、AP/Service の間でやり取りするデータ項目を完全に特定することは、個々の AP/Service の仕様に依存していることが多く、一般には困難である。本研究では連携を行う AP/service の間に業務制約を設け、予めその業務に仕様されるデータ項目を予想し連携情報として管理しておき、共有データオブジェクトにより連携情報から各データ項目のマッピングを行う方式を提案する。

Web service cooperation system through the common data object restricted within specified domain

DAIKI YAMAGUCHI[†] YOSHIHIKO SAKASHITA[†]

It's dependent on the specification of each application/service to specify completely the data item exchanged between application/service, although it is the service cooperation which is one of the important web technology in many cases, and generally it is difficult.

In this research, operating restrictions are prepared between the application/service which cooperates, and the data item beforehand used for the business is expected, it manages as cooperation information, and the system which maps each data item from cooperation information by a share data object is proposed.

[†] 湘南工科大学 工学研究科 電気情報工学専攻
Shonan Institute Technology

1. はじめに

現在、Web サービスは日々進歩しており、サービス間の連携は必要不可欠な技術である。主流とされているサービス連携方式には、W3C/WDDL などの枠組みが存在するが、サービス連携のために、各サービスによって異なる項目を繋ぎ合わせるため、設計者がシステムの詳細を把握し個々の項目の連携情報を取らなければならない。そのため、アプリケーションサービス間でやり取りするデータ項目を完全に特定することは、個々のアプリケーションの仕様や利用環境に依存していることが多く、一般には困難である[1]。

本研究では、対象業務内のアプリケーションで用いられるデータ項目名は、ある程度限定されると予測し、その範囲の中で可能なサービス連携に必要な連携情報を予め分析し[4]、そのデータを利用し、情報を変換させることによりサービス連携を行う。

AP 間で直にデータの変換を行う方法では、個々のアプリケーション間で実施するため、互いのデータを対応させる仕組みが必要となる。対象アプリケーションが現れる都度システム変更が必要となり、一貫性が保てなくなる。システムで共通的に用いる共有型の共通データ集合を備え、この集合と各個別アプリケーションの間でデータを対応させ、アプリケーション間でデータを変換する方式を本研究では採用する。共有データ集合を、汎用的に定義することは困難であるため、利用アプリケーションの範囲を限定する。方針として、業務依存性が高いことを予想し、業務又はビジネスパターンを基にした制約とし限定する方法を用いる。

2. 目的

共有データ集合と個別アプリケーションとの間でのデータ項目のマッピングを行うアルゴリズムを提案する。方式は既に共有データ方式を提案した[2][3]。

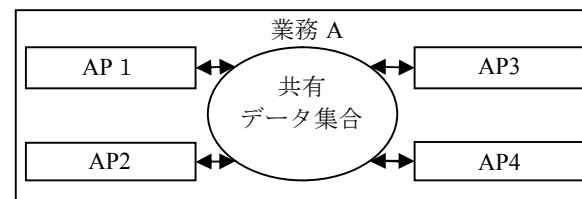


図1. 共有データ方式による連携図

図1に一般の共有データ方式の図を示す。共有データ集合には「業務 A」で利用が予想される項目を予め管理しておく。個々のアプリケーション間ではなく、「業務 A」に依存しているため、どのアプリケーション間でも精度良く変換することができる。しかし、一般的に項目名の対応付マッピングは精度が低いため、適用対象分野を特定して制約すれば、マッピングの精度は高まると期待できる。この為に、特定の業務を対象とすることで範囲を限定する。本研究は以下の方針で進める。

(1) 具体的なアプローチ

制約として範囲を限定しても、項目名は異なるが同じ用途、項目名は同じだが異なる用途の項目が存在する。これらの項目は、共有データ集合からマッピングの正確性を出し、確率の高い方で連携を行うものとする。

