社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2005－ＩＣＳ－139 (27)
2005／3／16

セマンティックアノテーションに基づく多言語コラボレーション支援

杉山　香織[†1]　船越　要[†3]　神田　智子[†1,2]　藤代　祥之[†1]　石田　亨[†1,2]

[†1]京都大学情報学研究科　〒606-8306 京都市左京区吉田本町

[†2]科学技術振興機構 CREST デジタルシティプロジェクト　〒606-0924 京都市中京区一之船入町 366

[†3]NTT サービスインテグレーション基盤研究所　〒180-8585 武蔵野市緑町 3-9-11

E-mail: kaori@kuis.kyoto-u.ac.jp, kf@cslab.kecl.ntt.co.jp, koda@desitalcity.jst.go.jp

fujishiro@kuis.kyoto-u.ac.jp, ishida@i.kyoto-u.ac.jp

あらまし 多国間でのソフトウェア開発などの協調活動では，参加者間での情報の共有を進めることによって，共通の理解を創出することが重要である．しかし，参加者間での情報の共有を妨げる要因として (1) 機械翻訳を用いて多言語化された環境においては，機械翻訳の誤訳によって，情報の不伝達や誤解が起こること (2)プロジェクトでは，多様なツールが利用されるため，情報が分散して管理されていることの二つがあげられる．そこで，ツール間で使用されるデータの相互運用を実現するために，プロジェクトで用いられるデータを記述するためのメタデータ集合を RDF スキーマによって定義し，このメタデータ集合でタグ付けされたデータに対してアノテーションをつけるための枠組みを提供する．本論文では，このメタデータ集合を用いた，ツール間でのデータの相互運用および，アノテーションによるコラボレーション支援について述べる．

キーワード　RDF，アノテーション，多言語コラボレーション

Supporting Multilingual Collaboration with Semantic Annotation

Kaori SUGIYAMA[†1], Kaname Funakoshi[†3], Tomoko KODA[†1,2],

Yoshiyuki FIJISHIRO[†1] and Toru ISHIDA[†1,2]

[†1]Department of Social Informatics, Kyoto University　Kyoto, 606-8501 Japan

[†2]JST CREST Digital City Project　Ichino-funeiri-cho 366, Nakagyo-ku Kyoto, 606-8501 Japan

[†3]NTT Service Integration Laboratory　3-9-11 Midori-cho, Musahino-shi Tokyo, 606-8501 Japan

E-mail:　†kaori@kuis.kyoto-u.ac.jp, kf@cslab.kecl.ntt.co.jp, koda@desitalcity.jst.go.jp

fujishiro@kuis.kyoto-u.ac.jp, ishida@i.kyoto-u.ac.jp

Abstract　In software development project among multiple countries, information sharing and collaboration among participants are crucial for the success of the project. But there are two problems that hind information sharing among participants. (1)In multilingual environment with machine translation, mistranslation by machine translation causes miscommunication and misunderstanding. (2)Various tools are used for a project and information is managed by each tool separately. To approach these problems, we defined a metadata set by RDF schema that can describe the data that is generated in a project to realize interoperability of the data among tools and also defined metadata to annotate the data that is described with the metadata set. In this paper, we propose a framework for interoperability of the data among tools and supporting collaboration with annotation.

Keyword　RDF, metadata, annotation, multilingual collaboration

1. はじめに

世界的なインターネット人口の増加に伴い，従来，英語が中心であったインターネット上の利用言語も実世界の言語の分布に応じ，多様化が進んでいる．特に東アジア圏でのインターネット利用者の増大は目覚しい．インターネットの普及は遠隔地間でのコミュニケ

ーションのコストを低下させる．今後，このような環境の下，アジア圏内においても国境を越えたソフトウェアの共同開発などのプロジェクトはますます増えていくものと思われる．

しかし，特に東アジア圏では，英語が公用語として十分に機能しておらず，かといって，たとえ隣国同士であっても互いの母国語はほとんどわからない．よって，複雑な思考を必要とするような作業は，それぞれは自分の母国語を用いて行いたいというのが実情であり，このような言語障壁が多国間での共同プロジェクトを難しいものにしている．

