

# 6 知識処理技術

## 6 Knowledge Processing Technology

### 6-1 知識処理基盤の概要

#### 6-1 An Overview of Knowledge Cluster Systems

是津耕司

ZETTTSU Koji

#### 要旨

今や Web は、社会や環境の活動を反映した膨大な情報の記憶媒体 (social memory) としての役割を果たしている。これらの膨大なデータを総合的に分析しながら、潜在的な法則性を発見したり、様々な分野や組織で個別に蓄積されてきた情報を横断的に連結・統合したりすることで、世界規模の知識ネットワークを構築し、様々な社会問題や環境問題を解決できるようになると考えられる。我々は、Web をより知的な情報獲得と分析の環境へと進化させることを目的として、次世代知識処理基盤の研究開発を行っている。本論文では、次世代知識処理基盤を構成する、相関分析による異種・異分野の横断検索技術と、グリッド基盤に基づくサービス指向知識処理技術について述べる。

Today, Web is regarded as a “social memory” of extremely-large amount of information on natural/social environments surrounding our daily lives. Constructing a global-scale knowledge network interconnecting multi-domain, heterogeneous information will be a crucial of knowledge discovery and analysis equipped with interdisciplinary data integration and information fusion. We conduct research and development of Knowledge Cluster Systems aiming at shifting the Web to a next-generation of knowledge information infrastructure. In this paper, we explain two major components of our knowledge information infrastructure; interdisciplinary correlation search of heterogeneous knowledge bases, and service-oriented knowledge information management based on grid computing.

#### [キーワード]

異分野相関検索, サービス指向アーキテクチャ, ナレッジ GRID, ナレッジクラスタシステム, 言語グリッド

Interdisciplinary correlation search, Service-oriented architecture, Knowledge GRID, Knowledge cluster systems, Language grid

### 1 はじめに

今日 Web 上では、従来の Web ページに加え、ブログやチャットなどネットワークを介して行われるユーザ同士のコミュニケーションの記録、オンラインショッピングやライフログのような人々の生活の記録、さらにセンサ等によって取得された時々刻々と変化する環境情報など、実に様々な

情報が配信されている。今や、Web は、自然環境や社会活動を反映した膨大な情報の記憶媒体 (social memory) としての役割を果たしている。これらの膨大なデータを総合的に分析しながら、潜在的な法則性を発見したり、様々な分野や組織で個別に蓄積されてきた情報を横断的に連結・統合したりすることで、世界規模の知識ネットワークを構築し、様々な社会問題や環境問題を解決で

きるようになると考えられる。我々は、Webをより知的な情報獲得と分析の環境へと進化させることを目的として、次世代知識処理基盤の研究開発を行っている。

我々の知識処理基盤システムは、グリッド基盤と異分野相関検索エンジンによって構成される(図1)。従来のWebのようにデータを配信・共有するだけでなく、特定のデータを意図的に集めたり、集めたデータの中から様々な話題や出来事に関する情報を抽出したり、抽出した情報を様々なつなぎ合わせたり、つなぎ合わされた情報を検索したり閲覧したりする機能を備えている。例えば、地球気候変動と治安問題のつながり(相関関係: correlation)を知りたい場合、従来の検索エンジンでは“地球気候変動 治安問題”というキーワードを含むWebページを検索し、検索結果の内容をユーザが直接自分で判断するしかなかった。一方、我々が開発しているシステムは、同じ質問に対し、例えば「この数年、北アフリカで異常高温と水紛争が頻発している」というような、キーワードで直接指定されていない内容にまで範囲を広げて相関のある情報を見つけ出すことができる。

相関検索エンジンを使って、Webページはもちろん、ニュース配信や気象観測データまで多岐にわたる大量のデータを対象に、それらの意味的なつながりと時空間的なつながりのあるデータを発見する。その際、従来のように、予め共通の辞書を用意しておく必要がないため、異種・異分野のデータの相関も拡張性高く発見することができる。相関検索エンジンは、時間、空間、主題に関する様々な種類の特徴量を使ってデータを索引付

け、問い合わせ処理の際は、相関を発見するのに最適な特徴量の組み合わせ(相関の文脈と呼ぶ)と、それらを用いて高い相関を示すデータ集合を同時に検索する。その結果、先程の例のように、「この数年(時間の特徴量)、北アフリカで(空間の特徴量)異常高温と水紛争が(主題の特徴量)頻発している」という文脈の中で高い相関を示すデータ集合を発見することができる。我々は、意味空間モデルと、Moving Phenomena 時空間モデルによる相関分析手法を提案しており、いずれも多次元空間における部分空間選択により最適な特徴量の組み合わせと相関データ集合を柔軟性高くかつ効率的に見つけ出すことができるのが特徴である。

我々の知識処理基盤システムの全ての機能は、グリッド基盤の上に実装されている。アジアやヨーロッパなど世界中の拠点をつないだ広域グリッドネットワークを構築し、その上にサービス指向アーキテクチャに基づいたグリッド環境を実現している(図2)。グリッド基盤上には、データ収集と情報抽出、相関分析、ユーザインタフェース・インタラクションに関する多種多様なソフトウェアサービスが、各拠点で並行して開発され配備されている。これらのサービスを組み合わせ連携させることで、様々な次世代Webアプリケーションを実現することができる。これまでに、Webをハイパーリンクとは無関係に相関関係に沿って閲覧する“Link-free Web Browsing”、場所や時期、テーマに応じて相関性のある情報を次々と推薦する観光案内システム、企業や組織の製品・サービス・技術に関する情報をWebサイトや特許から幅広く収集し、様々なテーマの下でマッチメイキングを行うパートナーシップマネジメントシステムなど、様々なアプリケーションを開発してきた。グリッド基盤は、アプリケーション開発のためのサービス連携機構と開発言語を提供している。また、アプリケーションでの使われ方に基づいて役立つサービスを検索する機構(サービス発見エンジン)の研究開発も行っている。

本稿では、**2**で相関分析による異分野横断検索技術について述べた後、**3**でグリッド基盤を用いたサービス指向知識処理について述べる。最後に、**4**でまとめと今後の展望について述べる。

