

制約自然言語の提案

The Proposal of Formal Natural Language

太田 吉美
Yoshimi Ota

大庭 英雄
Hideo Ohba

宇宙航空研究開発機構
Japan Aerospace Exploration Agency

The Japan Aerospace Exploration Agency (JAXA) is developing the Emergent Synthesis Environment (ESE) for next-generation space systems. ESE aims at innovation of the conventional operating process, the increase in efficiency of the R&D, and the improvement in reliability of space systems. We are developing the Generic Information and Knowledge Model (GIM) and the Formal Natural Language (FNL) used as the base of knowledge share environment of ESE. GIM is the general-purpose semantic network model which used the predicate as the core, and has the important feature which can also describe the syntax of natural language. The Formal Natural Language (FNL) describes GIM by the syntax of natural language. GIM/FNL enables it to aim at accumulation, share, and practical use of information or knowledge covering the life cycle of space systems. Thereby, the sharp improvement in efficiency of business and the reliability of space systems are expectable.

1. はじめに

21世紀の宇宙開発は、地球環境の保全、宇宙空間の平和利用、宇宙に関する新知識の獲得等を目指すとともに、開発された先端技術が産業界に波及することが期待される。

宇宙航空研究開発機構では、これらの期待に応えるべく、2003年10月に3宇宙航空関連機関を統合し、研究開発を強化した。しかしながら、過去および直近の相次ぐ人工衛星打ち上げの失敗や人工衛星の不具合に対する本質的な解決策やそれを推進するための経営戦略が不十分であると言わざるを得ない。

そこで、JAXAとしては、宇宙機システム開発のプロセスを抜本的に革新し、研究開発の効率化やシステムの信頼性を大幅に向上させる、技術知識基盤として、知的な一元情報システム ESE (Emergent Synthesis Environment) の開発を推進している。

知的な一元情報システムでは、(1)知識共有環境、(2)設計解析統合環境、(3)業務協調環境、(4)創造的発見支援環境などの開発テーマに分け、研究開発を行っている(図1参照)。宇宙機システムの開発を現在の5年から2年程度、開発コストを現在の1/3程度に削減するのが当面の目標である。

本論文では、知的な一元情報システムの基本である知識共有環境について報告する。

JAXAの技術知識基盤である知的な一元情報システムの知識共有環境の構築では、IMS-VIPNETプロジェクトで開発されたGPM (Generic Product Model) および GPM-DWH (Data Warehouse) [Ota 2002] に注目し、宇宙機システムの開発に利用することにした。GPM および GPM-DWH は、原子力発電プラントに適用され、実用化されている。しかしながら、宇宙機システムへの適用では、宇宙機システムについての用語の整備や知識表現への拡張、オントロジーとしての検討などが必要であった。

そこで、GPMを宇宙機システムの知識モデルに拡張し、GIM (Generic Information & Knowledge Model) と呼ぶことにした。ま

連絡先: 太田吉美, 宇宙航空研究開発機構 高度ミッション研究センター, つくば市千現 2-1-1, TEL:029-868-2141, Fax:029-868-2957 番号, ohta.yoshimi@jaxa.jp