このような遠隔地間での共同作業を支援するためには，E-mail やビデオ会議システムなどをはじめとするコミュニケーション技術の利用が有効である．また，言語障壁を克服するために，これらのツールと機械翻訳を組み合わせ利用することができる．

この際の問題点として，以下の二つの問題点に対応する必要がある．

機械翻訳ノイズによる情報共有の失敗　機械翻訳を介した多言語環境でのコラボレーションでは，機械翻訳の誤訳によるノイズの高い文章を理解したり，またそのようなたくさんの文章の中から必要な情報を見つけ出したりすることが必要となる．しかし，プロジェクトの参加者にとって，このような作業は負荷が高く，結果，内容が理解できなかったり，また，必要な情報を見逃してしまったりするといったことが起こりがちである．これによって，参加者間での情報の不伝達や誤解などが生まれ，プロジェクトの進行を妨害する結果となっている．

ツール間でのデータの相互運用の必要　一般にプロジェクト内部では E-Mail やビデオ会議システムをはじめとする各種のコミュニケーションツールやプロジェクトの進捗管理のためのツール，文章共有のためのレポジトリなど，多様なツールによって，プロジェクト内での情報が分散して管理されているという現状が上げられる．しかし，これらの情報はタスクの実行や話題において，互いに密接に結びついているにもかかわらず，これらの情報を統合する方法がない．

ここでは，これらの問題に対して，次のような方法を提案する．

まず，プロジェクトの中で用いられる各種ツールのデータを一元的に記述することのできるメタデータを用いることを考える．なぜなら，各種のツールは大きく特性が異なるものであるため，すべての機能が統合されたツールを作成するよりも，ツールは別個に作成し，データだけを必要に応じて相互運用するほうが容易だからである．

また，機械翻訳のノイズを克服するための方法としては，過去に機械翻訳を用いて掲示板上で議論を行った共同開発実験の分析から，議論の参加者間で現在のトピックは何かなどの，投稿内容を理解するうえでの文脈となる情報を共有することによって，情報共有が円滑に行われるということがわかった．(2 章)

そこで，本研究では,プロジェクトで用いられる各種の情報を記述することの出来るメタデータ集合を RDF スキーマによって構築した．次に，このメタデータに記述された情報に対してのコメントやコンテンツ間の関連，情報の属性などといったアノテーションを付与することができるようにこのメタデータの拡張を行った．各ツールがこのメタデータに沿ってデータを記述し一つのデータベースに保存することによって，ツール間でデータやアノテーションを相互に利用できることとなる．(3 章)

最後に，この枠組みの上で構築されたツール，および，これらのツールと同じプラットフォーム上で構築されたツール間での連携の例を示す．

2. アジアブロードバンド異文化コラボレーション実験

近年の東アジア圏でもインターネットの普及に伴い，多国間での共同プロジェクトを低コストで行える環境が整ってきている．その一方で，東アジア圏では英語のような公用語が存在せず，プロジェクトメンバ間の意志の疎通が難しい．機械翻訳やコミュニケーション技術を活用することによって，このような言語障壁を乗り越えるため，異文化コラボレーション実験が行われた[3]．　2003 年度に行われたアジアブロードバンド異文化コラボレーション実験[9]には日本と中国の7つの研究機関と大学から，研究員，教員，学生合わせて34名が参加し，機械翻訳によって多言語化された掲示板やビデオ会議システムなどを用いて，異文化コミュニケーションの支援方法やその実現方法などについて議論を進めるとともに，多言語環境における強調活動のためのツールを作成するためのプラットフォームの構築などを共同で行った．

2.1. 多言語環境におけるディスカッション

この章では実験結果の解析から導き出された結果および，そこから多言語環境でのコラボレーションを支援するための方法について考察する．

2.1.1. 非同期型多言語コミュニケーション実験の概要

プロジェクトでの議論は主に多言語化された電子掲示板である TransBBS（図 1　参照）を用いて行われ，

実験終了後にこの議論の解析が行われた．

TransBBSは機械翻訳を介することにより，参加者がそれぞれの母国語を用いて議論を行うことのできる掲示板である．投稿したメッセージは自動的に翻訳され，各国の言葉で原文と並べて表示される．また，利用者はメッセージを投稿する前に，その翻訳結果を参照し，正しく翻訳されるまで，投稿文の修正を行える機能がある．