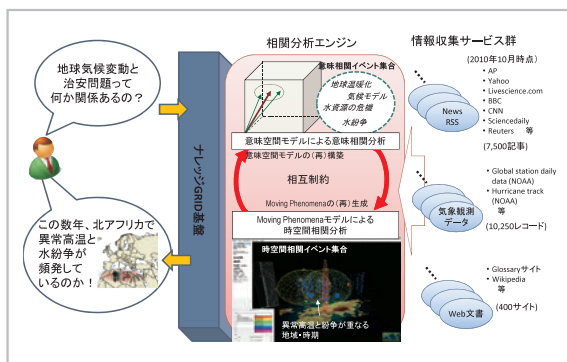


図1 ナレツクラスタシステムによる相関検索

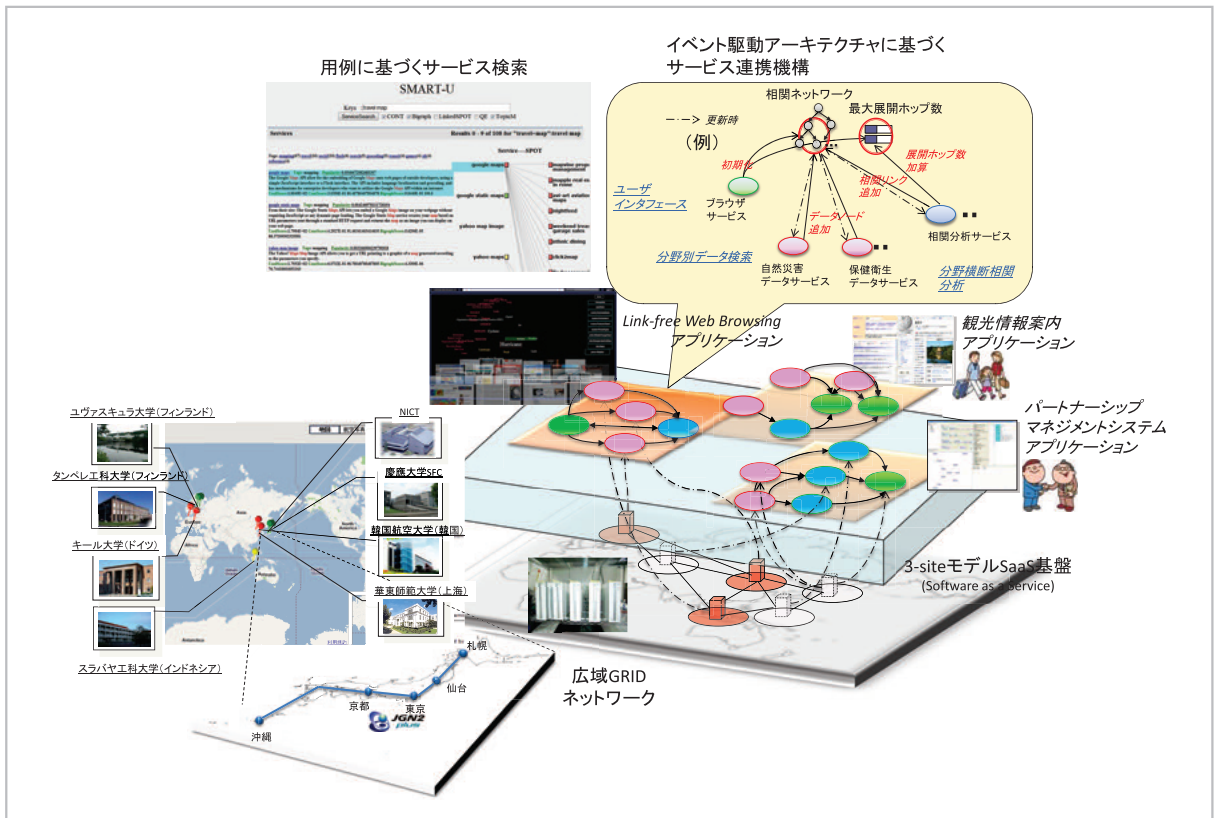


図2 サービス指向アーキテクチャに基づいたグリッド環境

## 2 関連分析に基づく異分野横断検索

一般的に、ある出来事に関する情報を整理し深い分析を行うためには、その出来事に特化した専門分野の情報を取得するだけでなく、そこから派生したり影響を受ける様々な分野の情報も関連性とともにも収集することが大切である。例えば、「地球温暖化」について知るためには、気候変動による自然災害だけでなく、各種産業、経済、エネルギー問題、さらには保健衛生や教育など多岐にわたる分野の関連情報を収集する必要がある。これらの情報は、様々なWebサイトで独自にまとめられ個別に提供されている。従来のWeb検索エンジンを使った場合、各分野ごとに検索キーワードを用意し、何度も検索を行いながら、ユーザ自身が膨大な検索結果を集約しなければならない。もし、相互に関連性のある情報を分野横断的に検索し分かりやすく提示できるシステムが実現できれば、Webを有効活用し実世界の様々な出来事についてより網羅的に把握することができるようになると思われる。

### 2.1 異分野知識ベース連携

これまでに我々は、分野の異なる様々な知識ベースを横断的かつ動的に連携させ、分野の壁を越えて関連性のある情報を検索・配信するための情報基盤“ナレッジクラスタシステム”[1]の研究開発を行ってきた。我々が提案する異分野知識ベース連携フレームワーク[2][3](図3)では、異分野の知識ベース間の連携を実現するために、以下の2つの基本機能を定義している：

**Intra operation:** ある知識ベースに閉じた相關検索を行う。入力として与えられた概念タームに対し、知識ベース内の概念タームとの相關の強さを計量し、相關値の高い上位の概念ターム集合を出力する。

**Inter operation:** 2つの知識ベースの間の相關検索を行う。例えば2つの知識ベースAとBの間のinter operationでは、知識ベースA、Bそれぞれに含まれる概念ターム集合の間の写像関係を記述した相關マトリクスを作成し、2つの知識ベースの概念ターム間の相關を、相關マトリクスによって構成



される意味空間（ベクトル空間）における距離として計算する。

これらの2種類の基本機能を様々な組み合わせることで、異なる知識ベースをまたがる相関検索を拡張性高く構成することが可能になる。検索結果は、図4に示すような意味相関ネットワークとして表現される。図4の例では、クエリとして与えられた“volcanic gas”から始まり、まず火山災害知識ベースのintra operationにより“lava flow”、“mud flow”、“CO2”、“SO2”、“pyroclastic flow”、等の火山災害分野の概念タームが導出される。なお、エッジに付与された数字は、相関値を表している。次に、これらの概念タームを入力として、火山災害知識ベースと環境知識ベースとをつなぐinter operationにより、環境分野で相関の高い“air pollution”、“hydrogen sulfide”、“heavy metals”、“ground water”、“rough fish”、“acid rain”、“global warming”等の概念タームが導出される。また同時に、火山災害知識ベースと医療知識ベースをつなぐinter operationにより、同じ入力に対し、医療分野で相関の高い“bronchitis”、“pulmonary edema”、“bronchial asthma”、“obstructive lung disease”などの概念タームが導出される。最後に、各分野の知識ベースに対応するintra operationにより、それぞれの分野内で相関の高い概念タームが導出される。結果として、クエリである“volcanic gas”と相関の高い環境分野と医療分野の概念ターム集合が、導出過程を示す意味相関ネットワークとともに得られる。