た、GIMのテキスト表現形式である制約自然言語 FNL (Formal Natural Language) も開発している。

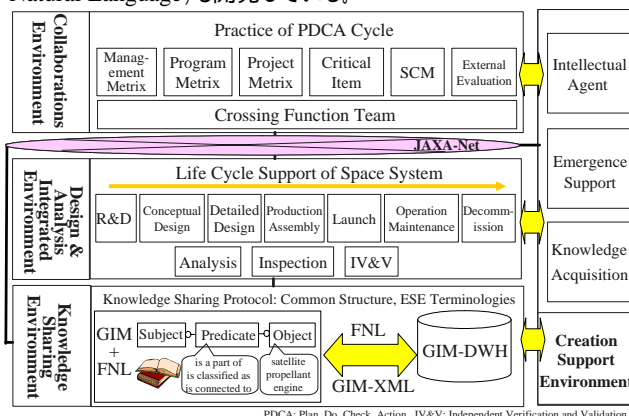


図1 知的な一元情報システム

2. 知識共有環境

知的な一元情報システムの知識共有環境の目的は、大きく二つに分けることができる。一つは、ビジネス・プロセスの革新で、もう一つは情報システムの革新である。

ビジネス・プロセスの革新としては、

(1) 情報システム活用による経営、プロジェクト、業務の継続的改善 (PDCA: Plan, Do, Check, Action) を実施する。

(2) 業務の途切れや断続現象をなくし、業務のシームレス化を実現する。

(3) 局所最適化ではなく、全体的最適化をねらう。などを ESE で実現する。

一方、情報システムの革新では、

(1) 情報システムの孤島をなくし、情報や知識を共有できる環境にする。

(2) 紙による管理をなくし、電子データでの管理をマスターとする。

(3) 重複入力などの無駄のない情報システムにする。

(4) 定型業務は極力自動化する。

- (5) ユーザー・フレンドリーな情報システムを実現する。
- (6) IT 発展に伴う陳腐化へ、対応可能な情報システムとする。などを ESE で実現する。

図2は、現状の業務および情報システム環境を示したもので、各業務および情報システムはそれぞれ個別で、情報もそれぞれ個人が所有している状況を示している。

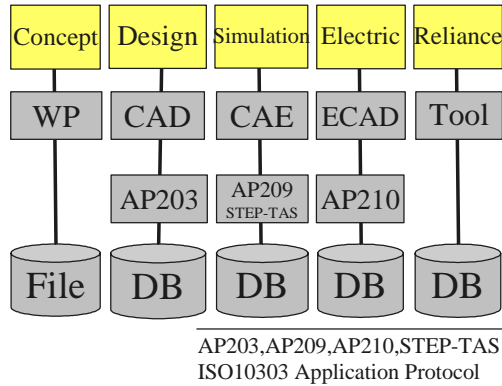


図2 現状の業務環境

個々のアプリケーションは、それぞれ個別のファイルやデータベースを持ち、それぞれ固有の形式でデータや情報を管理しているのが現状である。シームレスなビジネス・プロセスを実現するには、個別のファイルやデータベースの内容を知的に一元化し、共有できる環境が必要不可欠である。

データ交換の標準である STEP (Standard for the Exchange of Product Model Data, ISO10303) は、インターオペラビリティなどに問題があるが、方法論や知的資産として大いに利用する価値がある。また、自分たちに必要なプロトコルは、誰も作ってくれない、必要な人が自分で作らざるを得ない。

そこで、知的な一元情報システムの知識共有環境として、知識共有プロトコルである GIM/FNL および GIM/FNL をベースとするデータウェアハウス GIM-DWH を開発し、業務プロセスと情報システムの革新を図ることにした(図3参照)。

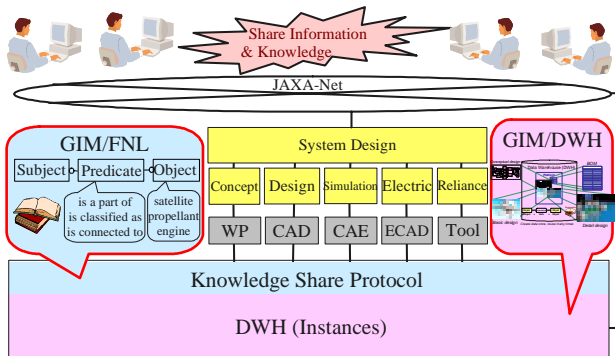


図3 協調的業務支援のための情報環境

宇宙機システムのライフサイクルに互る情報や知識を GIM/FNL で表現し、これを GIM-DWH で蓄積し、これを Web 上で共有、活用する。

データウェアハウス内に蓄積された情報や知識は、いろいろな視点で運用することが必要である。これは、蓄積された情報を基に、製品や技術・サービスのライフサイクルの各アクティビティで必要となる異なる情報表現(マルチビュー)が行えることである。また、蓄積された情報や知識は、一度作成するだけで、その後、何回も使える。