TransBBS を用いての議論は一ヶ月行われ，その間，各週に参加者には大まかな議論のテーマが与えられ，各議論のテーマごとにディスカッションルームが用意された．それぞれのディスカッションルームに司会者が選ばれた．この期間中1000件を超えるメッセージが投稿された．議論への参加者は国や大学が異なり，また実験前の事前の面識などはほとんどなかった．

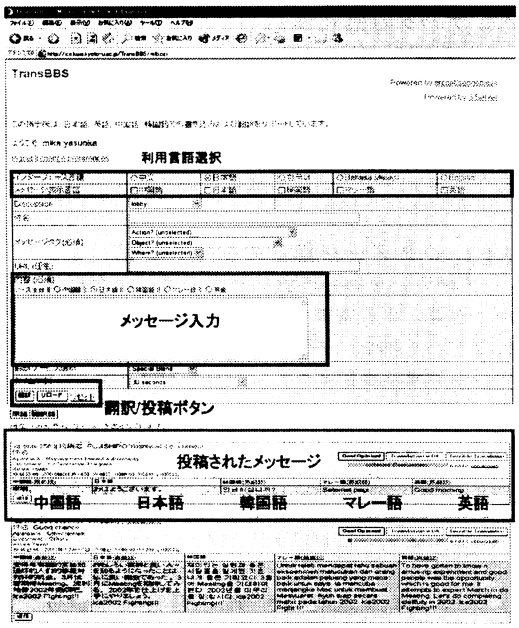

図 1:TransBBS

2.1.2. 実験からの考察

機械翻訳の利用による利点

実験後のアンケートから，機械翻訳を利用することは他国語や英語も不得手とする参加者にとって，難しい議論の内容でも自国語で発言できるため，たとえば英語などで議論する場合と比べて，自分が間違った言葉を使っていないかなどといった心配をすることなく積極的に発言できるといった，肯定的な意見が多く，言語障壁を取り除くことに役立っていることが見て取れた．また，形式的な文書の交換だけではなく実際に議論することによって，参加者間の一体感を得られるといった効果があることもわかった．

機械翻訳の利用による問題点：参加者間でのコンセンサス獲得の失敗

翻訳精度の悪い機械翻訳を介してのコミュニケーションは参加者間で共通の理解が得られないという事象を引き起こした．たとえば，実験終了後のアンケートによって，各参加者にそれぞれの議論でどのような結論に達したかということをたずねたところ，すべての参加者が議論で共通の結論に達したと認識しているにもかかわらず，それぞれが結論として考えている内容を比較すると日本人参加者と中国人参加者の間でまったく反対の結論に達しているという例が見られた．

つまり，議論において参加者間でコンセンサスが取れず，またそのことに参加者当人が気がつかないという重大な問題が発生した．

この議論の経過を調べたがスレッドが分断されているなどといった明らかに議論が破綻していることが一見してわかる特徴は見られず，日本人，中国人の間での応答も頻繁にみえるため，議論は順調に進んでいるように見える．

しかし，内容を詳細に見ていくと，二つの原因を見てとることができた．

1．トピックの取り違え

スレッドの構造など表面上は会話がつながっているように見えるにもかかわらず，内容を検証すると，返信のメッセージとその前のメッセージとのあいだで内容が分断してしまっている部分が散見された．

例えば図 2では，中国人参加者は画像処理についての話題を行っており，その一例として，商品について言及している．しかし，その翻訳結果を読んだ日本人参加者は商品の話をしているという勘違いをしてしまった．

誤訳を多く含んだ文章を読む際には，理解できた単語や話の流れを元に，推測しながら読むことになる．しかし，その結果として，反応が言葉尻を捕らえたものになりがちで，本来の意図が伝わらないということが起こってしまうのである．

2．意見の相違の修正がうまくできない

先に述べた例においても，意見が相違した場合には質問がなされたり，司会者によって議論で決まったことは何かということが確認された．このような質問や確認があったときは意見の相違に気づき，意見の相違を修正する機会となるはずであるが，このようなメッセージに対し，参加者からの反応がみられなかった．