我々の提案方式の特徴は、異分野の知識ベースを連結し横断的に相関検索を行う際、一連の相関関係をどのような文脈で形成するかを、inter operationによって柔軟に切り替えられることである。例えば、上記の例で、各inter operationが地球温暖化に関する情報から作られた相関マトリクスを用いていれば、検索結果の意味相関ネットワークは地球温暖化の文脈においてクエリと相関の高い各分野の概念タームの導出過程を表すことになる。その他に、我々は、特定の災害に関する文脈（例えば、“Unzen Fugendake,” “Miyake Jima”、“Sidoarjo Mudflow”など）での相関検索も実現している。これらを実現するために、我々はRSSデータから相関マトリクスを自動生

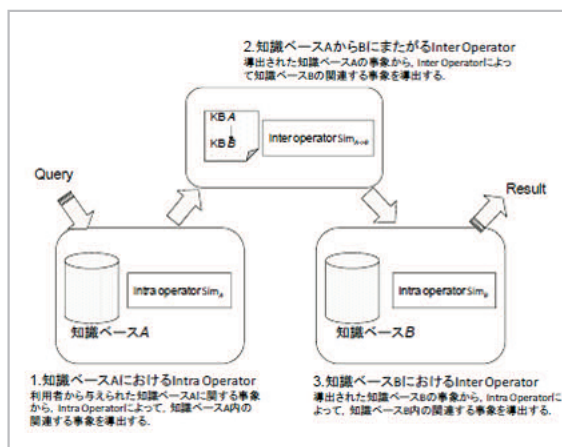


図3 異分野知識ベース連携フレームワーク

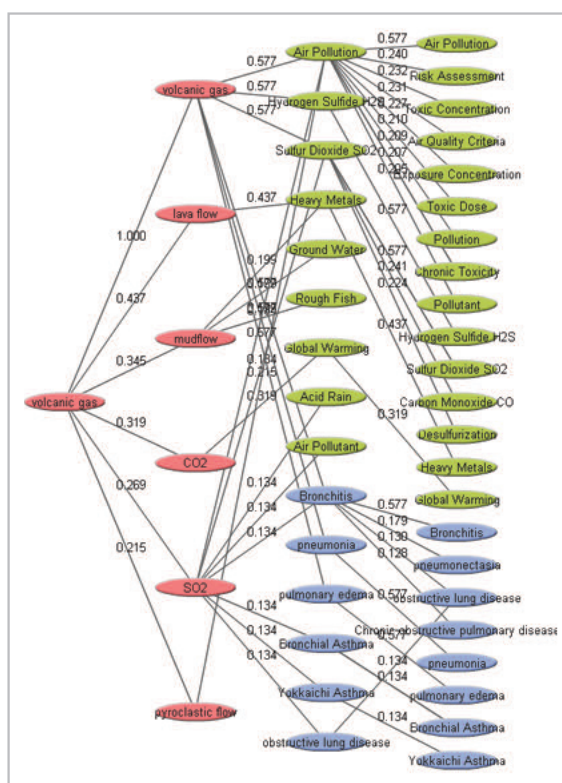


図4 火山災害分野の知識ベースから環境分野、医療分野の知識ベースへの相関検索結果を示す意味相関ネットワーク（クエリ：“volcanic gas”）

成する仕組みを実装した（図5）。RSSはWebのニュース記事配信等に広く用いられており、様々な分野の情報がRSSデータの中に網羅され、相関マトリクスを作成する情報源として有用性が高い。収集されたRSSデータは、相関検索の文脈に相当するテーマごとにグループ分けさる。Inter operationでつなぐ各々の知識ベースに含

まれる概念タームの組み合わせに対し、収集した RSS データの中にその組み合わせが多数出現していれば、相関マトリクスに正の相関を設定する。

## 2.2 Link-free browsing

Link-free browsing [3]-[5] は、異分野間知識

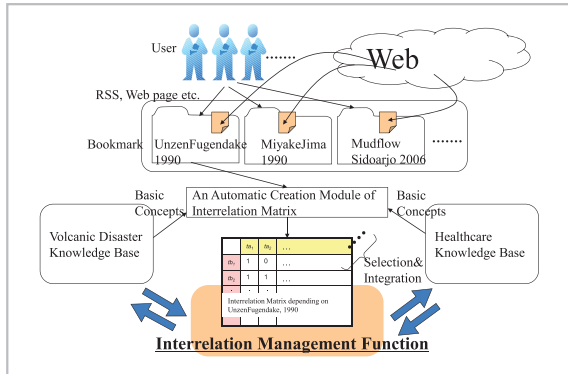


図 5 RSS データを用いた相関マトリクスの自動生成

ベース連携を Web 閲覧に応用したシステムである。従来の Web 検索では、たとえ関連性が高くてもユーザの専門外の分野についてはキーワードすら思いつかず、重要な情報にたどり着けない場合が多く、link-free browsing はこうした問題の解決につながると期待される。図 6 に、link-free browsing システムの概要を示す。Link-free browsing システムは、以下の様に動作する：

### 1. コンテンツ閲覧モード

通常の Web ブラウザと同様に、Web ページの内容を閲覧する。

### 2. ビューポイント選択

コンテンツ閲覧モードで表示されている Web ページから興味のある語句を選択してハイライトすると、システムがその語句から相関検索が可能な知識ベースを“ビューポイント”として提示す。ユーザがあるビューポイントを選択すると、対応する intra/inter operations がシステムに設定さ

図 6 Link-free browsing システム

れる。

3. 意味相関ネットワーク閲覧モード  
システムは、ビューポイントに基づいて、ユーザが選択した語句から様々な分野への相関検索を実行し、意味相関ネットワークを生成する。また同時に、意味相関ネットワークの各ノードが示す概念語に対応するWebページを、画面の下部にサムネイル表示する。サムネイルをクリックすると、Webページが表示されコンテンツ閲覧モードに移行する。

このように、link-free browsing では、静的なハイパーリンクによる通常の Web 閲覧と、相関検索による動的な意味相関ネットワークの生成・閲覧を切り替えながら、Web ページの閲覧と関連する概念の検索を繰り返し、段階的に理解を深めるような Web 閲覧が可能になる。