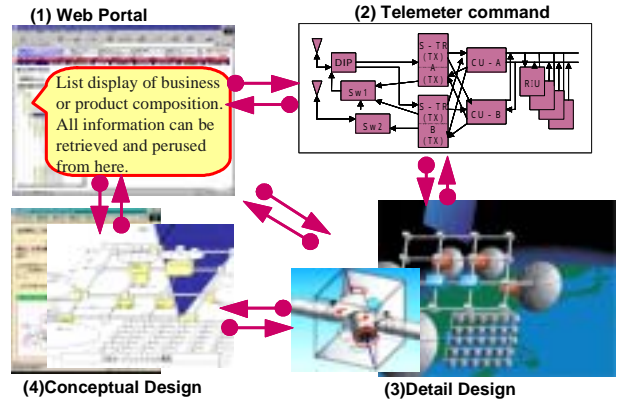


図4 協調的業務支援のための情報共有環境(イメージ)

データウェアハウス内に蓄積した各種設計情報を、インターネットを介して Web 上で、だれもが、いつでも、どこでも、検索・閲覧・活用することが必要である。辞書(リファレンス・ライブラリー)で規定した用語(製品の構成を表す用語)を表示(1)し、これをベースとして、ライフサイクルに互る情報や知識を検索・閲覧する。例えば、(1)からテレメータコマンド名をピックアップして、テレメータコマンド情報を表示(2)し、次にその系統の詳細設計情報を表示(3)し、さらに、(3)の概念設計情報を表示(4)することが容易にできるよう環境を開発している(図4参照)。

3. GIM/FNL

3.1 GIM/FNL の方法論

情報や知識を共有・蓄積するために必要なことは、情報や知識を表す用語(共通言語)を明確にすることである。このようなことは、STEP や他の国際標準でも認識されている。

GIM/FNL では、さらに積極的に、このような用語を汎用的に規定することを提案している。

GIM/FNL では、対象となる製品のライフサイクル全般にわたり、ワークフローを分析し(各作業の入力情報や出力情報を抽出)、情報モデルを構築し、情報や知識の共有・蓄積に必要な用語や用語間の分類、構成などの関連(これも用語)を規定する。このため、GIM/FNL では次の四段階の開発方法を採用している(図5参照)。

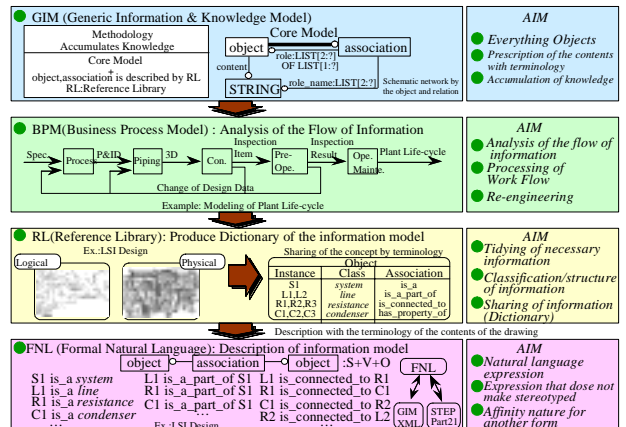


図5 GIM/FNL の方法論

(1) GIM

GIM は、いろいろなデータ構造の基本要素(オブジェクトとオブジェクトの関連(アソシエーション))という要素)を抽出したもの

で、この基本要素を組み合わせることにより、複雑な現実世界を記述しようとするものである(図6参照)。また、GIM は述語(動詞)を中心にした、意味ネットワークモデルであり、自然言語の構文(ステートメント)も記述できる重要な特徴がある。

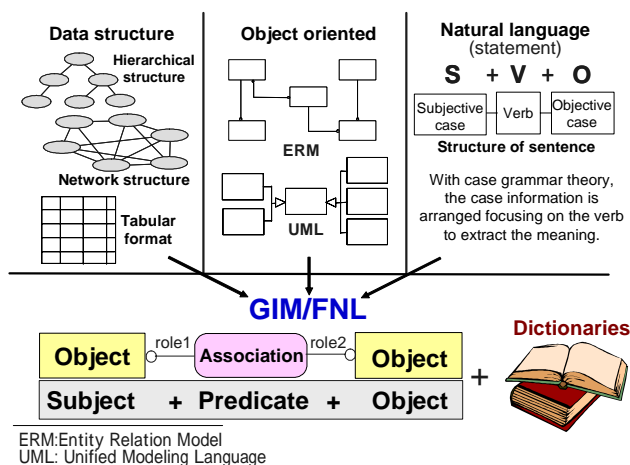


図6 GIM/FNLの原理

GIM は、あらゆる製品や技術・サービスなどのライフサイクルに亙り、情報や知識を蓄積・共有・活用・創出することを支援する知識モデルで、従来のデータ構造やデータベースなどの考え方を革新する技術でもある。