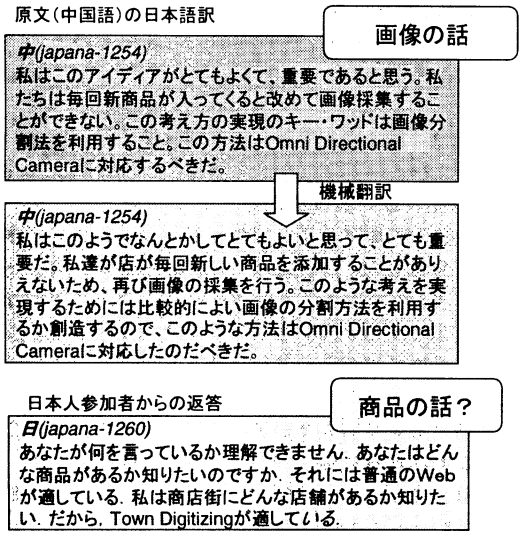

図 2：トピックの取り違えが起こった例

翻訳ノイズの克服：議論進行上の工夫

一方，議論がうまく一致した例を調べると，自分が発言する話題の種類に対して，文頭で，明確に述べる．もしくは発言内容に対して司会者が指定したカテゴリ（例：[Tag]，[Tool]，[Demo]など）を Subject 欄に書き込むなどというように，自分が話す内容を明確に表すように推奨された．

また，司会者は重要な発言に対し参加者に注意を向けるように促す発言や議論の要約と次の話題への方向づけを行う発言が 4,5 通の投稿に対して一件と非常に頻繁に行われている．

つまり，先に述べた問題に対して，それぞれ

1．発言者は発言する内容を明確に述べる

2．司会者が，重要な情報を見逃さないように参加者に注意を喚起する

という二つの方法で，うまく対処がなされていた．

セマンティクス支援の必要性

実験から，各メッセージのカテゴリを明確にしたり，重要なメッセージを示すなど，情報を理解するうえでの文脈となるような情報を豊富に提供することが，情報そのものの理解を容易にするといえる．

このような情報の提供は実験ではそれぞれの参加者が文中に記すことによってなされていた．しかし，この方法は議論に付きっ切りの当事者にはわかりやすいが，事前資料のように掲示板以外に重要な情報があったり，また，後で情報を見直したりするときにはあまりうまく働かない．そこで，情報の重要度やトピックのカテゴリなどの属性，注釈などのセマンティクスを明示的に情報に付与するための仕組みが必要であると考える．

3. 多言語コラボレーションを支援するためのメタデータ集合の構築

コラボレーションで生成される情報を有効にメンバ間で共有するためには，ツール間での情報の相互運用が不可欠である．

このためには，各ツールの情報を一元的に取り扱える枠組みが必要となる．そこで，まずツール間での情報の相互運用のために各ツールで生成された情報を共通のメタデータ集合に基づいてタグ付けし，データベースに保存することを考える．さらに，このデータの上にアノテーションや属性，データ間の関連性などを記述するためのメタデータを構築する

3.1. データの相互運用のためのメタデータ集合の構築

この章では，プロジェクト内でのコミュニケーションやプロジェクトの進行状況の管理などのために使われる各種のツールで扱われる情報をタグ付けし管理するためのメタデータの構成について述べる．[5]

3.1.1. メタデータ集合の構築の方法

メタデータ集合は各ツールでデータを記述するために必要とする語彙をそれぞれ過去のログデータの分析などから抽出，整理したものを元に，これらを統合する形で構築を進めた．メタデータは RDF スキーマを用いて構築された．このため，データはクラス階層によって表現される．

例えば，一見異なるように見える E-mail や BBS のスレッド，テキストチャットの一つのセッション，ビデオ会議システムにおける一つのセッションなどは同じ「議論のコンテキスト」としてまとめることができ，全て，参加者や開始日時，終了日時，セッションの中で取り交わされたインタラクションの内容など同様のプロパティを持っていると考えることができる．このように，類似する情報をまとめ，階層的なクラスに整理することによって，メタデータ集合は構築された．