### 2.3 Sticker: 時空間相関に基づくクラスタリング

異分野を横断的に検索する場合、意味的な相関だけでは十分ではない。例えば、未曾有の災害による影響は、むしろ時間的・空間的な周辺を調べた方が、意外な相関性を発見できることが多い。そこで我々は、時空間的な相関関係に基づいて情報をクラスタリングする方式として、Sticker (Spatio-Temporal Information Clustering and

Knowledge ExtRaction) を提案している [19]。図7にStickerのGUIを示す。Stickerでは、地理空間と時間軸を合わせた三次元空間の中に、位置と時刻が付与された情報(イベント情報と呼ぶ)をプロットし(図7(a))、この空間内での近さや位置関係により様々なクラスタリングを行う。Stickerの特徴は、クラスタを形成する際の様々な時空間的制約条件を、この三次元空間内での図形として表現していることである(図7(b))。例えば、興味のあるイベント群(points of interest: POI)を囲む極小矩形(minimal boundary box: MBB)、あるイベントを中心とした周辺範囲を表す円(circle)や球体(sphere)、これらの時間的な動きを表すpipeやcorn、さらに、これらを組み合わせた複合図形などである。また、ユーザがこれらの図形を操作することで、クラスタリング条件を視覚的に調節できることもStickerの特徴である。さらに、これらの図形の間のような包含関係を調べることによって、複数のクラスタ間の間接的な相関性も把握することができる。例えば、単純な重なり(overlap)から、時間が進むとともに重なる(join)、離れる(leave)、合併する(merge)、分離する(split)、あるいは2つの図形が一定の時間幅・空間幅を置いて並行する(along with)などが挙げられる。各クラスタに含まれる情報は、タグ・クラウド(主要なキーワード集合の提示)やトレンドグ

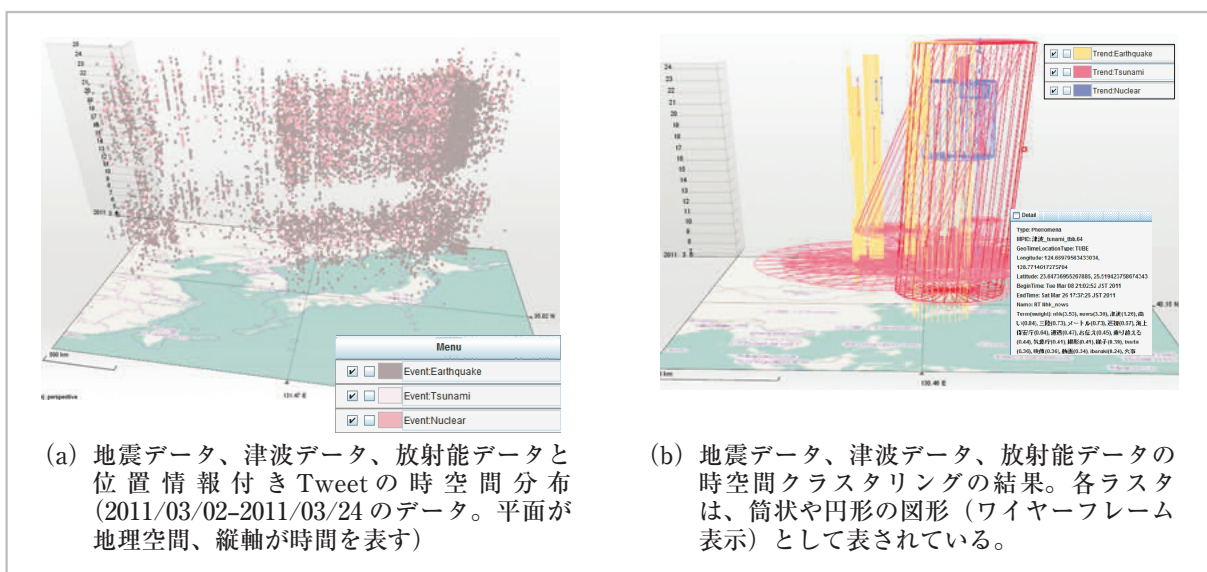


図7 Stickerを用いた時空間相関に基づく情報クラスタリング



ラフ（出現頻度の高いキーワードの時間的変遷を示したグラフ）など、様々な方法でアウトライニングされる。

StickerのGUIを通じて行われるクラスタリングやクラスタ操作は、背後にあるMoving Phenomena データベース管理システム (DBMS) [20]によって実行されている。Moving Phenomena DBMSは、SQLを拡張したデータ定義言語 (data definition language: DDL) とデータ操作言語 (data manipulation language: DML) を持ち、イベント情報のデータモデルと各種クラスタリング操作を定義している。

### 3 グリッド基盤によるサービス指向知識処理

#### 3.1 ナレッジGRID 基盤

今日のネットワーク社会において、知識集約的な作業では、組織や国、文化、分野の壁を越えたコミュニケーションや協働が頻繁に行われる。ナレッジGRID [1][6]は、分散知識処理環境で協調して問題解決を行うための情報基盤である。元来、グリッドコンピューティング基盤の上に分散知識発見のための機能レイヤーを追加した parallel distributed knowledge discovery and data mining (PDKDD) を表す Knowledge Grid [7]の概念を、我々は以下の点に焦点を当て拡張している：

**横断的集合知の活用：** Web上では、様々な人や組織がコミュニティを形成し、情報の共有や交換を行っている。こうしたソーシャルネットワークごとに様々な集合知が情報資産として形成されるが [8]、基本的にこれらの情報資産は分野やコミュニティごとに閉じたものである。しかし、災害対応や環境問題などでは、分野やコミュニティの壁を越えて、これらの情報資産を横断的に利活用できることが必要である。また、「参加のアーキテクチャ (architecture of participation)」により、個々の利用者が発信するデータを集約し、よく利用される知識は残り、そうでないものは消えていくという「利用者による価値の選定」を機能させることも重要である。

**仮想組織 (Virtual Organizations) の動的形**

**成：** 近年、産業界や環境科学でよく見られるように、異なる分野の組織が協調して問題解決にあたる際、連携する組織間で、データやツール等の情報資産をタスクごとに動的に共有することが行われる。こうした、横断的かつ動的な情報資産共有を実現するため、グリッド基盤上で複数の組織にまたがる仮想組織 (virtual organization: VO) を動的に形成する [9]。情報資産の利用規則 (誰が、何を、どう利用するか) を明確に定義し、この規則を共有する組織がVOに参加するという、セミオープンな情報共有環境を実現する。