GIM では、製品や技術・サービスのライフサイクル全般を支援する共通の知識モデルとして、オブジェクトとアソシエーション(関連)からなるコアモデルを規定する。コアモデルでは、対象とするものすべてをオブジェクトとして捉える。オブジェクトは、概念やモノなど定義対象の全てを記述するための入れ物(器)である。アソシエーションもオブジェクトの一つ(オブジェクトのサブタイプ)であり、オブジェクト間の関連付けを記述する器で、アソシエーションからオブジェクトには多数の手(ポインタ)を出すことができる。各々手にはそれぞれの役割があり、その役割の内容についても明示的に記述する(図7参照)。

オブジェクトとアソシエーションの中身は、後で述べるリファレンス・ライブラリー(辞書)で規定する。すなわち、分野ごとのリファレンス・ライブラリーを用意する。アソシエーションでは、知識モデルの主要な概念である構成や分類などの用語を規定する。

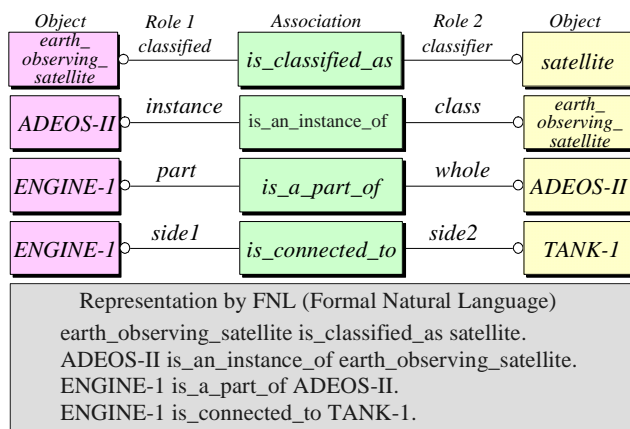


図7 GIM/FNLによる記述例

(2) ビジネス・プロセス・モデル

対象となる分野の業務内容や共有する情報について、関係者間で共通の概念、理解を持つことが必要である。このため、GIM/FNL では、ビジネス・プロセスの分析を行う。ビジネス・プロセスの分析では、対象とする分野のアクティビティの分析を行い、各アクティビティと情報の流れを表現し、アクティビティへの入力情報と出力情報を明確にする。

GIM/FNL では、ビジネス・プロセスの各アクティビティ一つのオブジェクトとして規定する。すなわち、各アクティビティもリファレンス・ライブラリーとして規定する。これにより、各アクティビティ単位の情報や知識をまとめて表現することができる。

(3) リファレンス・ライブラリー (Reference Library)

このステップでは、ビジネス・プロセスで分析した各アクティビティでの入出力情報をまとめ、対象とする分野での情報や知識についての知識モデルを作成し、これをリファレンス・ライブラリーとしてまとめる。

リファレンス・ライブラリーは、さらに概念やモノを規定するリファレンス・クラス・ライブラリー、アソシエーションを規定するリファレンス・アソシエーション・ライブラリー、JIS などの標準部品を記述するエンジニアリング・スタンダード・ライブラリーに分類する。

リファレンス・ライブラリーは、人間が理解できる電子ドキュメント(テキスト・ファイル)として自然言語形式で記述する。

リファレンス・ライブラリーでは、記述対象となる“もの(オブジェクト)”の構成要素やその分類についての用語(クラス)や、そのプロパティ(形状、材質、物理単位など)、アソシエーション(述語)を定義する。構成や分類、所有などの関係は、アソシエーションとして記述する。

(4) 制約自然言語 FNL (Formal Natural Language)