(2) 実施内容

対象となる範囲を限定したシステムを構築し、公開されているデータ及び予想できる項目から共有データ集合を構築する。共有データ集合は、用途別にグループ管理しておく。

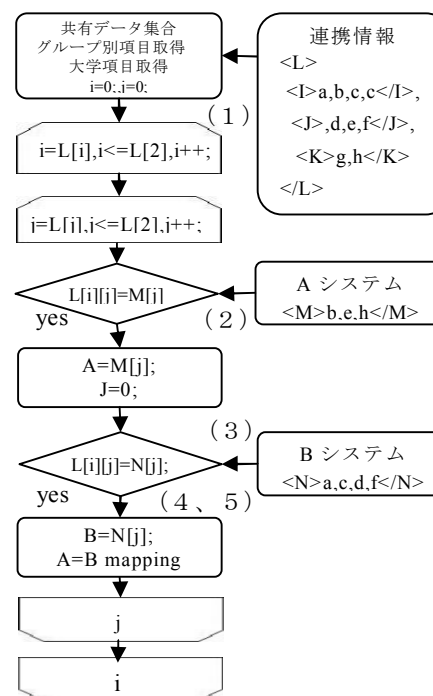
(3) 評価方法

各項目のマッピングをプログラムでテストし、どの程度の一致率なのかを確認する。このアルゴリズムを検証するために、大学の「学務」の業務を想定対象として用いる。

3. 項目間マッピングアルゴリズム

Webサービスの連携は範囲を限定せずに行えることが理想であるが、関係情報が管理するデータ量が増えると、マッピング処理回数増加すると予想される。本研究では「学務」として制約を加え、サービス同士を連携させる方式について図2に沿って説明する。連携情報は制約から、予想できる項目を、用途別(<I>,<J>,<K>)にまとめて管理しておく[5]。

- (1) 共有データ集合に連携情報及び対象システムの項目を取得する。
用途別グループ<i>,<j>,<k>を取得する。
- (2) 連携情報の項目とAシステムの項目が一致するか探索する。
図2では、項目L[i][j]と項目M[j]が一致する場合(L[i][j]=M[j])、項目を置き換えると、連携情報<I>のaとAシステムの<M>のbが対応する。
- (3) 一致した場合、項目を保管し、同グループからBシステムも探索する。
図2において、最初の一致は、iループ2回目のbが一致する。bを保管し、jループを初期化し、同グループでBシステムの項目も探索する。
- (4) 両方のシステムで一致した場合マッピングは完了する。
AシステムのbとBシステムのaが同じ<I>グループで存在したため、マッピングは完了する。
- (5) 項目が両方のシステムで一致し、その項目が複数存在する場合、連携情報全



体から見た一致率の高い組み合わせで対応付けを行う。

<I>からAシステムではb、Bシステムではaとcが存在する。図2の連携情報からAシステムのbと、Bシステムのa,cどちらかが対応付けを行うことになる。この場合、連携情報からより多く使用されている項目を選ぶ。aは一つに対し、cは二つ使用されているため一致率が高いAシステムのbとBシステムのcで対応付けとなる。詳細は5章で述べる。

図2. マッピングアルゴリズムの流れ図

4. 履修申告サービス

4.1 システムの概要

Webでは、様々な方法でデータ参照が行われ、参照結果をユーザに提供している。ユーザは視覚的にデータ参照が可能かどうかを判断し(カーソルが乗ると単語の色が変化など)その単語を選択することで、リンクによるWebデータ間の参照が行われたり、データベースに接続しデータを取得して参照するなどの方法がある。

大学の「学務」における履修申告方法は、Webシステム上で申告することが多い。しかし、履修申告システムにシラバスデータや自身の成績データを、うまく組み合わせていない場合、安易に科目を決定し、申告してしまうことが起こりえる。前提条件科目などを取得していないにも関わらず、履修してしまうと履修計画に支障が出てしま

う。

また最近では、興味のある科目が所属大学で開講されていない場合、他大学で受講することができ、単位を取得できる。その場合、前提条件科目の確認などは、シラバスは他大学、成績は所属大学のようにシステムが異なるため、項目を対応付けることができない。そのため、所属大学と他大学とで Web サービスの連携を行う必要がある。科目名や講義名などのように、大学によって項目名が異なる場合があるため、自律的に項目を繋ぎ合わせ連携できることが望まれる。

対象業務として履修申告 Web システム（履修システム）にシラバスデータや成績データを細かく組み合わせ、学生の履修申告を、円滑に正確に行えるようにすることができれば、科目担当者も受講者である学生も科目計画に集中することができる。そのためには、履修計画に支障を来す可能性のある科目取得の際、学生に警告として促し、サポートする必要がある。