3.2. データの相互運用のためのメタデータ集合の概要

メタデータ集合は RDF スキーマによって構築されている．このクラス階層を表 1 に示す．（なお，このメタデータ集合の完全な仕様は[4]を参照のこと．）

表 1:プロジェクトオントロジーのクラス階層

Primary Class	Secondary Class
pr:Project	
pr:Task	
pr:Plan	
pr:Resource	pr:Budget
	pr:Tool
pr:Agent	pr:Group
	pr:Person
	pr:SoftwareAgent
pr:Interaction	pr:Message
	pr:LogItem
pr:Content	
pr:Context	

これらのクラスは大きく分けて，プロジェクトの構造を表すためのクラス，人と資源のためのクラス，プロジェクト内での参加者の活動を表現するためのクラスの 3 つに分類される．以下，それぞれを説明する．

プロジェクトの構造のためのクラス

プロジェクトオントロジーの Project クラスはプロジェクトで管理される情報の全ての起点となるクラスで，管理されるプロジェクトの名称や説明，またプロジェクトのタスクやスケジュール，プロジェクトで利用できるリソースやユーザの活動の結果できた各種のデータへのリンクを持っている．（図 3）

プロジェクト内で実行されるべきタスクとその実行スケジュール，をあらわすために Task，Plan のクラスが用意されている．

人と資源のためのクラス

Resource クラスではプロジェクトを実行する上で利用可能な，予算やソフトウェアを記述するためのクラスである．

Agent クラスでは各プロジェクト参加者やタスク実行のためにできたグループ，ソフトウェアによるサービスなどプロジェクトを実行する主体についての情報を管理するクラスである．

参加者活動のためのクラス

協調作業ツールにおける参加者の全ての活動はインタラクションであるとみなせる．ここで，各インタラクションは，ドキュメントや E-mail のメッセージ本体などの内容および，BBS のスレッドやチャットのセッションなどのコミュニケーションが行われたコンテキストを持っている．よって，参加者の活動を記述するために，インタラクション自体を現す Interaction クラス，インタラクションの内容を表す Content クラス，およびインタラクションのコンテキスト文脈をあらわす Context クラスの三つを用いて表現される．さらに Interaction クラスは，参加者同士のインタラクションを表現する Message クラスと，機械とひとの間のインタラクションを表す LogItem クラスの二つのサブクラスを持つ．

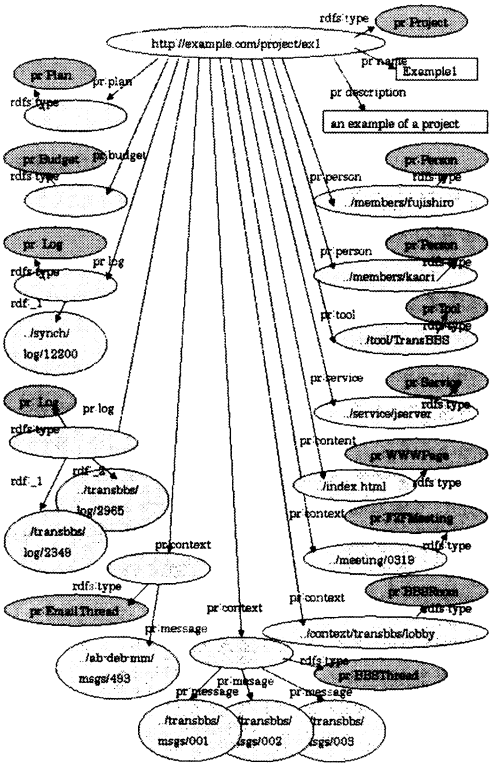

図 3:プロジェクトの記述例

3.3. アノテーション付けのためのオントロジーの拡張

前述のメタデータ集合によって，ツール間でのデータの相互運用ができるようになった．この章では，これらのデータをより有効に活用するために，この枠組みの上でデータ間を関連付けたり，各データに注釈を付加するためのメタデータを紹介する．

コンテンツ間の関連付けのためのプロパティ

コンテンツ間の関連付けを行うために，Content クラスはさまざまなプロパティを持っている．例えば，参照関係を示すための reference プロパティやコンテンツが改訂された場合，改訂されたコンテンツから古いコンテンツへのリンクである replaces プロパティである．

コンテンツの分類のためのクラス

コンテンツを分類するためのクラスとして Attribute クラスが用意されている．このクラスは例えば，「2 月 3 日の会議」「重要」「スケジュール」などといった，属