先にも述べたように、GIM には自然言語の構文も記述できる特徴がある。GIM を自然言語の構文で記述するのが制約自然言語である。図7で、主語、述語、目的語の順につなげると一つの文章(earth_observing_satellite is_classified_as satellite.)となる。文章や文書を組合せることにより、知識を記述することができる。

制約自然言語の基本構造は、主語 + 述語 + 目的語で、主語や述語、目的語の取り得る内容をリファレンス・ライブラリー(辞書)で制限する。これにより、一般の自然言語処理の難しさを排除する。辞書は、英語、日本語、中国語、韓国語などの多言語に対応できる。

FNL は、情報システム間で情報や知識を共有するため、リファレンス・ライブラリーとインスタンス(実体で固有名称や具体的な数値を表す)をコンピュータ内で処理するための外部表現形式である。FNL は、コンピュータだけでなく、人間にも理解できる自然言語形式である。

GIM/FNL では、人間の世界で規定したリファレンス・ライブラリーから自動的にコンピュータが解釈できる表現形式(FNL で記述したクラス・ライブラリ、アソシエーション・ライブラリー)に変換される仕掛けになっており、用語や用語間の関連について、その矛盾や誤りをパーサーでチェックできる。

インスタンスは、各種情報システムのトランスレータを介して、FNL として出力または、入力される。インスタンスに対しても同様なチェックを行うことができる。

3.2 GIM/FNL の特徴

GIM/FNL は、情報や知識を共有するための知識モデルである。GIM/FNL は、情報や知識の表現方法(オブジェクトとアソシエーションからなるシンプルな構造)を共通にし、オブジェクトやアソシエーションの内容をライブラリー(情報や知識を共有する

ための用語辞書)として規定(標準化)することにより、情報や知識を表現するものである。このライブラリーを拡張・拡充することにより、情報や知識を体系的に「積み上げる」(オントロジー)[溝口 1999]ことができる画期的な技術である。

GIM/FNL は、オブジェクトとアソシエーションから成る意味ネットワークモデルであり、人間の頭脳のように、ニューロン(オブジェクト)とニューロンをシナプス(アソシエーション)で結合する単純なモデルと類似している。このことはニューロンやシナプスを増やすことで、将来の発展や拡張に対応できる可能性を秘めた、成長モデルと見ることもできる。

GIM/FNL は、リファレンス・ライブラリーを替えることにより、他の産業分野でも利用できる。他の分野に適用するには、その分野固有の辞書を追加することになる。この辞書を豊富にすることにより、知的資産の相互運用性(Interoperability)が可能となる。

4. GIM/FNL の宇宙機システムへの展開

GIM/FNL のねらいは、あるドメイン(世界)の情報や知識を体系的に積み上げ、蓄積・共有・活用・創出できるようにする技術(体系化)を確立することである。

ドメインを記述するためには、

- (1)ドメインについての共通の認識・理解(概念の共有)
- (2)ドメインを明示的に記述する方法(モデル)

が必要である。GIM/FNL では、3章で述べたようにビジネス・プロセスの分析を通して、ドメインについての共通の認識・理解を得るを行っている。

GIM/FNL はドメインの構成要素をオブジェクトとし、それらの関連を示す述語からなる集合体として明示的に記述するものである。

ドメインをさらに分析すると、オブジェクトは、概念(クラス)と実体(インスタンス)に分けることができる。概念と実体は関係しており、is_an_instance_of(=is_a)なる述語で記述する。実体(インスタンス)の記述では、分類関係を表す述語 is_classified_as は現れない(図8参照)。

GIM/FNL では、概念(クラス)や述語などを用語辞書として定義し、必要に応じて拡張できる仕掛けになっている。

宇宙機システムのライフサイクルにわたる情報や知識を体系的に積み上げるためには、宇宙機システムについてのオントロジーが必要不可欠である。図9は、GIM/FNL による宇宙機システムの構成管理を記述したもので、右側のブロック内の四角形がクラスで、左側のブロック内の四角形が実体(インスタンス)で、六角形が述語で、is_a(=is_an_instance_of)は、インスタンスとクラスを関連付ける述語である。また、構成は、is_assemble_from(逆の関係は is_a_part_of となる)なる述語で記述する。クラス間の分類には、is_classified_as という述語を用いる。辞書では、宇宙機システムの構成要素以外に、物理量や形状などを表す属性やアクティビティ、組織、述語などの用語を記述する。辞書は拡張でき、新たな概念や知識の追加が可能である。