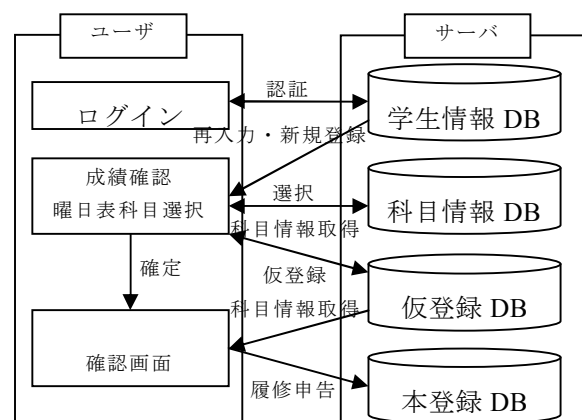


図3. 履修申告支援システム流れ図

4.2 シラバス内データ参照機構

科目選択の際、重要な項目は警告し、その他の項目は学生が確認しやすく提供できるようシラバスデータを履修システムに組み込む。

学生が履修したい科目を選択すると、シラバスデータを取得する。シラバスデータの学生に提供すべき項目である履修条件・前提科目などを取得しているかを成績 DB と照合する。履修条件・前提科目を取得済みであれば選択された科目のシラバスを提供し、未取得ならば履修条件・前提科目を取得していないことを学生に警告し、シラバスデータを提供する流れとなる。データ参照機構は、履修システム内に組み込まれる。

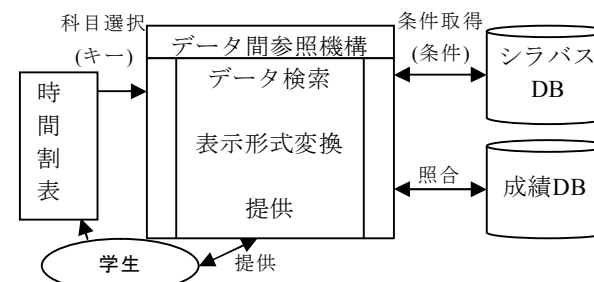


図4. 履修システムでのデータ参照機構の役割

図4に履修システムのデータ参照機構の役割を示す。例として、電子回路2を学生が選択した場合の流れを示す。

1. 「電子回路2」を選択（キー）
2. 「電子回路2」のシラバスデータをシラバス DB から取得
3. 履修条件科目項目から「電子回路1」を取得
4. 「電子回路1」を取得しているかログイン時の ID である学籍番号から成績 DB と照合
5. 「電子回路1」を取得済みであれば警告する必要があるため学生に「電子回路2」のシラバスデータを提供する
6. 「電子回路1」を未取得ならば学生に未取得であることを警告し、「電子回路1」のシラバスデータを学生に提供する

履修前提科目であれば、あくまで目安なので問題ないが、条件科目の場合でシラバスデータ、成績データが組み込まれていない場合重大なミスとなり、履修計画、科目計画に支障をきたすため、履修システムにはこのような役割を持ったデータ参照機構は必須といえる。また、例では条件・前提科目の項目は重要なため自動参照している。他の重要度の低い項目は重要度の高いデータ提供時に視覚的に解るように参照先として提供しておく。

4.3 データ参照機構による分野評価

科目を選択する際、履修計画に則り科目を選択する、または今後履修計画を立てるため、自身の成績を分野別評価した表や図を見ながら行うことが望ましい。本節では、学生に得意分野、不得意分野を複数の手法で提供できる様、履修申告サービスを構築する。複数の方法で提供する理由として、得意・不得意科目を決める観点は学生それぞれであるためである。A 評価が多い分野が得意、取得科目数が多い分野が得意と様々である。いくつかの分野評価方法を以下に記す。

- (1) すべての取得科目情報を分野別に一覧表示

学生はログイン後、学籍番号をキーに成績 DB から取得した取得科目一覧表を確

認することで、各科目の細かい情報を見ることができる。

(2) 分野別に棒グラフで表示

得意・不得意分野を確認する際、評価数で判断できる。

(3) レーダーチャートによる評価

評価別にポイント換算し確認できる。棒グラフとの違いは、各分野の評価が1つの図にまとまる。分野によって取得した科目数に差がある場合、判断しにくいいため5段階評価にしてある。他分野に比べある分野はどれだけポイントを稼いでいるかという視点で判断できる。