性ごとにコンテンツを分類する．このクラスは属性名を表すプロパティとこの属性に分類されるコンテンツへのリンクを持っている．

注釈付けのためのクラス

コンテンツへの注釈付けのためのクラスとして，Annotation クラスがある．注釈もまた，参加者によって作り出されるコンテンツの一種であるため，Annotation クラスは Content クラスのサブクラスとして，定義されている．このクラスは注釈の文章そのもの，注釈の作成者，注釈の作成日などの値を持つためのプロパティをもっている．

4. アノテーションを用いた多言語コラボレーション支援

先に述べたメタデータによって，ツール間での情報の共有が可能になる．図 4は RDF データベースを介して，各ツールの間で情報が共有される様子を示す．プロジェクト内では BBS やビデオ会議システムなどを介して参加者間のコミュニケーションが進む．これら各ツールのなかで交わされたデータは全て RDF データベースに保存される．この際に各データには URI が振られ，タグ付けされる．

共通の部分については同じメタデータに沿ってデータが記述されるため，ツール間でのデータの共有も可能となる．例えば，参加者の個人データなど，プロジェクト全体で共有されるデータも同一のものを利用することが可能となるため，ツールをまたいで同一の作成者によるデータを引き出すといったことも容易である．また，あるツールで作成されたコンテンツを別のツールで利用することもできる．どのツールによって作られたのかにかかわらずアノテーションの付加を行えるので，コンテンツ管理システムなどで，例えばある会議に関連する E-mail, BBS のスレッド，ドキュメントのすべてに[会議]といったアノテーションをつけておけば全てをまとめて引き出してくるといったことも可能になる．

5. おわりに

機械翻訳を介した多言語環境でのコラボレーションでは誤訳が多発することが参加者の意思の疎通を妨げる．これを克服するために，多種のツールにまたがる情報の共有を進めるための仕組みとして，プロジェクト内の情報全てを記述できるメタデータの構築を行い，さらにこれらの情報を参加者が活用するためにアノテーション付けを行い整理する方法について提案した．最後に，この枠組みの上で構築されたツール間の連携の例を示した．

文　献

[1] Dublin Core Metadata Element Set, Version 1.1: Reference Description, 1999, DCMI Recommendation

[2] 野村早恵子，石田亨，船越要，安岡美佳，山下直美:アジアにおける異文化コラボレーション実験2002：機械翻訳を介したソフトウェア開発 ,情報処理 Vol.44, No.5, pp.503-511.

[3] アジアブロードバンド異文化コラボレーション実験(2004).
http://ice.kuis.kyoto-u.ac.jp/ice/pub/ABB2003Report.pdf

[4] Project Vocabulary Specification
http://ice.kuis.kyoto-u.ac.jp/project_rdf/rdf/

[5] Kaname Funakoshi, Kaori Sugiyama, Toru Ishida, Takashi Yoshino, Jun Munemori, Haijun Zhang and Zhongzhi Shi. Semantic Interoperability in Tools for Intercultural Collaboration. Third International Conference on Active Media Technology (AMT-05), 2005.

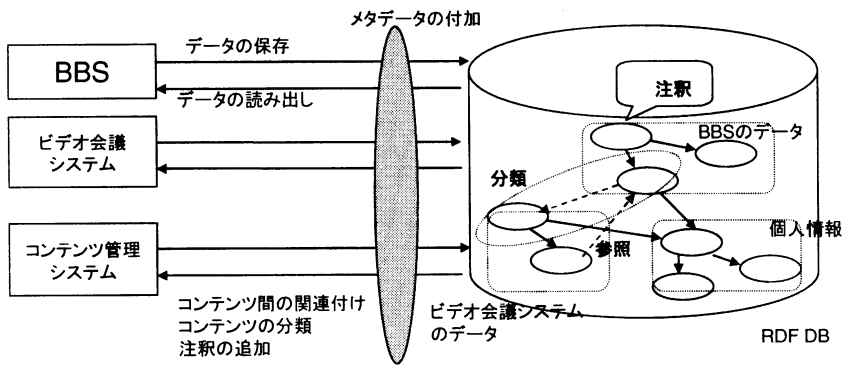

図 4：ツール間の連携