**サービス指向知識処理：** サービス指向アーキテクチャ (service-oriented architecture: SOA) [10]は、あらゆる情報資産をサービスとして統一的に扱うパラダイムであり、従来のコンピューティング環境に大きな変革をもたらすものとして広く認知されている。サービス指向知識ユーティリティモデル [11]は、いつでも、どこでも、誰でも知識処理を共有材として使えるようにするための基盤技術であり、ユーザの要求や課題に応じてサービス化された知識処理が動的に組み合わせられ、実行環境に合わせ構造や振る舞いを変化させる。また、知識ベースも、従来の関係DBから、より柔軟な知識表現を扱えるグラフDBやトライプルDBへと移行する。

#### 3.2 システム概要

ナレッジGRID上の知識サービスは、以下の3種類に分類される：

**Knowledge discovery サービス：** 情報資源の知識記述を提供することを目的として、分類、要約、オントロジーなど、様々なデータマイニング手法によるメタデータ生成の仕組みを実装する。

**Knowledge association サービス：** knowledge discovery サービスが提供する情報資源の概念レベル記述 (メタデータ) を用いて、情報資源の間の様々な相関関係を生成する。このサービスは、主に 2.1 で述べた異分野知識ベース連携フレームワークに基づいて開発される。

**Knowledge delivery サービス：** knowledge discovery サービスと knowledge association

サービスからの出力を構造化したり、可視化したりする機能を提供する。例えば、2.2で述べた link-free browsing は、この knowledge delivery service に相当する。

ナレッジ GRID のアプリケーションは、上記のサービスを様々に組み合わせることで実現される。ナレッジ GRID では、以下の2つのサービス連携方式が提供されている：

**ワークフローモデル：** サービス連携の手順 (procedure) を定義し、一般的なプログラミング言語に見られる条件分岐 (if-then-else) やループ (while) 等の実行フロー制御を行う。ナレッジ GRID では、Web サービスのワークフロー連携におけるデファクトスタンダードである Web service business process execution language (WS-BPEL) [12] を用いている。

**イベント駆動モデル：** サービス間で共有するデータを決めておき、どの共有データのどのような変化に関心があり、その変化が起きた時どう振る舞うかを宣言的に定義する。ナレッジ GRID では、従来のイベント駆動モデルを、複

数サービスが協調して情報資産を利活用する目的に特化させ、データ中心のイベント駆動サービス連携を実現する仕組み (Service MeshUp) を提供している。

これら2つのモデルは相互に変換可能であるが、ビジネストランザクション等の手続き的なプロセスに適した従来のワークフローモデルとは異なり、我々の Service MeshUp モデルは、集合知形成や状態監視など、自己組織的なプロセスや長い期間継続するプロセスに適している。

本節の後半では、データ中心イベント駆動モデルである Service MeshUp について説明する。図8に Service MeshUp の記述例を示す。Service MeshUp では、1つのアプリケーションは aspect の集合として構成される。各 aspect は、アプリケーションの中の単一のタスクや機能を表し、サービス間で共有されるデータ (aspect properties) を持つ。アプリケーション開発者は、各 aspect に組み込まれるサービスを指定し、1) それぞれのサービスが

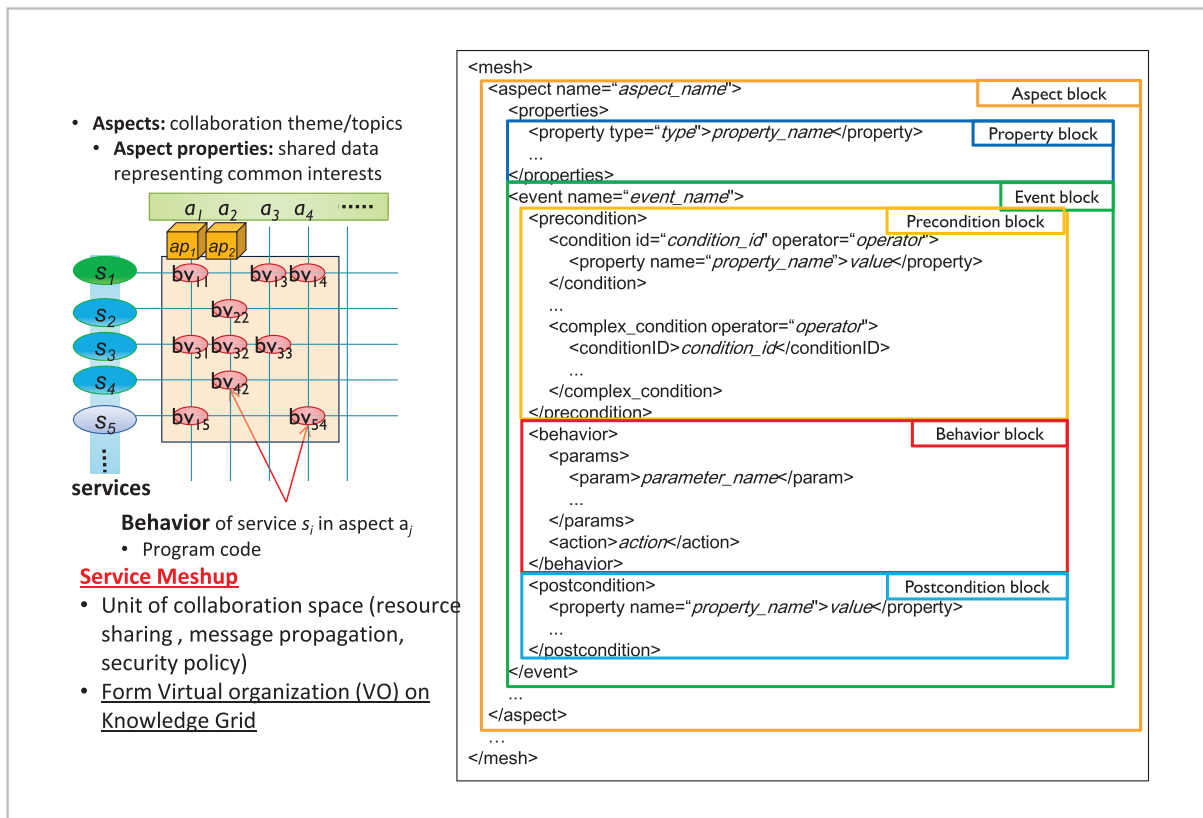


図8 Service MeshUp 記述の例



呼び出されるための aspect property の条件 (precondition)、2) その条件で呼び出された際にサービスが実行する内容 (behavior)、3) サービス実行後の aspect property の更新 (postcondition) を定義する。