(4) 円グラフによる比率表示

合計取得科目数から各分野の比率を出し提供する。比率の高い分野を得意分野と判断することもできる他、各分野をバランス良く取得したい場合、偏りを防ぐことができる。

学生によって判断基準が異なるためこれ以上の深入りはしない。あくまで解りやすく提供することを目的としているため図表以外での方法は対象外とした。データ間参照機構は、特定の大学のみを対象としており、他大学では項目名の違いなどから運用することができない。そのため制約により限定した範囲の中であれば自由に連携が行える方式を次章で述べる。

5. 制約を利用したサービス連携

初めに現在Webに公開されている各大学のシラバスデータを元に、同じ利用用途である項目名を連携情報としてまとめる。例えば図5に示すように、学籍番号、在学番号のように項目名は異なるが、同じ利用用途であるため管理者が判断し共有データ集合として管理する。これらの連携情報で、A大学、B大学の項目名を繋ぎ合わせる。

図6に連携時のデータ移動の流れ図を示す。B大学の学生がA大学の講義を受講したい場合、前提条件科目などの確認を行うため大学間で連携を行うとする。連携情報からA大学の「学籍番号」、「在学番号」などの項目名を繋ぎ合わせ、A大学のDBへの登録を許可する。これによりA大学の学生と同様にシラバスデータと成績データなどを組み合わせ、前提条件科目などの確認を行うことが可能となる。

5.1 連携情報の取得

Web上に一般公開されているシラバスデータから、同じ利用用途である項目をグループ分けする。項目をまとめる際、連携の際に生じる問題について考慮しながら行う。

1. 項目名が同じだが、利用用途が異なる項目 項目名=項目名 用途 同
2. 項目名が異なるが、利用用途が同じ項目 項目名≠項目名 用途 異
3. いくつかの項目をまとめている項目 例 「曜日・時限・単位」

4. 複雑な項目名

以上が考えられる問題である。

例 「授業科目名[英文授業科目名]」

5.2 項目のマッピング

項目名が同じだが、用途が異なる例として、項目名「科目区分」などが考えられる。この場合、連携情報からデータ取得、各大学での利用項目の取得から再び連携情報と照し合せ、どの項目同士が相応しいかのチェックを行う。項目名「科目区分」は、シラバスデータから様々な大学が使用している項目名であるが、中のデータの用途は異なる場合がある。必修選択の区分、ネットワーク関連分野の区分などのようにである。この場合の解決方法として、シラバスデータから項目をまとめる際、データの用途を照らし合わせて、必修の区分、分野の区分、それぞれの確立を出し、連携の際何%の確立でこの連携は正確性があるという方法で行う。

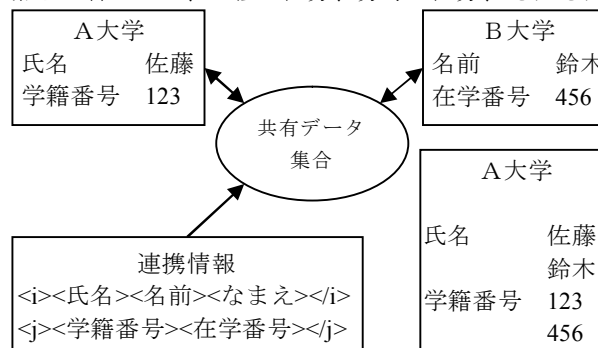


図5. 学務でのマッピング流れ図

	A大学	B大学	C大学	D大学
I 項目名	科目区分	科目区分	科目区分	科目区分
I 用途	必修選択	必修選択	必修選択	分野区分

上記のようなデータが取れた場合、「科目の区分」が項目名として出てきたら、75%の確率で用途必修区分のグループとなる。すべての連携情報を繋ぎ合わせた後、全体の正確性を出すことでA大学とB大学で行ったサービス連携は何%の正確性を持っているとすることができる。項目名が異なるが、利用用途が同じ項目についても同様の方式で対処する。

5.3 項目の分割

次に、いくつかの項目をまとめている場合の対処方法を記述する。

科目名/英文科目名	単位と開講時間
-----------	---------

例として上記の場合、一つの項目に用途が複数存在するため、データごとにグループ分けすることは、様々なパターンが存在するため困難である。そのため本研究では、

項目名の分割を行い各項目ごとにデータをまとめるものとする。

	A大学		B大学	C大学
項目名	科目名/英文科目名		授業名	講義名



	A大学		B大学	C大学
項目名	科目名	英文科目名	授業名	講義名



	D大学	E大学	F大学	
項目名	科目名	授業科目名	講義名・講義名(英文)	