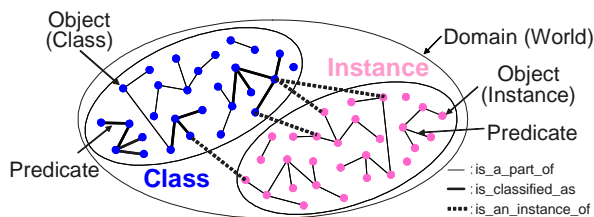


図8 概念(クラス)と実体(インスタンス)

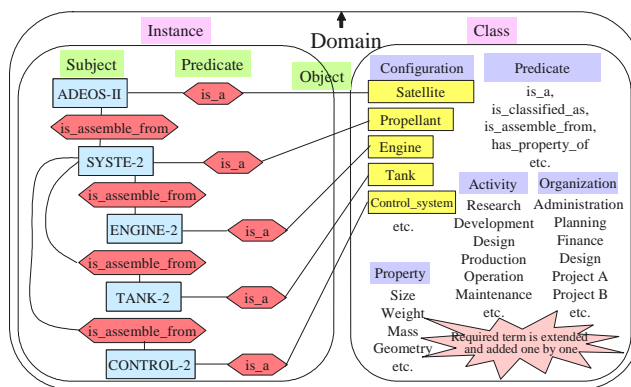


図9 宇宙機システムの記述例

GIM/FNL の宇宙機システムへの展開では、既存のオントロジーとして、STEP に注目した。STEP では、製品や技術についての用語の整備が既に行われており膨大な知的資産となっている。そこで、STEP の基本リソースの部分を、GIM/FNL のリファレンス・ライブラリーとして展開している。現状では、形状や物理単位などの共通化できる用語を中心に約 2000 語の整備が進んでいる。

GIM/FNL は、述語を中心とする知識モデルであり、知識処理の理論である述語論理として取り扱うことができる。また、GIM/FNL は一つのオントロジーであり、知識としての意味を表現できる。論理学に「意味の理論」であるオントロジーを加えることにより、様々な問題解決を試みることの可能性が広がる。今後は、GIM/FNL の応用としてこれを検討したい。

5. まとめ

宇宙開発のライフサイクルに亘る業務支援システムは、知識共通プロトコルで一元的に連携され、データウェアハウスにより、一元的に情報や知識を蓄積・管理し、これを Web で共有することにより、新たな創造性や創発性が発揮できる環境を実現する。

GIM/FNL は、情報や知識の表現方法(オブジェクトとアソシエーションからなるシンプルな構造)を共通にし、オブジェクトやアソシエーションの内容をライブラリー(情報や知識を共有するための用語辞書)として規定(標準化)することにより、情報や知識を表現するものである。このライブラリーを拡張・拡充することにより、情報や知識を体系的に「積み上げる」(オントロジー)ことができる画期的な技術である。GIM/FNL は、宇宙航空分野だけでなく、あらゆる分野の技術知識基盤の中核として適用できる。

以上のようなことができようになると、業務も大きく変わる。例えば、宇宙開発に係わる情報や知識を、だれでもが、いつでも、どこでも、個別なシステムを使わずに、Web 環境で検索・閲覧・活用できる。

また、情報や知識の共有により、業務をシームレス(整流)化でき、業務の効率化が図れる。また、情報や知識の共有により、エンジニアリングの質や信頼性が大幅に向上する。さらに、調達先やユーザーとの情報連携も可能となる。

参考文献

[Ota 2002] Yoshimi Ota, Yuji Naka, others: Construction of Technology Knowledge Infrastructure (TECHNOINFRA), WWDU2002, 6th International Scientific Conference on Work With Display Units (2002-5)
 [溝口 1999] 溝口理一郎: オントロジー研究の基礎と応用, 人工知能学会誌, Vol.14, No.6, 1999