ナレッジ GRID は、グリッドネットワーク、サービスプラットフォームおよびアプリケーションの 3-tier 構造で構成される。グリッドネットワークは、業界標準の Globus Toolkit [13] ミドルウェアを用いて実装される。従来のハイパフォーマンス計算機クラスタを指向したグリッドとは異なり、ナレッジ GRID のノードは、Web サーバのように、世界中に分散している。これらの GRID ノードは、セキュアな仮想ネットワークによって接続され、それぞれのノード上で各種知識サービスが独立かつ並行して開発されている。一方、サービスプラットフォームは、知識サービスの実装、配備、検索、連携の役割を担う。サービスプラットフォームは、以下のモジュールから構成される (図 9) :

**Mesh runtime:** Service MeshUp 記述を解釈し知識サービス連携を実行する

**Mesh repository:** Service MeshUp 記述を保存し検索する

**Service runtime:** ローカルあるいはリモートの GRID ノード上にある知識サービスを呼び出す

**Service repository:** 知識サービスのカタログを作成し、Service MeshUp 記述で使われるサービス名と Service runtime がサービス呼び出しに使う end-point-references の対応付けを行うこれらのモジュールは、以下の様に動作する。まず、Mesh runtime は、Service MeshUp 記述をインタープリットし、Service repository を検索してサービス名から end-point-references (EPRs) を解決する。これらの EPR 集合に基づいて、Mesh runtime はサービスが配備された GRID ノード群から構成される VO を作成する。この VO の中で、セキュアな aspect property データの共有やサービス呼び出しが行われる。VO が形成されると、Mesh runtime により aspect property が初期化され Service Me-

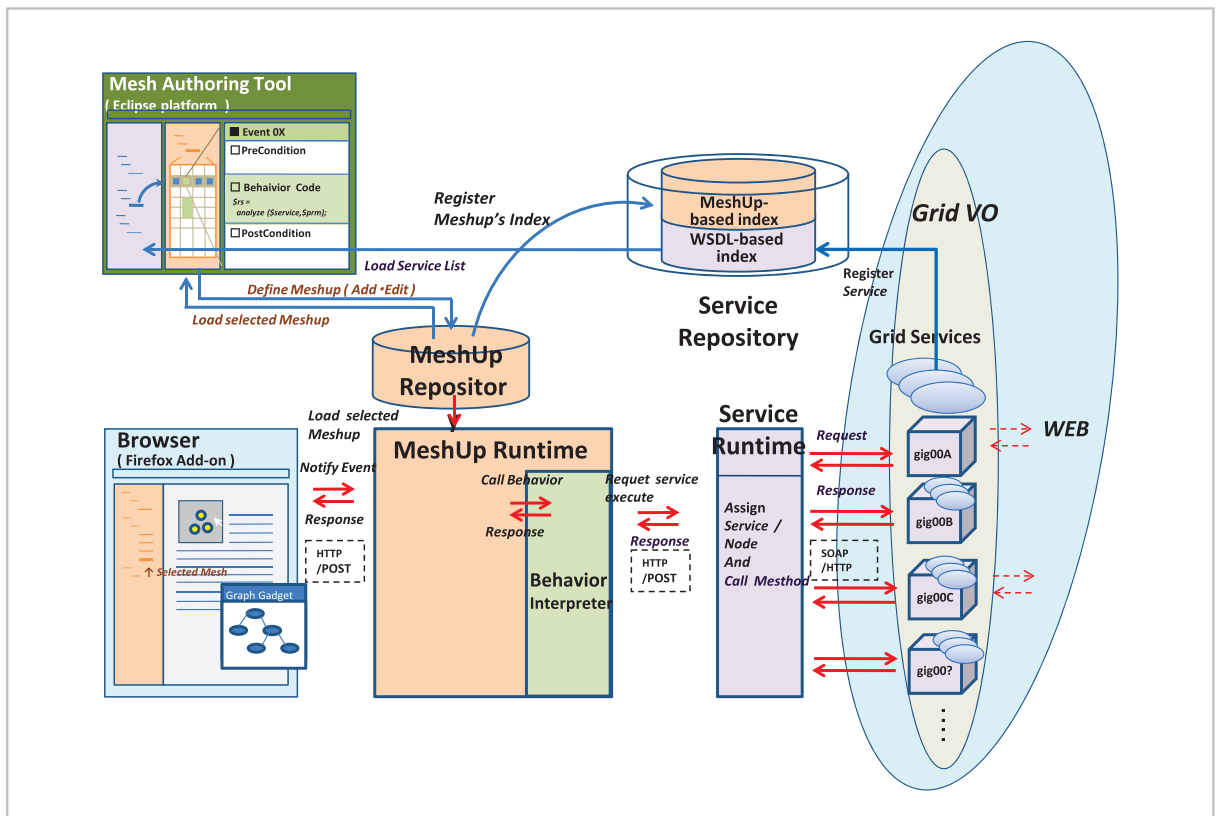


図 9 ナレッジ GRID サービスプラットフォームのシステム構成

shUp の実行プロセスが開始される。実行中は、Service runtime によって知識サービスが呼び出される。Service repository は非集中的に管理され、知識サービスの情報はまずローカルの GRID ノードに登録され、それから他の GRID ノードへと伝搬される。

### 3.3 言語グリッド

言語グリッド [14] は、異文化コラボレーションを支援するためのサービス指向の多言語基盤である。辞書などのデータや機械翻訳などのソフトウェアが言語資源として CD やダウンロードサービスで配布されている現状から、インターネットに接続すればすぐに利用できる言語サービスへと転換し、利用者が異文化コラボレーションの環境に合わせて言語サービスを容易に組み合わせられるようにすることを目的としている。言語グリッドは、言語サービスの収集と共有、連携を可能とする「基盤ソフトウェア」と、基盤ソフトウェアに登録された言語サービスを容易に利用可能とする「異文化コラボレーション環境」を提供する(図 10)。言語グリッドでは、特に以下の点に注力している：

**サービス指向の多言語基盤の構築：**言語サービスを蓄積し共有するには、標準インタフェースを持つ原子サービスに基づいてサービスを連携する基盤ソフトウェアが必要である。さらに、利用者がそれらの言語サービスを用いて異文化コラボレーションを支援するアプリケーションシステムを簡単に開発できなければならない。言語グリッドは、P2P サービスグリッド、原子サービス、複合サービス [15]、アプリケーションシステムといった 4 つの階層 [16] からなる。P2P サービスグリッド [17][18] により、コアノード間で情報共有が行われることで、サービス利用者はいずれのコアノードからも同じサービスを利用でき、サービス提供者はどのコアノード上でも統一したアクセス制御を行うことができる。