	D大学	E大学	F大学	
項目名	科目名	授業科目名	講義名	講義名(英文)



<I><科目名><授業名><講義名><授業科目名></I>

<J><英文科目名><講義名(英文)></J>

項目を分割することで単体のデータとして連携情報を取得する。項目を分割することにより「科目名」、「講義名」などは用途が同じグループとして扱う。公開されているシラバスデータなどから他の用途は考えられないため一致率100%となる。基本となる項目は人間が判断し分割するが、テストの際は、Web上のため人間の目ではなく共有データ集合が判断して項目を分割する必要がある。

共有データ方式ではSPLIT関数を使用し、文字列を分割する。上記のプログラムは項目の分割に使用されるデリミタなどを予想して構築してある。\$strに格納されている「科目名/英文科目名」の「/」に着目し\$maeには科目名、\$ushiroには英文科目名のように項目を分割することができる。分割した各項目ごとに、連携情報から同じ用途の項目を繋ぎ合わせ、連携の正確性を出すものとする。

5.4 複雑な項目

業務制約「学務」のシラバスデータでは、どの大学においても「科目名」、「担当者」、「授業の目標」など基本的な項目は共通して使用されていることが、公開されているシラバスデータから明らかになった。このような項目は、大学の数の分だけデータがあるため、連携情報も多く正確性を出しやすい。しかし大学によってカリキュラムが異なるため、ある特定の大学にしか存在しないような項目がある場合がある。それらの項目は他の大学の項目と接点が無いことが多く、連携情報を元に項目を繋ぎ合わせることができる可能性が低い。

6. 結果

5.2節で取得した連携情報の結果、「学務」の範囲に限定した場合、データの取得のため参考にした約50大学ほぼすべてで使用されている「科目名」、「担当教員」などは項目名こそ多少の違いはあるものの、他の用途は存在しなかったため一致率は100%となる。科目区分の場合、必選区分、分野区分という用途で使用されていることが解った。50大学のうち「科目区分」を使用している9大学から、どちらの用途かの確立を出し、確立の高いグループを優先してマッピングを行った。

用途 科目名グループ

<I>科目名、講義名、授業科目名</I>

用途 必選区分グループ

<J>科目区分(6)、単位区分(3)、区分(2)、授業区分、履修分類、必選の別</J>

用途 分野区分グループ

<K>科目区分(3)、単位区分(1)、区分(1)、分野、エリア</K>

上記グループを元に連携を行うテスト項目

A大学

講義名

科目区分

単位区分

B大学

エリア

授業科目名

必選の別

上記の共有データ集合グループが取得できたため、以下の手順でA大学からB大学へのデータを渡すサービス連携のテストを行った。A・B大学はデータ取得に使用した50大学以外の大学であり、テスト項目として3項目を選んだ。

(1) A大学の項目名「講義名」が管理されている<I>グループからB大学で使われている項目があるか探索する

まずA大学の項目名「講義名」に注目する。「講義名」は<I>グループ以外では使用されていない項目名であるため、<I>グループからマッピング項目を探索する。一致項目は「授業科目名」のみのため「講義名」、「授業科目名」で連携1の正確性100%で本決定とする。両方の項目が一つのグループにしか存在しない場合のみ、他に連携の可能性が無いため、現在の連携情報が本決定と言える。

(2) いくつかのグループで管理されている項目は、利用大学が多いグループから探索する

次に A 大学の「科目区分」に注目し、連携情報を見ると「科目区分」は用途必選区分、分野区分両方のグループで使用されていることが解る。<J>グループの用途の「科目区分」は 6 大学が使用している。<K>グループの用途の「科目区分」は 3 大学が使用していることから、<J>グループから探索する。一致項目は、「必選の別」であることから、「科目区分」、「必選の別」の連携 2 を仮決定する。