**運用モデルの制度設計：**サービス指向集合知には、多様なステークホルダーが存在する。利用者はそれぞれ異なる要件を持ち、提供者も異なるポリシーを持つ。運営者はこれらの異なるステークホルダーを協調させるよう、両

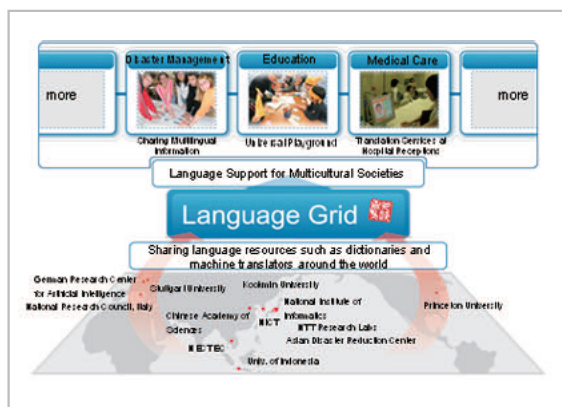


図 10 言語グリッドの概要

者のインセンティブを考慮した運用モデルを設計しなければならない。言語グリッドでは、異なるステークホルダーを協調させるための運営モデルを提案している。この運営モデルは、サービス利用者のインセンティブをサービス提供者のインセンティブに合致するように設計されている。さらに、言語サービスのアクセシビリティを向上させるために、複数の運営組織による連携を可能にする連邦制運営モデルを提案している。

**ユーザ参加型デザインの実践：**提供される言語サービスが多ければ多いほど、利用者はそのサービスによる利益を享受できる。つまり、サービス指向の集合知を形成するには、利用者とコミュニティを積極的に参加させることが必要である。言語グリッドでは、サービス指向のアプローチと汎用の異文化コラボレーション支援ツールにより、ユーザ参加型デザインを加速させている。事例として、利用者である学校や NPO が、言語グリッドプレイグラウンド (図 11 a)、言語グリッド Toolbox を用いて専用の異文化コラボレーション環境 (図 11 b, c) を開発した事例が挙げられる。

言語グリッドは、異文化コラボレーションを支援するための、サービス指向集合知に基づく多言語基盤であるが、その基盤ソフトウェアや運用モデルは言語に特化したものではない。新たにサービスのインタフェースを定義することで、他のドメインへの応用が可能である。今後は、大規模なサイエンスデータやアーカイブデータを対象としたデータサービスや、大規模データを分析するた

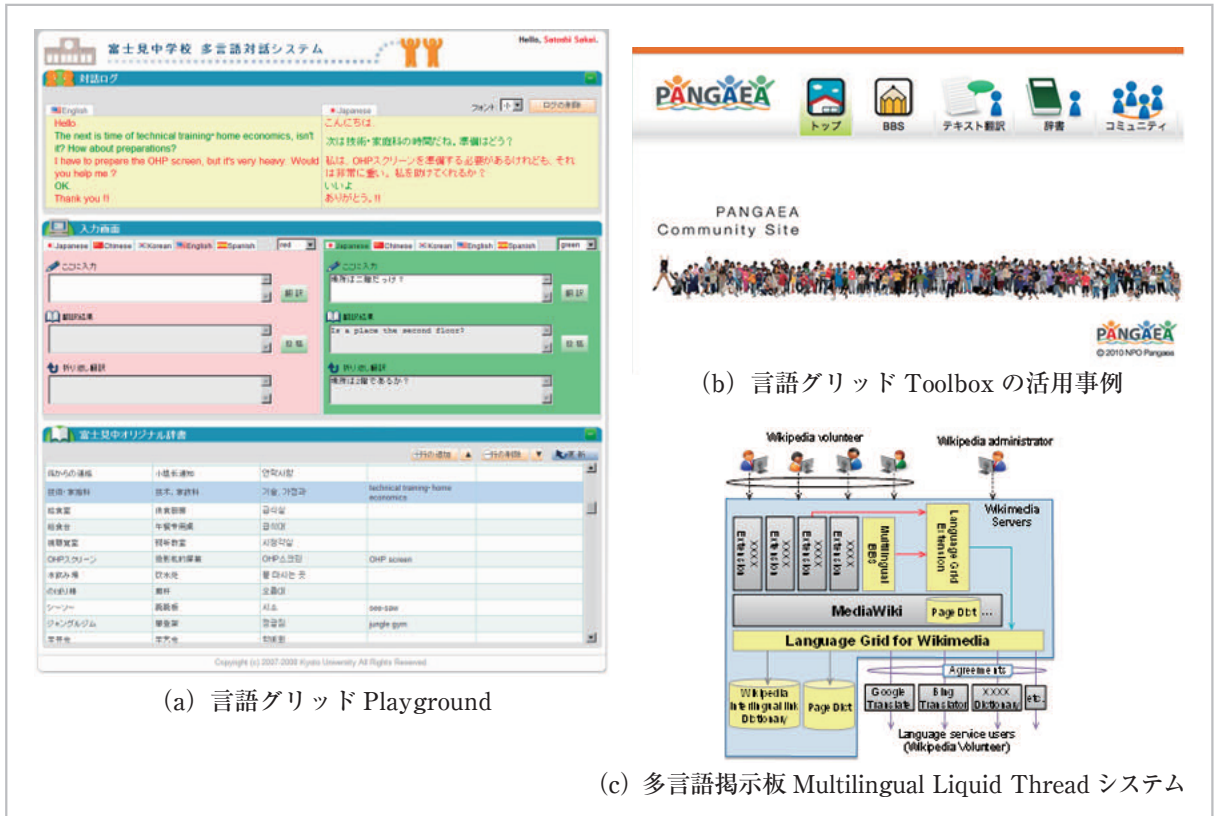


図 11 言語グリッドの応用例

めのサービスを体系化することで、サービス指向集合知のアプローチに基づき、ビッグデータの活用を促進するサービス基盤の開発に取り組んでいく予定である。

#### 4 今後の展望

eScience やデータ中心科学は、実験科学、理論科学、計算科学に次ぐ第四の科学と言われる、様々な分野にわたる大量のデータを分析することで潜在的に存在する法則性や関連性を発見・検証する科学である。社会的必要性を重視し、これまでの科学的探究により得られた知見を実社会での意思決定や行動支援に生かすことを目的に、科学関連のデータと社会関連のデータの相関を応用指向で発見する。例えば、「ある自然災害によって影響を受ける社会活動は何か？」というような質問を扱う。このような質問に対しては、従来のような科学的モデルの構築が困難なため、異種・異分野のデータの相関を柔軟かつ拡張性高く発見する技術が有効に働くことが期待される。こうした