(3) 一致項目が多数ある場合、全体による使用数から高い確率の項目を選ぶ。

最後の A 大学で使用されている「単位区分」も<J>、<K>グループ両方に存在するため、より多く使用されている<J>グループから B 大学の項目を探索する。一致項目は「必選の別」で、「単位区分」、「必選の別」の連携 3 を仮決定する。すべての項目が仮決定されたが「必選の別」が連携 2、3 両方で使用されている。この場合「必選の別」は他グループでは使用されていないため、「必選の別」に注目して探索する。一致項目として「科目区分」と「単位区分」が候補として見つかる。候補が複数ある場合全体の項目数から使用確率の高い方を優先する。<J>グループは全体で 14 項目のデータが存在する。「科目区分」は 6/14(43%)、「単位区分」は 3/14(21%)であることから確率の高い「科目区分」と「必選の別」の連携 2 を優先し、連携 3 の仮決定は破棄する。

再び「単位区分」に注目して<J>グループから探索するが「必選の別」は先ほど優先度が低かったため、無視する。同じ項目が来た場合に備えループ分の途中で break することも考慮して構築する。<J>グループでは一致しなかったため、<K>グループを探索する。新たに「エリア」が一致したため「単位区分」、「エリア」の連携 3 を仮決定する。

(4) 残されている組み合わせはすべて探索し、確率の高い項目同士でマッピングとする。

最後の分岐である「エリア」に注目した場合の、「科目区分」、「単位区分」は全体の項目数から、それぞれ 3/7(42%)、1/7(14%)となる。「科目区分」の 42%を優先したいが、連携 2 で「科目区分」は 43%なため連携 2 を優先し、14%である「単位区分」と「エリア」で仮決定する。すべての分岐が終わり、正確性が出たため、仮決定から本決定となる。

以上のようなテストを行った場合、「講義名」、「担当教員名」、「単位数」などの他に用途が見られなかった項目に関しては、データの一致率は 100%であり、問題は無い。しかし、「科目区分」などの他に用途がある項目に関しては、データの対応付けの一致

率は 40%ほどで課題が残った。

7. おわりに

本研究の対象である「学務」では公開されているシラバスから項目をグループ分けして管理することができた。このデータを元にサービス連携のテストを行うことができたが、対象が違う場合、独自の強みを持った企業がシラバスデータのように、各データを公開するとは考えにくく、連携情報の取得が困難なことが予想される。対象ごとに連携情報を取得する手法の検討も必要である。

「学務」の場合、公開されているシラバスデータから連携情報を取得できたものの、大学によって項目数が異なり、ある項目によってはデータ数に偏りが見られた。そのような項目に関しては、シラバスデータに加え、人間による予想データも必要であると考え。例えば「科目名」、「講義名」、「授業科目名」とデータが取れたとき、「授業名」という項目が予想できる。これによりデータの一致率が上がるかどうかテスト回数を増やし、検討する必要がある。

また「授業の目的」、「授業の概要」、「授業の計画」などは、項目名も複雑であるため、連携情報にヒットしないことが多い。また大学によって概要の中に目的を加えて記述してある場合などそれぞれである。しかし、これらの項目については、多少データにずれがあったとしても人間が判断できる範囲であるためあまり考慮しない。必要である場合は 5.22 項の項目の分割のように「の」などに注目してその後の文章をワイルドカードで照し合せる方法などで行う。本研究の方式である程度の一致率が出たため、今後、必選や履修条件科目などの重要な項目の一致率を高められるよう改良していく。

参考文献

- [1] 大谷真, 木下正博, 嘉数侑昇: 自律的 WEB サービスにおけるビジネスプロトコルの動的生成について, 気情報通信学会誌, D-I, 情報・システム, I-情報処理 J87-D-I(8), 824-832, (2004)
- [2] 山口大貴, 坂下善彦: シラバス内データ間参照機構を備えた履修計画支援システム, 第 8 回情報科学技術フォーラム, M-035(2009)
- [3] 山口大貴, 坂下善彦: ビジネスパターンに基づく共有データオブジェクト方式の WEB サービス連携, 情報処理学会第 72 回全国大会, 4ZC-7,
- [4] 日本の大学講義要綱検索システム: [HTTP://WWW.ASAHI-NET.OR.JP/~GB4K-KTR/SYLLABI.HTML](http://www.asahi-net.or.jp/~GB4K-KTR/SYLLABI.HTML)
- [5] PAVEL HRUBY, JESPER KIEHN, CHRISTIAN VIBE SCHELLER: ビジネスパターンによるモデル駆動設計, 日経 BP ソフトプレス(2007)