背景に基づき、我々は、分散情報処理基盤における横断的データマネージメントサービスの研究開発において、宇宙・地球環境の科学データと新聞記事などの社会データを対象に研究開発を進めている。

一方、グリッド基盤は、新世代ネットワークにおける「価値を創造するネットワーク」サービス連携基盤としての展開を進めている。従来のサービスコンピューティングのパラダイムを拡張し、サーバ、ストレージ、ネットワーク、端末から、ソフトウェア、通信手段、人手処理まで含むあらゆる ICT 資源をサービスとして抽象化し、ネットワークの制約を意識することなくシームレスにサービス連携が可能になる技術の開発に取り組んでいる。また、サービス連携を構成する要素技術（サービス・アドレッシング、メッセージング、サービス発見、協調制御など）を、ネットワーク基盤上にダイレクトに実装することで、サービス同士の水平連携やサービスと ICT 資源間の垂直連携を、高いパフォーマンスとスケーラビリティで実現する方法の開発にも取り組んでいる。



## 参考文献

- 1 Zettsu, K., Nakanishi, T., Iwazume, M., Kidawara, Y., and Kiyoki, Y., "Knowledge Cluster Systems for Knowledge Sharing, Analysis and Delivery among Remote Sites," *Information Modeling and Knowledge Bases*, vol. XIX, IOS Press, pp. 282–289, 2008.
- 2 Nakanishi, T., Zettsu, K., Kidawara, Y., and Kiyoki, Y., "A Context Dependent Dynamic Interconnection Method of Heterogeneous Knowledge Bases by Interrelation Management Function," *Information Modelling and Knowledge Bases XXI*, IOS Press, pp. 208–225, March 2010.
- 3 Nakanishi, T., Zettsu, K., Kidawara, Y., and Kiyoki, Y., "Approaching to Interconnection of Heterogeneous Knowledge Bases on a Knowledge Grid," In *Proceeding of The International Conference on Semantics, Knowledge and Grid (SKG 2008)*, Beijing, China., pp. 71–78, Dec. 2008.
- 4 Nakanishi, T., Zettsu, K., Kidawara, Y., and Kiyoki, Y., "SAVVY Wiki: A Context-oriented Collaborative Knowledge Management System," *Proc. of ACM Intl. Symp. on Wikis and Open Collaboration (Wikisym2009)*, P. 106, Oct. 2009.
- 5 Iwazume M, Kaneiwa K, Zettsu K, Nakanishi T, Kidawara Y, and Kiyoki Y., "KC3 Browser: Semantic Mashup and Link-free Browsing," *Proceedings of the 17th International World Wide Web Conference (WWW 2008)* 1209–1210, 2008.
- 6 Zhang, R., Zettsu, K., Kidawara, Y., and Kiyoki, Y., "A Decentralized Architecture for Resource Management of Group-based Distributed Systems," *Journal of Frontiers of Computer Science in China (FCSC)*, pp. 224–233, 2008.
- 7 Cannataro, M. and Talia, D., "The Knowledge Grid: Designing, Building, and Implementing an Architecture for Distributed Knowledge Discovery," *Communications of the ACM*, Vol. 46, No. 1, pp. 89–93, 2003.
- 8 Zettsu, K. and Kiyoki, Y., "Towards Knowledge Management based on Harnessing Collective Intelligence on the Web," in *Proc. of the 15th International Conference of Knowledge Engineering and Knowledge Management -- Managing Knowledge in a World of Networks -- (EKAW2006)*, *Lecture Notes in Computer Science*, Vol. 4248, 2006, pp. 350–57.
- 9 Foster, I., Kesselman, C., and Tuecke, S., "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," *International Journal of High Performance Computing Applications*, Vol. 15, No. 3, pp. 200–222, 2001.
- 10 Papazoglou, M. P. and Georgakopoulos, D., "Service-Oriented Computing," *Communications of the ACM*, Vol. 46, No. 10, pp. 24–28, 2003.
- 11 *Future of European Grids, Grids and Service- Oriented Knowledge Utilities, Next Generation Grids (NGG) Expert Group Report 3*, 2006.
- 12 Fu, X., Bultan, T., and Su, J., "Analysis of Interacting BPEL Web Services," in *Proc. of the 13th International Conference on World Wide Web*, 2004, pp. 621–630.
- 13 The Globus Alliance, <http://www.globus.org/>
- 14 T. Ishida, "Language Grid: an infrastructure for intercultural collaboration," *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96–100, keynote address, 2006.
- 15 A. Bramantoro, T. Tanaka, Y. Murakami, U. Schäfer, and T. Ishida, "A Hybrid Integrated Architecture for Language Service Composition," *IEEE International Conference on Web Services (ICWS-08)*, pp. 345–352, 2008.
- 16 Y. Murakami and T. Ishida, "A layered language service architecture for intercultural collaboration," *International Conference on Creating, Connecting and Collaborating through Computing (C5-08)*, 2008.
- 17 Y. Murakami, M. Tanaka, D. Lin, and T. Ishida, "Service Grid Federation Architecture for Heterogeneous Domains," *International Conference on Services Computing (SCC-12)*, 2012.

- 18 T. Ishida, A. Nadamoto, Y. Murakami, R. Inaba, T. Shigenobu, S. Matsubara, H. Hattori, Y. Kubota, T. Nakaguchi, and E. Tsunokawa, "A Non-Profit Operation Model for the Language Grid," International Conference on Global Interoperability for Language Resources, pp. 114–121, 2008.
- 19 K.-S. Kim, R. Lee, and K. Zettsu, "mTrend: Discovery of topic movements on geo-microblogging messages," In Proc. of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS), pp. 529–532, 2011.
- 20 K.-S. Kim, K. Zettsu, Y. Kidawara, and Y. Kiyoki, "Moving Phenomenon: Aggregation and Analysis of Geo-time-Tagged Contents on the Web," In Proc. of the 9th International Symposium on Web and Wireless Geographical Information Systems, pp. 7–24, 2009.

(平成 24 年 6 月 14 日 採録)



ぜっつ こうじ  
是津耕司

ユニバーサルコミュニケーション研究所  
情報利活用基盤研究室室長  
博士 (情報学)  
データベース、データ工学、情報マネ  
ジメント、情報検索  
zettzu@nict.go.jp